Engineered Ferritin as a Nanoparticle Platform for Vaccine Delivery Via Molecular Dynamics Simulation and Experiment

Yiran Qu

A Thesis submitted for the degree of Doctor of Philosophy

School of Chemical Engineering and Advanced Materials

Faculty of Engineering, Computer and Mathematical Sciences

The University of Adelaide

October 2021

I dedicate this Thesis to my parents

Yan Qu and Quan Dong

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university of other tertiary institution to Yiran Qu and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this Thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my Thesis to be made available on the web, via the University's digital research repository, the Library Search and through web search engines, unless permission has been granted by the University to restrict access for a period.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Scholarship.

Name: Yiran Qu

Signed..... ..... Date.....23/10/2021.....

EXECUTIVE SUMMARY

Researchers have spent decades developing safe and efficacious vaccine to protect and save lives. The current COVID-19 pandemic has sharply drawn attention to the need for innovative vaccine development. Traditional vaccines are based on live attenuated virus strains, or inactivated (killed) pathogens. Significant weaknesses with traditional vaccines include, low immunogenicity, and high risk of the virus replicating. Modern vaccines are developed to overcome these challenges. Recent studies have shown that epitope-based chimeric (EBC) vaccine is practically promising as one option because there is a broad selection of molecular sizes, highly repetitive structures to induce immune responses, and importantly, scalable and cost-effective production processes that are well developed.

Human ferritin heavy chain (HFn), is structured by 24 identical subunits with each of molecular weight 21 kDa to form a spherical assembled structure with outer and inner diameters of 12 nm and 8 nm, respectively. HFn has a number of advantages as a vaccine carrier, including: it 1) has a robust thermal and chemical stability; 2) displays antigens in a well-organised manner to induce potent immune responses; and 3) is safe with good biocompatibility, biodegradability and low toxicity. Therefore, there is increasing interest in it as a protein nanocage to develop EBC. EBC vaccine consists of 3 main parts, protein nanocage (NPC), epitope and linker.

Currently, approaches to develop vaccine rely significantly on extensive experiments that can result in high failure rates and costs. Additionally, there are limited studies focusing on molecular design of HFn (insertion sites, linker design and variant study) in order to enhance protein stability and boost vaccine immunogenicity. Significantly, a comparison of immunogenicity between HFn and other NPC, such as Hepatitis B Core (HBc) virus like

particles (VLPs), is not yet reported.

The overarching aim of this Thesis work was therefore to bridge these research gaps and apply molecular dynamics simulation (MDS) and judicious experiments to develop engineered HFn as a protein nanocage to develop novel EBC vaccine. The research focussed on 5 coordinated steps: 1) Molecular design of engineered ferritin inserted with epitopes at 2 different locations (N-terminus and C-terminus) and stability investigation; 2) Flexible linker length design to affect engineered ferritin stability against thermal- and chemical denaturants; 3) Engineered ferritin variant study via simulation and experiment; 4) Immunogenicity study of engineered ferritin inserted epitope at N- and C-terminus; and, 5) Immunogenicity study of HFn and HBc carrying the same epitope.

In Step 1), model epitope Epstein-Barr Antigen 1 (EBNA1) was inserted at 2 locations, namely, N-terminus (E1F1) and C-terminus (F1E1). Protein surface hydrophobicity and thermal stability were predicted by MDS and validated by experiments. In Step 2), the effect of linker length on protein stability, short (3 residues) and long (15 residues) flexible linkers were inserted between the epitope and protein cage at N/C-terminus to form $E_1L_{15}F_1$, $F_1L_{15}E_1$, $E_1L_3F_1$, and $F_1L_3E_1$, respectively. Protein surface hydrophobicity and protein stability against thermal- and chemical denaturants were assessed experimentally. In Step 3), hot spots were predicted by MDS and variants (C1, C2, C3, C4 and C5) were designed to replace hot spots with other residues. C1 and C2 were built by replacing predicted hot spots with non-charged, or positive charged, residues. C3, C4 and C5 were constructed by replacing native hydrophobic residues with more hydrophobic or more hydrophilic residues. Molecular characterization, protein surface hydrophobicity and thermal stability by experiments were demonstrated.

To investigate the effect of insertion site of HFn on vaccine immunogenicity, Step 4), characterization study was demonstrated to compare the structural difference between E1F1 and F1E1. IgG titer, proliferation index and memory T cells differentiation were performed to determine humoral and cell-mediated immune response induced by E1F1 and F1E1. In Step 5), a comparison of HFn NPC with HBc NPC, the immunogenicity of E1F1 (HFn NPC) was compared with E1H1 vaccine (HBc NPC) inserted with the same model epitope EBNA1.

The key findings from this Thesis are:

1. A combined approach of MDS with experiment has been successfully demonstrated to design protein structure and predict protein stability. This approach can significantly reduce experimental cost and failure rate of designed vaccine.
2. Both MDS and experiment have shown that C-terminus insertion significantly boosts protein stability. This may potentially enhance vaccine efficacy over N-terminus insertion *in vivo* study. A more comprehensive vaccine efficacy study is required to further validate. However, this preliminary study is useful to provide guidance for design of NPC with multiple insertion sites.
3. Long flexible linker has less impact on protein stability against thermal and chemical denaturants when compared with short flexible linker that provides useful guidelines on designing linker length in EBC vaccine.
4. C-terminus variant study highlighted the importance of helix E on stabilising assembled protein conformational and thermal stability. This finding has increased understanding of HFn molecular structure in designing stable vaccine.
5. Compared with HBc-VLP, HFn is advantageous to significantly boost cell-mediated immune responses because of the stronger binding to T cell immunoglobulin. However, HFn induces weaker humoral and proliferative responses.

Collectively these findings will significantly improve understanding in development of HFn as a vaccine carrier through processes of molecular design, vaccine stability and immunology.

Research findings will therefore be of immediate practical benefit and interest to a wide range of researchers and manufacturers for innovative EBC vaccine development.

ACKNOWLEDGEMENTS

I gratefully acknowledge my principal supervisor Associate Prof. Jingxiu Bi for her support in assisting development of my professional skills as a researcher and in encouraging me. I gratefully thank my co-supervisor Prof. Anton Middelberg for his consistent supervision, kind encouragement and constructive suggestions on experiments design, critical skills development, manuscript polishing and future career development. I am particularly grateful to my co-supervisors Prof. Yan Sun and Dr Yan Jiao for their helpful support on computer simulation and in proof reading manuscripts. I thank Associate Prof. Kenneth Davey, FIChemE, CEng, CSci, for his support in developing my academic writing in terms of logical structure and accuracy.

I gratefully appreciate my parents Yan Qu and Quan Dong for their endless love and encouragement. Thanks to their love and care, I always hold my belief on completing PhD and keep improving myself to be better. My deep and sincere gratitude to my grandmother Guiliang Li, my husband Weijun Zhu, my aunts Jie Qu, Jun Dong, Xin Dong, and my lovely cousins Zhaoyu Ren, Bo Yang and Shuang Wu, for their continuous love, support and care. It is my pleasure to thank my lovely friend Lihong Xiao who always look after me like my sister.

I thank Prof. Yongdong Liu, Prof. Songping Zhang, Prof. Yingli Wang, Prof. Robert John Falconer and Dr Luis Toronjo Urquiza for providing support in designing experiments. I thank also Dr Fabien Voisin from the HPC support team for providing useful inputs to quickly obtain desired simulation results.

I would like to especially thank my best friend Shuang Yin for always supporting me and providing and help in amending my manuscripts. I would also thank my team members Bingyang Zhang, Nhat Hoang Huynh, Lukas Gerstweiler and Lijie Wang for their contribution on my research development; and Seonho Yun, Afshin Karami, Thai Thao Ly, Yechuan Zhang, Xue Liu, Yanmin Song and Hong Luo for collaboration, support and friendship.

I am grateful to all staff in School of Chemical Engineering & Advanced Materials for both Administration and Laboratory services. I would like to acknowledge the University of Adelaide for providing Adelaide Graduate Research Scholarship, and Adelaide Graduate Center in supporting me at each research milestone.

TABLE OF CONTENTS

LIST OF FIGURES

Figures in appendix are not listed here because repeating figures have been presented in Chapter 2.

LIST OF TABLES

Tables in appendix are not listed here because repeating tables have been presented in Chapter 2.

ABBREVIATIONS

In the order of alphabet.

3D	Three-dimensional
ADCC	Antibody-dependent cell-mediated cytotoxicity
APC	Antigen presenting cells
AS	Ammonium sulfate
AYEwt	Human procarboxypeptidase A2
AWSEM	Associative memory water-mediated structure and energy model
BSA	Bovine serum albumin
C_α	Central carbon atom
CD	Circular dichroism
CEACAM	Carcinoembryonic antigen-related cell adhesion
CGMD	Coarse-grained molecular dynamics
CTL	Cytotoxin T lymphocyte
Cyt	Cytochrome
D	Aspartate (Asp)
DC	Dendritic cells
DhlA	Haloalkane dehalogenase
DLS	Dynamic light scattering
DS	Discovery Studio

DSC	Differential scanning calorimetry
E	Glutamate (Glu)
EBC	Epitope based chimeric vaccine
EBNA1	Epstein-Barr nuclear antigen 1
EBV	Epstein-Barr Virus
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetra acetic acid
E1F1 ($E_1L_{15}F_1$)	EBNA1-HFn with 15 residues linker
$E_1L_3F_1$	EBNA1-HFn with 3 residues linker
ELISA	Enzyme-linked immunosorbent assay
EV71	Enterovirus 71
F1E1 ($F_1L_{15}E_1$)	HFn-EBNA1 with 15 residues linker
$F_1L_{15}E_1$	HFn-EBNA1 with 15 residues linker
FRET	Forster resonance energy transfer (Spectroscopic Ruler)
GdnHCL	Guanidinium chloride
H	Histidine (His)
HA	Hemagglutinin
HB	Hydrogen bond
HBcAg	Hepatitic B core antigen
HBc	Hepatitis B virus core

HbsAg	Hepatitic B surface antigen
HCP	Host cell protein
HFMD	Hand, foot and mouth disease
HFn	Human ferritin heavy chain
HIC	Hydrophobic interaction chromatography
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HPC	High performance computing
HPV	Human papillomavirus
I	Isoleucine (Ile)
IEC	Ion exchange chromatography
IL-2	Human interleukin-2
K	Lysine (Lys)
LFn	Human ferritin light chain
LN	Lymph nodes
MALS	Multiple angle laser light scattering
MBP	Myelin basic protein
MD	Molecular dynamics
MDS	Molecular dynamics simulation
MFIC	Maximum fluorescence intensity change (%)

MHC	Major histocompatibility complex
MM-PBSA	Molecular mechanical Poisson-Boltzmann surface area
MOG	Myelin oligodendrocyte glycoprotein
MW	Molecular weight
mRNA	Messenger RNA
N	Asparagine (Asn)
NP	Nanoparticle protein
NPC	Nanoparticle protein cage
NNA	Normal node algorism
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PDB	Protein data bank
PEG	Polyethulene glycol
P*f*CelTOS	Protein for ookinetes and sporozoites
pI	Isoelectric point
Q	Glutamine (Gln)
R_g	Radius of gyration
RES	Reticuloendothelial system

RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SASA	Solvent accessible surface area
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
TBM	Template-based modelling
TCR	T cell receptor
TEM	Transmission electron microscopy
TFM	Template-free modelling
TIM-2	T cell immunoglobulin and mucin domain-2
TMB	Tetramethylbenzidine
VLP	Virus like particles

CHAPTER 1 INTRODUCTION

1.1 Background

Epitope-based chimeric (EBC) vaccine is constructed by nanoparticle protein cage (NPC), inserted epitope/antigen, and linkers. There is an urgent need to find innovative vaccine carriers of EBC vaccine to replace traditional vaccines, such as live attenuated strains of pathogen or inactivated killed pathogen vaccines which are facing the challenges of the safety risks and low immunogenicity [1]. To overcome these challenges, there is a significant potential in developing EBC vaccine.

EBC vaccine is promising to be developed because of: 1) a broad selection of particle sizes from 5 to 100 nm [2], 2) cost-effective production [3], and; 3) highly repetitive structures to induce immune response (s) [3-5].

Human ferritin heavy chain (HFn) is one of examples of NPCs. Wild-type HFn comprises 24 identical subunits (each 21 kDa), forming a spherical hollow cage structure 12 nm in diameter [6]. As a vaccine carrier, HFn has been investigated with 3 potential insertion sites, namely, the N-terminus, C-terminus and the flexible loop region. The N-terminus of HFn is located outside the ferritin cage, while C-terminus is located inside the ferritin cage [7]. HFn demonstrates a number of advantages as a nano vaccine carrier, including: it 1) can carry antigens and expose immunogens in a repetitive and well-organised manner by genetic modifications [8,9]; 2) is thermally and chemically stable [10], and can be engineered for enhanced physicochemical and biological properties [11]; and, 3) is safe with low toxicity, good biodegradability and biocompatibility [12].

Besides NPC, linker is another important factor to design a stable and efficacious EBC vaccine. The linker is essential to provide significant separation between NPC and epitope to maintain

individual functionality. A lack of appropriate linker leads to problems, for example, protein misfolding [13,14], low yield in protein production [13,15], and/or impaired bioactivity [13,16,17]. In addition, the immunogenicity of epitopes presented on a NPC is significantly influenced by length and rigidity of the linker [18].

Molecular dynamics simulation (MDS) has been widely applied in biological research to solve these problems, for example, epitope immunogenicity prediction [19,20], fusion protein structure prediction [21-23], and protein stability prediction [24,25]. Therefore, by MDS prediction, the successful rate of EBC vaccine is likely increased and the period for the innovative EBC vaccine development can be shorten. MDS can filter and establish the most stable molecular structure for vaccine candidates and later validated via experiment with reduced time and cost.

There are 5 research gaps of developing engineered HFn as a vaccine carrier,

1) Vaccine development purely based on experiments results in higher failure rates and costs.
2) Limited knowledge about the connection between the different insertion sites of HFn and corresponding protein stability and vaccine immunogenicity.
3) Limited knowledge about the impact of the flexible linker length to the stability of engineered HFns.
4) Limited knowledge about key amino acid residues of helix E in HFn C-terminus affecting the molecular properties of formed HFn-based EBC vaccines.
5) Knowledge gap of understanding the immunogenicity difference between different protein cage sizes as EBC vaccine carriers, for example, HFn (12 nm) and Hepatitis B virus core (HBc) (35 nm).

Therefore, developing a stable and highly immunogenic HFn-based EBC vaccine is essential to bridge these research gaps. This project for the first time demonstrated development of HFn-based EBC vaccine through, 1) molecular design insertion sites and linker length; 2) purify and characterize protein stability; and 3) immunology study.

1.2 Aims and Objectives

In this PhD project, the overreaching aim of this work is to design and develop an engineered HFn-based vaccine carrier by using Epstein-Barr nuclear antigen 1 (EBNA1) as a model epitope through MDS and experiment. As is illustrated in Figure 1.1, this project is conducted by 3 main components, molecular design, protein purification and characterization, and vaccine immunology study.

Molecular design is to construct protein structures considering 2 different insertions sites (N- and C-terminus) and linker length. Additionally, molecular design is used to predict key residues affecting HFn stability and therefore design corresponding variants. Protein purification and characterization are to produce purified engineered HFn for examining protein stability. Vaccine immunology study is to determine the immunogenicity of engineered HFn, and to compare with other EBC vaccine, for example, VLP hepatitis B virus core.

The objectives and tactics are as follows:

1. To generate molecular designs of engineered HFns with inserted EBNA1 epitope at two (2) different positions (N- and C-terminus), namely, E1F1 and F1E1; and to predict protein hydrophobicity and thermal stability by MDS and validate by experiment (**Chapter 3**).
2. To design flexible linker length (3 or 15 residues) for engineered HFns-based EBC vaccine and characterize protein hydrophobicity and stability against thermal and chemical denaturants by experiment (**Chapter 4**).
3. To predict key residues affecting HFn stability by MDS, and therefore to design variants and study protein stability by MDS and experiment (**Chapter 5**).

4. To compare immunogenicity of N- and C-terminus insertion by vaccine immunology study (**Chapter 6**).
5. To compare immunogenicity between HFn-based EBC vaccine and VLP-based EBC vaccine (**Chapter 7**).

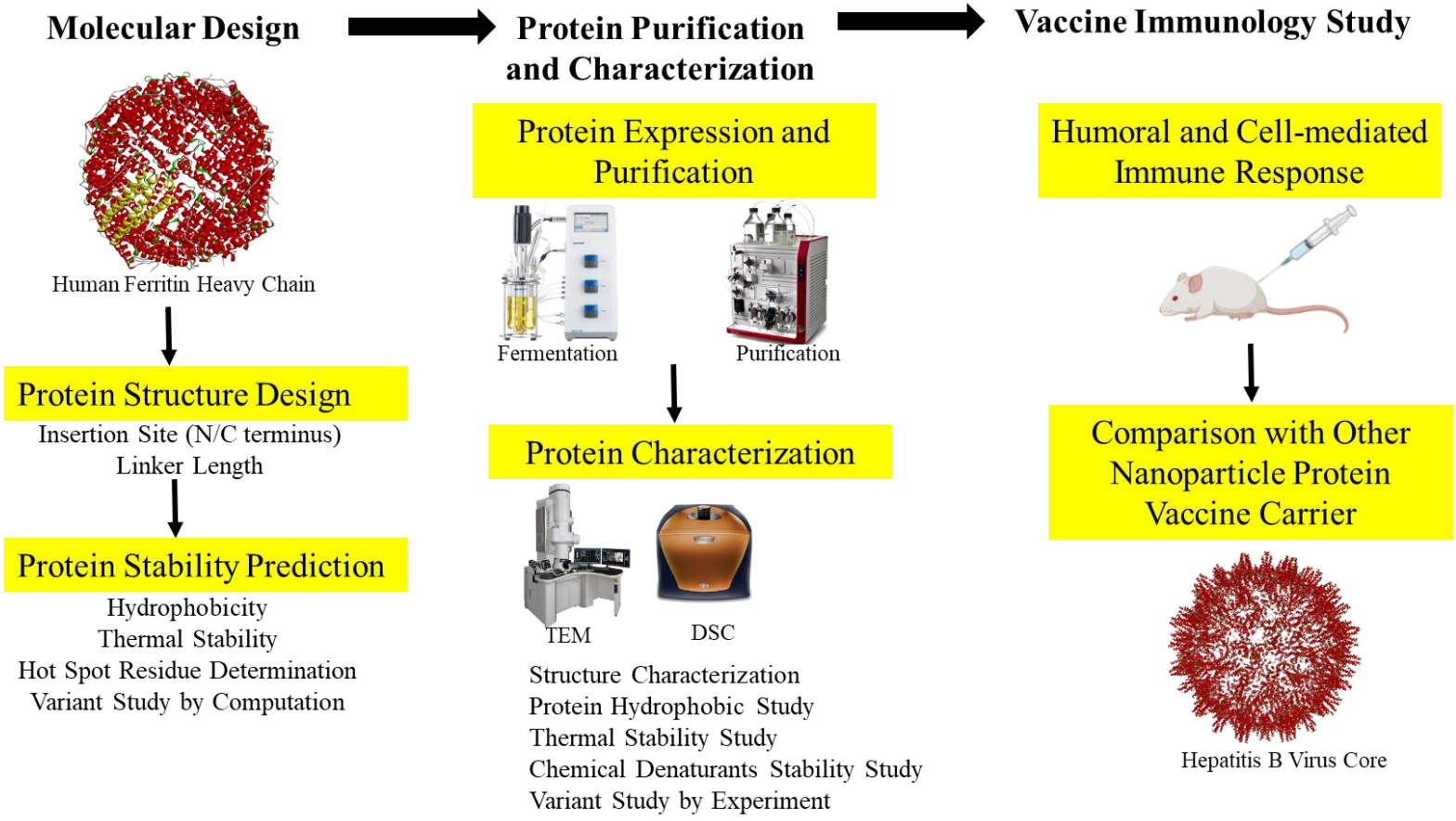


Figure 1.1 The design and experimental structure of the PhD project.

1.3 Thesis outline

This Thesis consists of eight chapters (Table 1.1). The content of each chapter is summarised below. Chapter 3, 4, 5, 6 and 7 are presented by the format of publication.

In Chapter 1, Introduction. This chapter includes the background, current research gaps of using HFn as a vaccine carrier, project importance, aim and objectives, Thesis outline of this PhD project.

In Chapter 2, literature review. This chapter describes limitations of other innovative vaccine carriers, advantages of nanoparticles as vaccine carriers, advantages and potentials of HFn as vaccine carriers, application of MDS on protein stability predictions. Furthermore, linker design, impact of epitope insertion sites and variants study associated with HFn's stability and immunogenicity are critically reviewed.

In Chapter 3, two insertion sites at human ferritin heavy chain (HFn) nanocage are studied to understand impact of different insertion sites on resulting protein stability. The model epitope Epstein-Barr nuclear antigen 1 is used to construct engineered ferritins E1F1 (N-terminus insertion) and F1E1 (C-terminus insertion). Protein hydrophobicity and thermal stability of these two engineered ferritins with inserted EBNA1 epitopes are predicted by MD simulations first and validated by experiments. Additional experimental studies on protein stability against pH and chemical denaturants are demonstrated to further understand the impact of insertion sites on protein stability in detail.

In Chapter 4, the impact of flexible linker length on stability of engineered ferritins is studied. Two groups of linkers, long (15 residues) and short (3 residues) are designed to insert at N-

terminus ($E_1L_{15}F_1$, $E_1L_3F_1$) or C-terminus ($F_1L_{15}E_1$, $F_1L_3E_1$) of engineered ferritin carrying EBNA1 epitope. Protein hydrophobic study and stability study against thermal and chemical denaturants are demonstrated to understand the impact of linker length in detail.

In Chapter 5, five variants (C1, C2, C3, C4 and C5) of engineered ferritin carrying EBNA1 epitope at C-terminus ($F_1L_3E_1$) are synthesised to understand the importance of C-terminus on engineered ferritin stability. MDS is performed to predict hot spots in helix E of C-terminus of HFn. Five variants are designed to replace hot spots in order to alter original electrostatic interface (C1 and C2) or hydrophobic interface (C3, C4 and C5). Protein hydrophobicity, conformational and thermal stability are predicted by MDS and validated by experiments to determine key residues to maintain stability of C-terminus in HFn.

In Chapter 6, the immunogenicity difference between E1F1 and F1E1 is compared from *in vivo* immune responses studies. Protein characterization and protein crystallization are performed to understand structural difference between E1F1 and F1E1. Humoral and cell-mediated immune responses induced by E1F1 and F1E1 are studied through IgG titers, T cell proliferation, lymphocyte activation and differentiation.

In Chapter 7, the immunogenicity of HFn is compared with the other nanoparticle protein, hepatitis B virus core (HBc), carrying same epitope, EBNA1. Humoral and cell-mediated immune responses induced by EBNA1-HFn (E1F1) and EBNA1-HBc (E1H1) are studied through IgG titers, T cell proliferation, lymphocyte activation and differentiation.

In Chapter 8, Conclusions and future directions. Conclusions derived from key findings in previous chapters and future research directions for HFn as a vaccine delivery platform are

included.

Appendix A A review paper of performance prediction of epitope-based chimeric vaccine using molecular dynamics simulations.

Table 1.1 Thesis outline.

Chapter	Type	Chapter title
1	Introduction	Introduction
2	Literature review	Literature review
3	Research output (publication)	Stability of engineered ferritin nanovaccines investigated by combined molecular dynamics simulation and experiments
4	Research output (publication)	Impact of flexible linker length on protein stability of engineered ferritin as a vaccine carrier
5	Research output (publication)	Engineered design of helix E on ferritin nanoparticles
6	Research output (publication)	Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope
7	Research output (publication)	Immunogenicity and vaccine efficacy boosted by engineering human heavy chain ferritin and chimeric hepatitis B virus core nanoparticles
8	Conclusion	Conclusion
Appendix A	Review paper	Performance prediction of epitope-based chimeric vaccine using molecular dynamics simulations: a critical review

1.4 References

1 Sahdev, P. *et al.* Biomaterials for nanoparticle vaccine delivery systems. *Pharmaceutical Research* **31**, 2563-2582, doi:10.1007/s11095-014-1419-y (2014).

2 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine: Nanotechnology, Biology and Medicine* **2**, 137-149, doi:https://doi.org/10.1016/j.nano.2006.07.009 (2006).

3 Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances* **41**, 107547, doi:https://doi.org/10.1016/j.biotechadv.2020.107547 (2020).

4 Bachmann, M. F. & Jennings, G. T. Vaccine delivery: A matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **10**, 787-796, doi:10.1038/nri2868 (2010).

5 Mohsen, M. O. *et al.* Major findings and recent advances in virus–like particle (VLP)-based vaccines. *Seminars in Immunology* **34**, 123-132, doi:https://doi.org/10.1016/j.smim.2017.08.014 (2017).

6 Han, J.-A. *et al.* Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine: Nanotechnology, Biology and Medicine* **10**, 561-569, doi:https://doi.org/10.1016/j.nano.2013.11.003 (2014).

7 Harrison, P. M. The structure and function of ferritin. *Biochemical Education* **14**, 154-162, doi:https://doi.org/10.1016/0307-4412(86)90203-7 (1986).

8 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: Epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

9 Li, C. Q., Soistman, E. & Carter, D. C. Ferritin nanoparticle technology: A new platform for antigen presentation and vaccine development. *Industrial Biotechnology* **2**, 143-147, doi:10.1089/ind.2006.2.143 (2006).

10 López-Sagaseta, J. *et al.* Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

11 Deshpande, S. *et al.* Thermostable exoshells fold and stabilize recombinant proteins. *Nature Communications* **8**, 1442, doi:10.1038/s41467-017-01585-2 (2017).

12 Zhen, Z. *et al.* Ferritins as nanoplatforms for imaging and drug delivery. *Expert Opinion*

on Drug Delivery **11**, 1913-1922, doi:10.1517/17425247.2014.941354 (2014).

13 Chen, X. *et al.* Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews* **65**, 1357-1369, doi:https://doi.org/10.1016/j.addr.2012.09.039 (2013).

14 Zhao, H. L. *et al.* Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Expression and Purification* **61**, 73-77, doi:10.1016/j.pep.2008.04.013 (2008).

15 Amet, N. *et al*. Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharmaceutical research* **26**, 523-528, doi:10.1007/s11095-008-9767-0 (2009).

16 Bai, Y. *et al.* Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proceedings of the National Academy of Sciences of United States of America* **102**, 7292-7296, doi:10.1073/pnas.0500062102 (2005).

17 Bai, Y. & Shen, W. C. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharmaceutical Research* **23**, 2116-2121, doi:10.1007/s11095-006-9059-5 (2006).

18 Wei, Y. *et al.* Evaluation of lumazine synthase from Bacillus anthracis as a presentation platform for polyvalent antigen display. *Protein Science* **26**, 2059-2072, doi:10.1002/pro.3243 (2017).

19 Hasan, M. *et al.* Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infection, Genetics and Evolution* **70**, 140-157, doi:https://doi.org/10.1016/j.meegid.2019.03.003 (2019).

20 Deng, H. *et al.* Development of a multivalent enterovirus subunit vaccine based on immunoinformatic design principles for the prevention of HFMD. *Vaccine* **38**, 3671-3681, doi:https://doi.org/10.1016/j.vaccine.2020.03.023 (2020).

21 Shamriz, S. & Ofoghi, H. Design, structure prediction and molecular dynamics simulation of a fusion construct containing malaria pre-erythrocytic vaccine candidate, PfCelTOS, and human interleukin 2 as adjuvant. *BMC Bioinformatics* **17**, 71, doi:10.1186/s12859-016-0918-8 (2016).

22 Mobini, S. *et al.* Computational design of a novel VLP-Based vaccine for hepatitis B virus. *Frontiers in Immunology* **11**, doi:10.3389/fimmu.2020.02074 (2020).

23 Cheung, N. J. & Yu, W. De novo protein structure prediction using ultra-fast molecular dynamics simulation. *PLOS ONE* **13**, e0205819, doi:10.1371/journal.pone.0205819 (2018).

24 Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports* **7**, 44651, doi:10.1038/srep44651 (2017).

25 Muneeswaran, G. *et al.* Molecular dynamics simulation approach to explore atomistic molecular mechanism of peroxidase activity of apoptotic cytochrome c variants. *Informatics in Medicine Unlocked* **11**, 51-60, doi:https://doi.org/10.1016/j.imu.2018.04.003 (2018).

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

The COVID-19 pandemic has underscored the importance of reliable vaccines. Vaccines are powerful solutions to challenges of new viral strains [1]. Traditional vaccine uses live-attenuated strains of a pathogen or inactivated killed pathogens. An attenuated vaccine, which contains live, whole bacterial cells or viral strains, is treated in such a way to have reduced virulence but maintain the ability to result in an immune response [2]. Live attenuated vaccine strains are highly immunogenic and effective at achieving high avidity and long-term immune responses [3], however, they bring many safety issues on replicating or reverting into viruses and causing disease. For example, dangerous pathogens (e.g. Human immunodeficiency virus (HIV)) are highly risky to completely be translated into inactivation process. There are significant potentials for virulent forms reversing [4,5]. Consequently, there may be more virulent strains generated in host organism due to mutagenic events. Inactivated virus vaccines cannot replicate or revert into forms that are more virulent. However, they induce a weaker immune reaction and require multiple dosages of administration [6]. Therefore, modern vaccine delivery platforms are urgent to be developed to solve these problems.

Recently, there are increasing interests for nanoparticles to be developed as vaccine delivery carriers because of their desirable properties for antigen delivery: 1) a precise control over particle physical properties such as size, shape, functionality and surface properties [7]; and 2) as adjuvants and as mimics of viral structures to enhance antigen presentation and possess strong immunogenicity [1]. Common nanoparticle-based vaccine platforms include liposome, polymers and epitope-based chimeric (EBC) vaccine.

Liposomes are part of lipid-based delivery vehicles ranging in size from tens of nanometres to several micrometres in diameter [8], and have been applied commonly for drug and vaccine

delivery application [3]. Liposomes have good biocompatibility, biodegradability, low toxicity and also are possible to be modified on surface or size [9]. Recent research focuses on developing liposomes-based vaccines by modifying physical chemical characterizations such as selection of lipid, charge, size and location of antigens [10]. For example, cationic liposomes were involved in a depot effect by ionic interactions with negatively charged cell membranes to extend antigen release at the injection site [11]. In addition, liposomes are found to be able to deliver both antigen and adjuvant to the same antigen presenting cell, which are important to induce potent immune response [3]. However, the major issue associated with liposomes is that plain liposomes are very quickly opsonised and sequestered by cells of the reticuloendothelial system (RES), mainly by liver [12].

Polymers have been widely used as vaccine adjuvants or carriers for vaccine delivery [13]. The size range is from 10 to 500 nm [14]. Polymers are very stable with low toxicity. It has been reported that, because the similarity of polymers with viruses regarding to the size and surface properties [15], polymers are able to increase the strengths of both B and T cell responses [16]. Additionally, polymers have significant loading capacity for single or multiple antigens and adjuvants [17]. However, the major challenge associated with polymers is that the accumulations are unexpected [18]. There are a range of polymeric materials with disadvantages including poor biodegradability, bad thermal stability and low degradation speed [14].

Epitope-based chimeric (EBC) vaccine can prevent above mentioned limitations as antigen display platform [19]. EBC vaccine consists of three (3) parts, nanoparticle protein cage (NPC), linker and epitope (Figure 2.1). NPC is also known as vaccine carrier, which is the key component needing to be carefully designed to achieve a desirable protein stability. Protein stability is important because: 1) structural stability of proteins is critical to presentation of

antigens on major histocompatibility complex (MHC) to induce potent immune response [20,21], and; 2) stable proteins can be readily expressed and purified [22,23]. A stable protein therefore ensures that EBC vaccine is safe and efficacious. Ferritin and virus-like particle (VLP) are two well-known examples of NPCs.

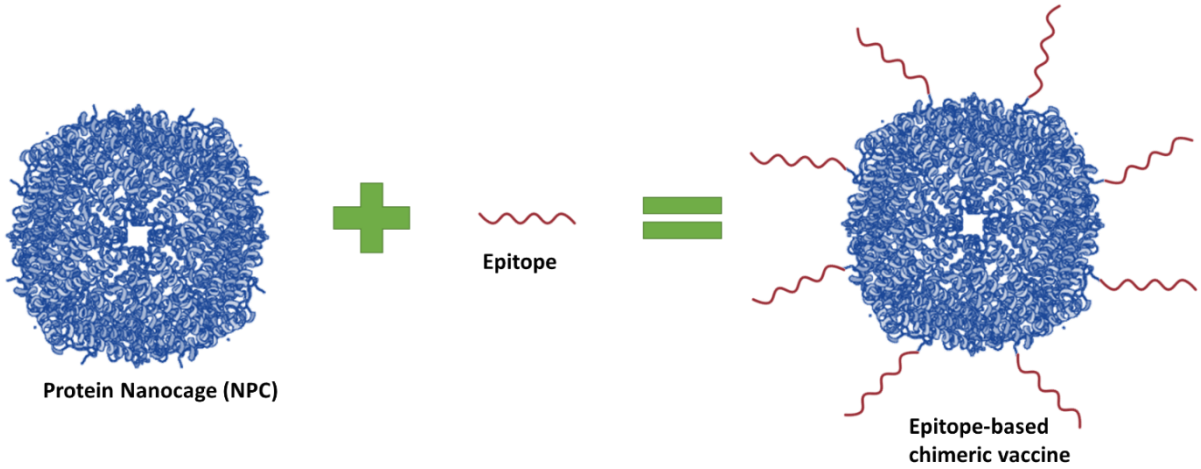


Figure 2.1 Schematic of epitope-based chimeric vaccine consisting of protein nanocage and epitope in BioRender©.

Ferritin is mainly investigated in this project to develop an EBC vaccine because of its three (3) main advantages, 1) well-organised structure to present antigens; 2) robust thermal and chemical stability; and 3) safety with low toxicity, great biocompatibility and biodegradability. However, despite an increasing global interest, there has not been a substantial review to assess the design of ferritin-based EBC vaccine in detail. For example, 1) limited knowledge about the connection between different insertion sites of ferritin and corresponding protein stability and therefore vaccine immunogenicity; 2) limited knowledge about design of linker to ensure ferritin-based EBC vaccine stability; 3) limited knowledge about key residues affecting ferritin stability; and 4) immunogenicity comparison between ferritin and other NPCs, such as VLP, as vaccine carrier. Besides that, 5) the design of stable and efficacious vaccine relying on experiments introduces high failure rates and costs [24]. Molecular dynamics simulation (MDS)

is a powerful tool to support design of stable and efficacious EBC vaccine. MDS is used to determine physical movement of three dimensional (3D) particles in a given dynamic environment and an applied force field. It was developed in the late 1970s to simulate atoms with biological relevance [25-27]. Over time because of parallel developments in high-performance computing (HPC), large protein molecules can now be investigated *in silico* using MDS.

In this chapter, these five (5) needs are addressed and critically evaluated, shown as highlighted sections in Figure 2.2. To address the first need, concept including ferritin structure and insertion sites are introduced. Recent studies related to ferritin insertion sites are assessed and evaluated. The second need refers to linker design. Classification of linkers and key criteria in flexible linker design are introduced. The third need is addressed by variant study. Recent ferritin variants studies are summarised and assessed to determine potential key residues in HFn affecting protein folding and stability. The fourth need is addressed by comparing ferritin with VLP, such as hepatitis B virus core (HBc), as the vaccine carrier. The benefit and limitation of HBc are assessed in this section. The last need is addressed by reviewing recent applications of MDS on predicting EBC vaccine performance through epitope immunogenicity prediction, protein structure prediction, and protein stability prediction.

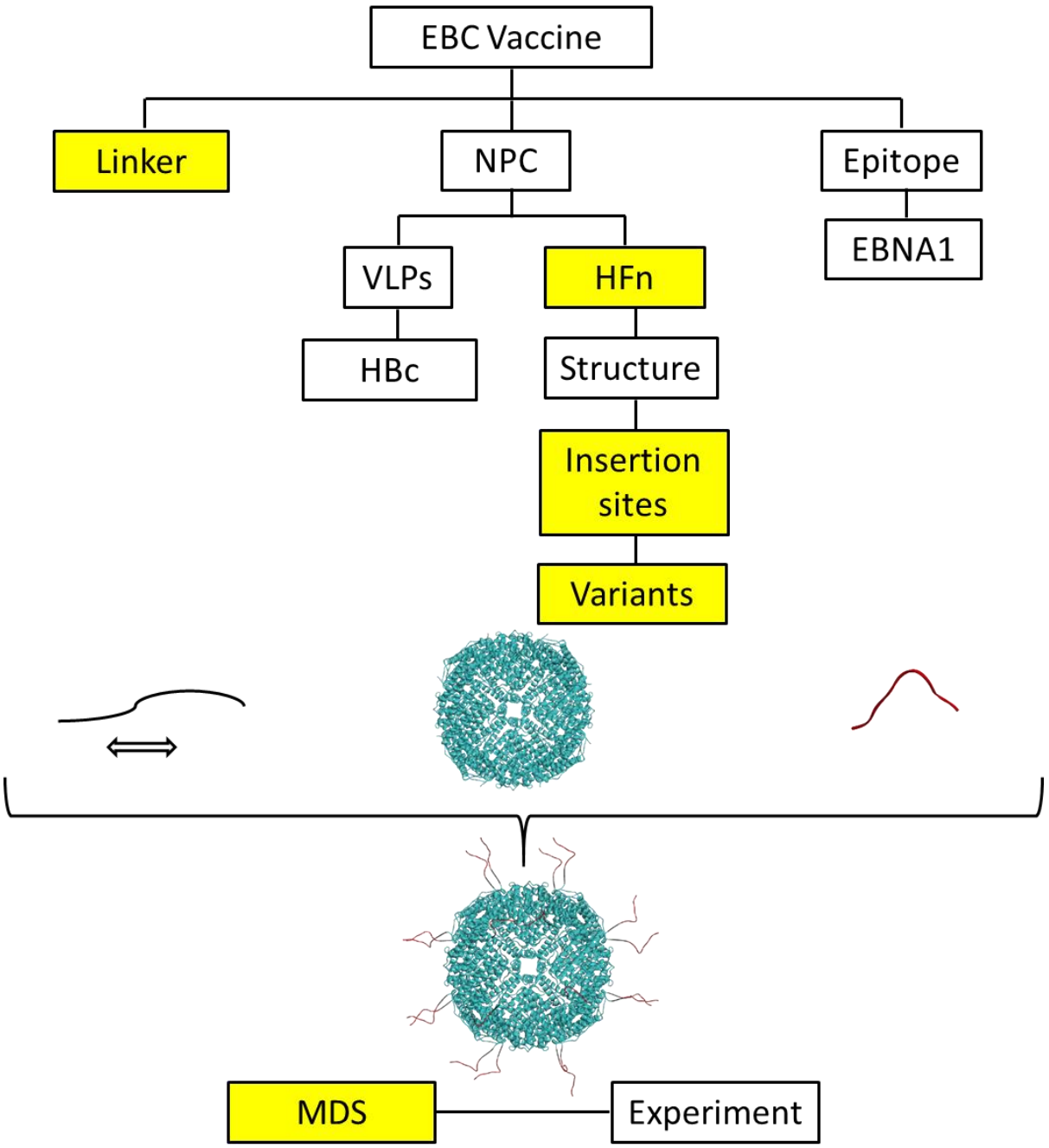


Figure 2.2 Flowchart of Chapter 2 structure. EBC stands for epitope based chimeric; VLPs stands for virus like particles; HFn stands for human ferritin heavy chain; HBc stands for hepatitis B virus core; and MDS stands for molecular dynamics simulation.

2.2 Ferritin-based NPC

2.2.1 NPC

NPCs are self-assembling biological nano cages with uniform size, structure and composition[28]. There is a broad selection of particle sizes at the nanometre scale (5 to 100 nm) [29], which ensures optimal interactions with various cells of the immune system [19]. NPCs can be produced via various cost-effective production systems, for example, bacteria, plant, insect and mammalian cells [28]. By both genetic and chemical modification techniques, nanocage is able to perform unlimited alterations to achieve desirable surface properties [29]. NPCs are highly repetitive structures with appropriate size to induce immune responses [15,28,30]. Ferritins and virus like particles (VLPs) are two well-known examples of nanoparticle proteins.

2.2.2 Ferritin structure

Ferritin is naturally occurring self-assembling protein nanoparticles, which can be found in plants, bacteria and animals [31]. The main roles of ferritin are storing iron and homeostasis [32]. Mammalian ferritin is a large protein (~ 504 kDa), and it has two encoding genes, namely, heavy chain ferritin subunit (HFn) and light chain ferritin subunit (LFn) with a molecular weight of 21 kDa and 19 kDa, respectively [33]. HFn plays a role in oxidising Fe^{2+} and Fe^{3+}, in contrast, LFn is responsible for iron nucleation [34]. Wild type HFn, comprises 24 identical subunits, forming a spherical hollow cage structure of 12 nm in diameter [35]. The inner cavitation of the spherical hollow cage structure has a diameters of 8 nm, (Figure 2.3A) [31]. Every ferritin subunit is made up of five helices, A to E, with a long BC loop and three (3) short loops (AB, CD and DE) connecting other helices (Figure 2.3B) [36].

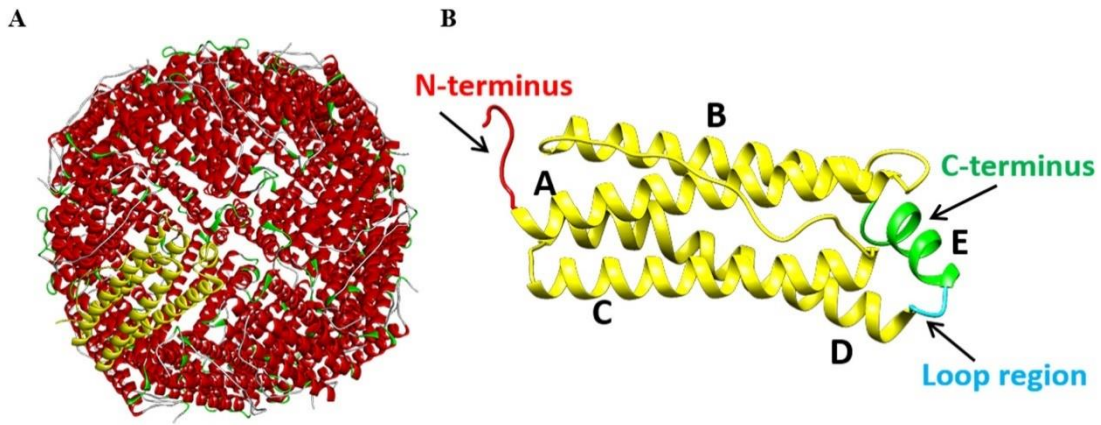


Figure 2.3 Human ferritin heavy chain mimic visualized with Discovery studio (DS). A) The native ferritin assembly structure. B) One subunit of ferritin structure made up of five helices A to E.

The native structure of ferritin is defined as *flip* structure (Figure 2.4), which is the most energetically favourable state. However, if exogenous peptides of heterologous proteins inserted at C-terminus are too large to be packed inside the ferritin cage, the helix E points outside creating a *flop* conformation (Figure 2.4) [37]. It has been approved by Jappelli *et al.* that fusion of 10 kDa peptide at C-terminal favoured ferritin into *flop* structure and the E-helix and the antigenic epitopes were extruded outside [38,39]. Based on the research taken by Luzzago *et al.*, the yield and stability for ferritin *flop* structure was similar to the wild type [37]. However, the immunogenicity difference for ferritin vaccine platform with *flip* or *flop* conformation is unknown.

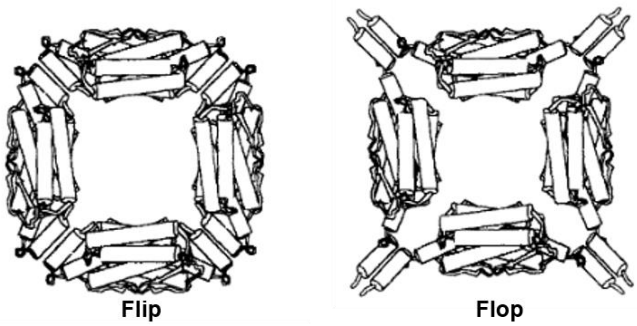


Figure 2.4 Schematic representation of ferritin flip and flop conformations reproduced from Jappelli *et al*. [37], copyright 1992, doi.org/10.1016/0022-2836(92)90905-Y.

2.2.3 Advantages of HFn as a vaccine platform

There is a significant potential to develop HFn as the vaccine platform because of various benefits. First, by genetic modifications, ferritin can carry antigens and expose immunogens in a repetitive and well-organised manner. Ferritin can display exogenous peptides [40,41] and heterologous proteins, such as the influenza virus hemagglutinin (HA) [42] and native-like HIV-1 envelope glycoprotein trimers.[43] Both examples have indicated that ferritin-based protein nanoparticles enhance the potency and breadth of virus immunity [40-43].

Second, ferritin is a naturally self-assembling nanoparticle with robust thermal and chemical stability [19], which makes it facile and economical to produce and store.

Third, HFn as one example of ferritin family, is highly safe with low toxicity, great biocompatibility and biodegradability [44].

Last, HFn can strongly bind to T cell immunoglobulin and mucin domain-2 (TIM-2) to regulate cellular immunity [45-47]. Lee *et al.* demonstrated that, compared with other nanoparticle proteins (e.g. *Escherichia. coli* DNA binding protein, *Thermoplasma acidophilum proteasome* and HBV capsid), ferritin vaccine carrier rapidly migrated to lymph nodes (LNs) with a short incubation time (< 1 min) and the accumulation of HFn in the LNs lasted for an adequately long time (6 days) [47]. This is presumably caused by the interaction between human ferritin and T lymphocytes [47]. The prolonged accumulation of HFn ensures that antigens can be sufficiently exposed to immune cells in LNs [47], which may in turn enhance the immunogenicity of a vaccine [48].

These advantages demonstrate that ferritin is applicable to be developed as a vaccine carrier.

2.2.4 Insertion sites of ferritin

There are 3 insertion sites in ferritin, N-terminus, C-terminus and the flexible loop region between helix D and E (Figure 2.3B). Table 2.1 summarises the research for inserting peptide into ferritin at various locations in the past decades. It can be concluded that ferritin nanoparticle is a robust platform to display peptides at various locations (N-terminus, C-terminus and flexible loop region). N-terminus insertion is an epitope insertion site in early research exploring ferritin as a vaccine carrier. For example, research conducted by Kanekiyo *et al*. [42], Sliepen *et al*.[43] and Wang *et al*.[49] six years ago selected the N-terminus as the insertion site of epitopes. Other groups frequently selected C-terminus or loop region as the insertion site when using ferritin as drug delivery carrier [47,50-53]. They often select flexible loop region between the 4th and 5th helices (D and E) for the insertion sites [54-56]. Up to date, to see if other insertions would result in a better dendritic cells (DCs) processing [35], some researchers started selecting the C-terminus or loop region for epitope insertion [40,51-53,57].

There is still limited knowledge on the relationship between the insertion site and immunogenicity when using ferritin as vaccine carrier. The research by Wang *et al.* displayed different length of epitopes from Enterovirus 71 (EV71) at N- and C-terminus and loop region (163 position) of ferritins to compare the immune responses [40]. They found that peptides inserted in loop region proceeded stronger immune response than N-terminus and C-terminus [40]. The other study demonstrated by Han *et al.* also used ferritin protein cage as the antigen carrier for dendritic cell (DC)-based vaccine development [35]. They observed that epitopes inserted at the loop region was not processed as efficiently as the epitopes at the C-terminus. It can be clearly noticed that the insertion site is significantly important to induce potent immune response. However, there are still limited studies investigating the potential relationship between the insertion site and corresponding immunogenicity.

In this Thesis, N-terminus and C-terminus are two key insertion sites to be investigated, because they represent two different protein conformations. Epitopes inserted at N-terminus are presented outside the protein cage, while epitopes inserted at C-terminus are packed inside the protein cage. The answer to the question ‘if epitopes exposed on the surface of nanoparticle are processed more efficiently compared with ones hidden inside the nanoparticle’ is still unknown.

Table 2.1 Insertion of peptide in ferritin case studies.

	Inserting target	Carrier	Insertion site	Treatment
Kanekiyo *et al.* [42]	Influenza virus hemagglutinin (HA) trimers	Ferritin from *Helicobacter pylori* (GenBank accession no. NP_223316)	N-terminus	Vaccine
Sliepen *et al.* [43]	Native-like HIV-1 envelope glycoprotein trimers (BG505 SOSIP.664)	Ferritin from *Helicobacter pylori* (GenBank accession no. NP_223316)	N-terminus	Vaccine
Wang *et al.* [1 49]	*N. gonorrhoeae* peptides	Ferritin from *Helicobacter pylori*	N-terminus Flexible loop between helices A and B	Vaccine
Han *et al.* [35]	OT-1 and OT-2 peptides from ovalbumin	Ferritin	C-terminus The 146 position	Vaccine
Wang *et al.* [2 58]	Enterovirus 71	Ferritin	N-terminus C-terminus Loop zone	Vaccine
Kang *et al.* [1 54]	Thrombin Cleavage peptide (GGLVPR/GSGAS)	Ferritin from *Pyrococcus furiosus* (Pf_Fn)	The 146 position of Pf_Fn, middle of the flexible loop connecting helices D and E.	Drug

Kang *et al.*[2] [55]	Fc-binding peptide (FcBP) (GGGGGGDCAWHLGELVWCTGGGGGAS)	Pf-Fn	The 146 position of Pf_Fn, middle of the flexible loop connecting helices D and E.	Drug
Jeon *et al.* [56]	Interleukin-4 receptor (IL-4R)- targeting peptide, AP-1	Ferritin-L-chain (FTL)	The 157 and 158 position in the exposed loop region between helices D and E.	Drug

2.2.5 Engineered ferritin variants study

Variant study is an efficient tool to understand important structure elements on protein folding or unfolding [59]. Research has shown that substitutions of amino acid residues can affect both protein stability and protein folding pathway [59-64].

Shown in Table 2.2, the current studies regarding mutations of ferritin focus on modifications at the C-terminus of ferritin, because the C-terminal region has been proven to play a major role in protein stability and assembly [65]. The N-terminal region, BC loop and threefold axis contribute minimally to ferritin folding and assembly, while the DE loop and C-terminal region impact more significantly on protein folding and stability. Based on the study demonstrated by Ingrassia *et al.*, the last six (6) non-helical residues (Figure 2.5) had no obvious effect on ferritin stability, and the extension of this area slightly reduced the solubility and capacity of assembly ferritin cages [65]. This research particularly highlighted the importance of Helix E on helical structure of the molecule along with protein folding. Fan *et al.* found similar results. They indicated that the BC Helix in BFR (a protein cage similar to ferritin) was less important than the Helix E for self-assembly [66]. Removal of Helix E resulted in a destabilized protein [66,67]. However, Luzzago *et al.* claimed that E helices were not important for human ferritin H-chain assembly [37], because they found that the mutations of the E helix exhibited similar yield and stability compared with wild type ferritin. The contradictory results from these studies highlights the importance of further understanding the role of Helix E on ferritin stability.

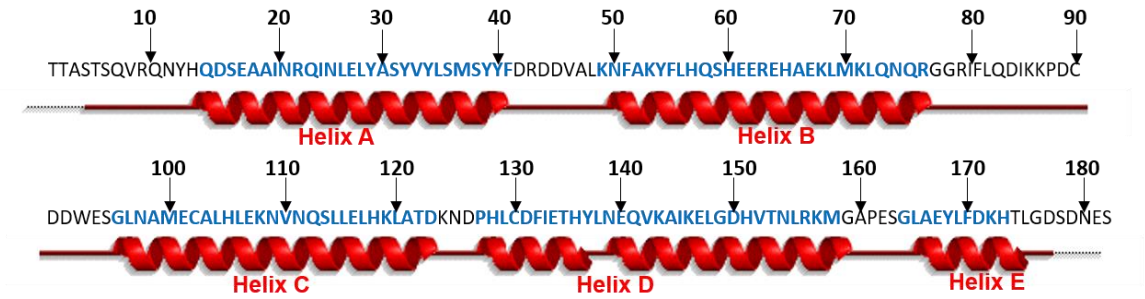


Figure 2.5 Schematic representation of ferritin human heavy chain secondary structure by EMBL-EBI. Residues within alpha helices are blue-bold shown in sequence.

Table 2.2 Variant study summary for human ferritin heavy chain.

Region	Variants	Results	Reference
N-terminus	Deletion of first 13 amino acid residues	Assembly	[68]
BC loop	Deletions of single amino acid residue (hydrophobic patch: L^{82} or I^{85} or P^{88})	• Disassembly (Delete L^{82} or I^{85}): shift polar residues into apolar ones • Assembly (Delete P^{88}): no shift the frame of hydrophobic interaction	[69]
	Duplication (P^{88}-L^{106}+$D^{91}D^{92}$ to NV)	Assembly: stability conserved	
DE loop	Deletion of A^{160} to S^{163}	Disassembly: DE loop is required for protein folding and stability	[38]
	Replace G^{159}, A^{160}, or P^{161} with positively charged residues	Assemble into native protein shell and form insoluble aggregate	
Threefold axis	Human ferritin H chain: replacement of D^{131} and E^{134} by H and A	Assembly: no major effect	[69]
C-terminus	Replacement of last 10 residues with L chain	Assembly: Improve stability	[69]
	Fusion α-peptide of β-galactosidase at E helix; Random variants at E α-helix	Assembly: reduce thermostability; interactions around E α-helix were eliminated	
	Deletions of last 7 amino acid residues	Assembly	
	Deletions of last 22 amino acid residues	Assembly	
	Deletions of last 28 amino acid residues	Disassembly	
	Deletions of amino acid residues (G^{163} to H^{173})	Assembly	[37]
	Deletions if last four amino acids in D helix (L^{155} to M^{158})	Disassembly	[37]

2.3 VLP based NPC

VLPs are composed of viral capsid proteins and self-assemble to produce hollow nanocage with the lack of infectious nucleic acids [28,70]. Most VLPs are in size range of 20 to 100 nm in diameter [71], which ensure them to freely enter the lymphatic vessels and optimal uptake by antigen presenting cells (APCs) [72]. VLPs are significantly applicable as vaccine platforms because they display antigenic epitopes in a multi-meric, repetitive, and highly spatially organized conformation, and demonstrate high epitope density [73]. This special property ensures VLPs induce potent humoral and cell-mediated immune responses [15,28,74-76].

One of the most well characterized VLPs is the hepatitis B virus core protein (HBc). HBc is made up of 180 equal subunits (T=3) with outer diameters of 27-31 nm [28,77]. There is also a proportion of HBc comprising 240 equal subunits (T=4) with an external diameter of 35 nm [78]. The molecular weight for each subunit is approximately 21 kDa. Because of its strong immunogenicity and extraordinary flexibility allowing insertion of up to 300 amino acids without affecting its ability to self-assemble into VLP structures [79-81], HBc has potential to develop as a vaccine carrier. Moreover, VLPs can be engineered to express foreign proteins by fusion or conjugation [82,83]. As a result, VLPs can protect against both virus and heterologous antigens [84].

However, there are four (4) main challenges involved with VLPs as vaccine carriers. First 1), owing to VLP structural complexibility, it appears to have limitations on expression systems. *E. coli* is a common cost-effective expression system as a primary laboratory workhorse bacterium. It is often preferred for expressing small proteins (usually < 100 amino acids long),

while for larger proteins (~ 27000 amino acids) with complex structures like VLPs, a more complex expression system such as mammalian cell, is used [85]. It has been reported that applying bacteria, yeast or insect cells expression systems may reduce soluble VLPs expression and form inclusion bodies [85]. Other complex expression systems such as mammalian cells, the successful rate for soluble proteins expression may be increased, however, a very high production cost will be involved. Second 2), the solvent conditions for VLPs are critical to maintain structure stability [85]. The lack of the viral genome reduces the stability of VLPs when the conditions are changed during purification [86]. VLPs are highly sensitive to pH conditions and salt [86]. Third 3), VLPs are very easy to aggregate [87]. At higher protein concentrations, VLPs may form insoluble precipitate [85]. Last 4), numerous impurities from host cells such as cell debris, host cell proteins (HCPs), DNA and lipids [86], are considerably difficult to be removed from VLPs during purification, which reduce VLP antigenicity significantly.

2.4 Linkers

The linker is a key component in design of EBC vaccine. Linkers are essential to provide separation distance between NPCs and epitopes to maintain individual functionality. A lack of appropriate linker therefore leads to problems, for example, protein misfolding [88,89], low yield in protein production [88,90], and/or impaired bioactivity [88,91,92]. Additionally, the immunogenicity of antigens presented on a nanoparticle protein cage is impacted by length and rigidity of the linker [93]. The selection of linker therefore directly affects protein stability [88, 94]. Therefore, the linker is important in design for stable epitope-based chimeric vaccine.

There are three groups of linkers, flexible, rigid and cleavable [95] (Figure 2.6). Flexible linkers are designed for protein cages that require a particular degree of movement, or interaction [88]. Compared with flexible linkers, rigid linkers, such as polypro line motifs [96] and α-helical linker [97], separate protein cage and epitopes more efficiently in rigid structures [88]. Rigid linkers have the potential to interfere with protein folding [98]. Cleavable linkers are used for releasing free functional domains *in vivo* [88].

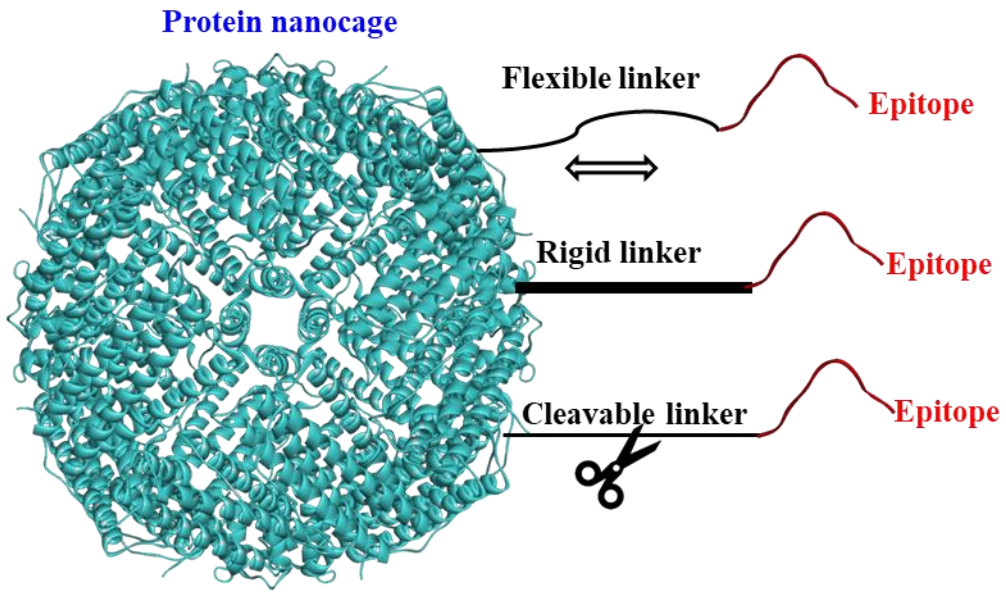


Figure 2.6 Schematic of the three (3) linker types by using software Discovery Studio 2021. Human ferritin heavy chain is used as example for the protein nanocage and is marked in blue-

colour. Epitope is marked red. Linkers are marked black. 1) *Flexible link* provides flexible distance between epitope and protein nanocage; 2) *Rigid linker* gives fixed distance between epitope and protein cage, and; 3) *Cleavable linker* releases epitope from protein nanocage.

Flexible linkers are the most commonly applied in epitope-based chimeric vaccine, because these: 1) can be readily designed with negligible changes on conformational stability of the protein nanocage, and; 2) provide significant distance between epitope and protein nanocage to function independently [42,43,88].

Amino acid composition and length of flexible linker are two (2) key design parameters. Flexible linkers are synthesised by small non-charged residues [99]. Gly and Ser residues, known as 'GS linker', are widely applied in fusion proteins because of good solubility and flexibility [88]. The property of a GS linker is altered by adjusting glycine content. Gly-rich linkers are highly significantly flexible [100]. Sabourin *et al*. designed a flexible linker $(Gly)_8$ between the open reading frame of yeast gene and nine (9) Myc epitopes [101]. This type of linker increases accessibility of an epitope to antibodies and/or to improving protein folding [88], [101]. Rosmalen *et al*. reported that Forster resonance energy transfer (FRET) efficiency was overall lower for linkers with less glycine [94]. This finding explains that stiffness of polypeptide linkers increases with decreasing glycine content. The minimum and maximum linker length, based on recent reports, is between 2 and 31 amino acids. Linker length is commonly selected from 5 to 11 residues [100]. This length needs to be optimal to maintain protein folding and stability. Changes in linker length and composition make a significant difference to protein folding kinetics [102].

There is however a present lack of understanding on the impact of flexible linker length on protein stability. Robinson *et al*. pointed out that the linker with 19 residues synthesised the most stable protein [102]. The adding or deleting of amino acids decreased protein stability. Nagi *et al*. demonstrated that that there is an inverse correlation between linker length and protein

stability [103]. Rop protein was used as a model system, and a series of flexible linkers with 1 to 10 Gly residues was applied. They reported that all variants maintained functionality. Stability against thermal and chemical denaturation was reduced with increasing flexible linker length. However, experimental results reported by Chen *et al*. were contentious. This is because it was claimed that longer flexible linker improves activity and stability of displayed esterase over shorter flexible linker [104]. The mechanism of impact of linker length on protein folding and stability is not understood. These studies however underscored that the length of flexible linker is significant to design and expression of stable folded fusion protein.

2.5 Model epitope Epstein-Barr nuclear antigen 1 (EBNA1)

Epstein-Barr Virus (EBV), belonging to the human herpesviruses family [105], infects B lymphocytes and some epithelial cells, which results in various disease counting with Burkitt's lymphoma, Hodgkin's disease, lymphomas and lymphoproliferative diseases [106]. There are more than ninety percent of human adult populations infected by EBV and approximately 1.5% of all cancers in the world associated with EBV [105]. The percentage causing cancers may increase for co-infection of EBV and other oncogenic viruses. EBV and human papillomavirus (HPV) are linked to 38 % of all virus-associated cancers [107,108], inclusive of cervical cancer, breast cancer, prostate cancer and lung cancer [107].

There is limited therapies for prevention of EBV-associated disease and no licensed vaccine available for EBV [105]. EBNA1 is selected to be the target, because it is expressed in all EBV-associated tumours to persist the viral genome in the cells when they multiply [105]. In addition, EBNA1 is preferentially recognized amongst latent EBV antigens simulating $CD4^+$ T cells [109]. It is also suggested by Münz *et al*. that EBNA1-specific $CD4^+$ T cell immunity should be enhanced to prevent and treat EBV-associated diseases [109]. Research demonstrated by Destro *et al*. confirmed that cytotoxic T lymphocytes (CTLs) specific for the HLA-B35/B53-presented cut EBNA1 epitope (amino acids 407-417: HPVGEADYFEY) was detectable in most HLA-B35 individuals and recognized EBV-transformed B lymphocytes [110]. The development of EBNA1 vaccine potentially prevent EBV infection and EBV related cancer.

In this Thesis, the linear epitope (HPVGEADYFEY) was truncated off from EBNA1 comprising aa 407 to aa 417, which is the main immunogenic domain of EBNA1 [110].

2.6 Molecular dynamics simulation (MDS)

2.6.1 Background of MDS in EBC vaccine design

There are however significant practical challenges to development of EBC vaccines[19,111-113]. Development includes but is not limited to: 1) molecular design/selection of epitopes; 2) design of energy-minimized fusion protein structures; 3) protein expression by host cell lines; 4) protein purification, and; 5) *in vitro* and *in vivo* testing.

Significant time-consuming and costly experiments are needed to develop desirable vaccine candidates. This is because there are many unknown factors, including epitope immunogenicity and protein stability. The result is often a low rate of successful vaccine production [24].

Molecular dynamics simulation (MDS) is an emerging design tool that appears to reduce the need for experimental testing and has potential to increase the likely success rate in vaccine development. MDS is used to determine physical movement of three-dimensional (3D) particles in a given dynamic environment and an applied force field. It was developed in the late 1970s to simulate atoms with biological relevance [25-27]. Over time because of parallel developments in high-performance computing (HPC), large protein molecules can now be investigated *in silico* using MDS.

A schematic showing how MDS is used to design and predict performance of proposed nanoparticle protein-based vaccines is given as Figure 2.7. It is seen from the figure there are three (3) serial steps, namely, prediction of: 1) epitope immunogenicity (Figure 2.7, M1); 2) protein structure (M2), and; 3) protein stability (M3), via *in silico* analyses.

Prediction of epitope immunogenicity by MDS (Figure 2.7, M1) filters the suitable epitope which is, most, 1) specific to a target disease, and; 2) efficacious. Epitopes *in silico* that show the strongest binding with receptors and greatest stability are the most likely to induce significant immune response [114]. Step M1 significantly improves success rates in later experimental immunological studies (Figure 2.7, E3) [114]. Step M2 is achieved via construction of energy-minimized molecular-structure for nanoparticle proteins (NPs) inserted with epitopes. If the NPCs have registered structural information in the Protein Data Bank (PDB) [115], M2 can be readily applied following assessing of structural stability data.

However where NPCs have not been registered, and limited molecular structural information is available, MDS becomes a critical tool to simulate protein folding and to predict energy-minimized protein structure with insertion of the epitopes [116]. Step M2 ensures protein structural resonance to support an optimal vaccine production (Figure 2.7, E1).

Step 3, M3 stability studies in MDS, have been found to be highly consistent with experimental data. This consistency between MDS prediction and experiment is advantageous to design the most stable vaccine with reduced time and cost. For example, protein thermal stability *in silico* agreed well with experimentally determined protein melting temperature using differential scanning calorimetry (DSC) [117,118]. MDS can provide protein secondary structure *in silico* information [119,120]. This is advantageous as it can replace experiment using circular dichroism (CD) [121, 122]. Protein stability studies via MDS achieves the same purpose, and it gives the most stable structures for experiment (E2) with both optimized time and cost. It is clear that there is therefore practical benefit in using MDS to predict performance of nanoparticle protein-based vaccine. Despite this however it has not been reviewed in detail.

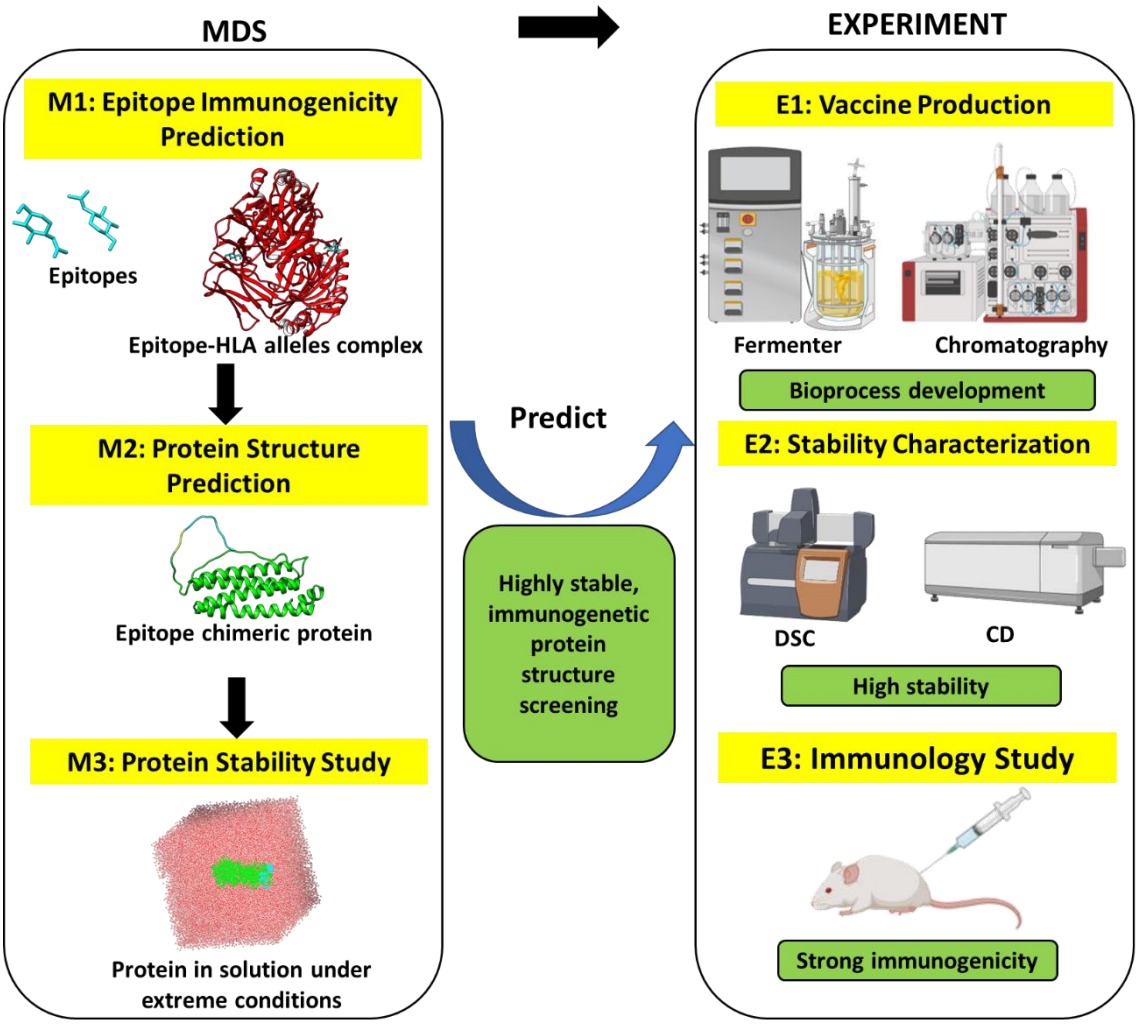


Figure 2.7 Schematic of Molecular Dynamics Simulation (MDS) for monitoring performance of nanoparticle protein-based vaccines in which DSC = Differential Scanning Calorimetry and CD = Circular Dichroism. Protein structures were created in UCSF Chimera. Cartoons of equipment prepared in BioRender©.

2.6.2 Prediction of epitope immunogenicity by MDS

It is most important in vaccination to achieve high immunogenicity by inducing potent immune responses. Therefore selection, optimization and validation of suitable epitopes is important to ensure efficacy and safety.

A flowchart for the method to epitope design and immunogenicity prediction via computation is presented as Figure 2.8 [123-127]. This shows serial steps including: 1) viral strain selection, 2)

protein sequence preparation, 3) epitope prediction, 4) vaccine structure construction, 5) allergenicity and toxicity, 6) population convergence analysis, 7) molecular docking analysis, and; 8) MDS.

Steps prior to molecular docking analyses, are performed based on particular case-to-case using a number of prediction tools including, IEDB [128], SYEPEITHI[129] and ProPred1[130]. However, they are all designed to achieve the structure of the most potent selected epitopes [123-127]. The immunogenicity for specific epitopes is then predicted by considering interactions between epitopes and viral strains.

Together, molecular docking analysis and MDS, are of particular interest in this Review. This is because molecular docking and MDS quantify interaction between epitope and alleles i.e. Molecular docking is the step prior to MDS to prepare a 'ligand-receptor' complex [131].

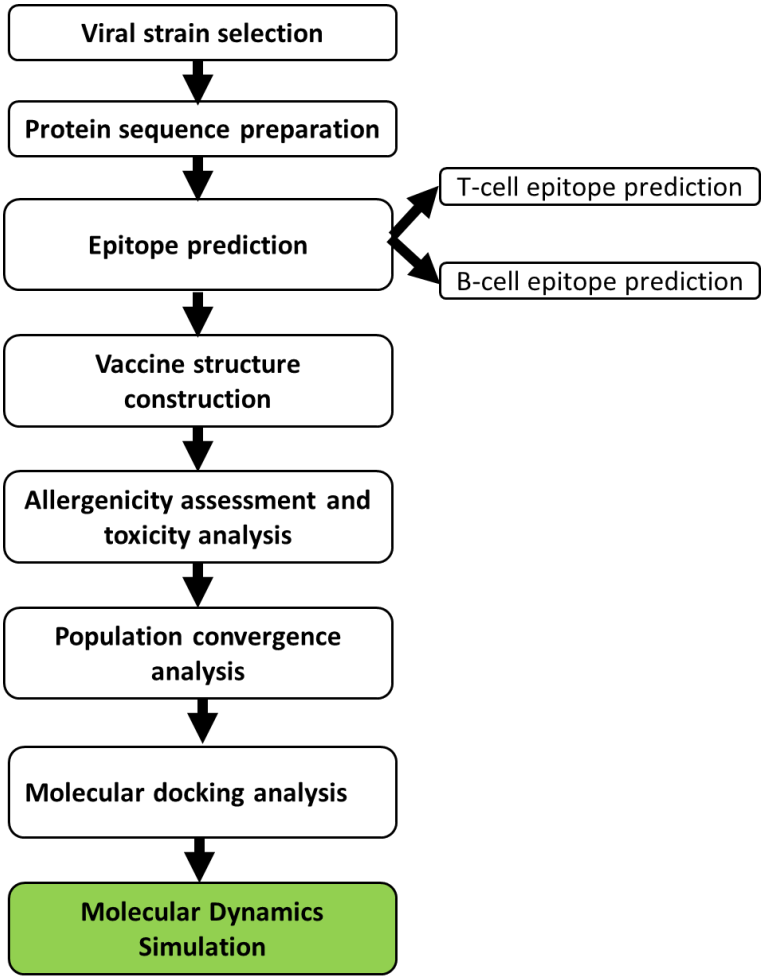


Figure 2.8 Flowchart for prediction of epitope immunogenicity.

Epitopes, or antigens, are identified as 'ligands', whilst different alleles and immune receptors are treated as 'receptors', Figure 2.9. The docking analysis scores these complexes based on binding affinities. The complex with the most negative value for binding affinity is transferred to MDS to validate the binding interaction [124], and determine stability [123].

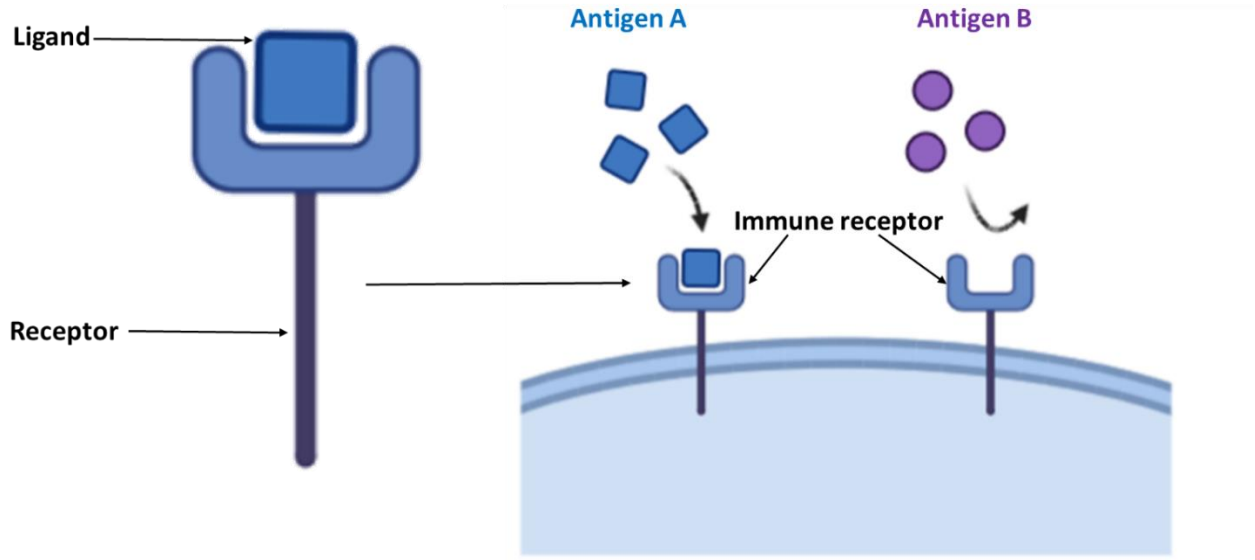


Figure 2.9 Schematic of antigen-immune receptors interaction in BioRender©. The antigen is considered as 'ligand' and immune receptor as 'receptor'.

2.6.2.1 Application of MDS in epitope immunogenicity prediction

Hasan *et al*. conducted research to develop vaccine against Marburg virus [123]. Through epitope prediction tool PEP-FOLD3, they obtained twelve (12) T-cell epitopes, six (6) from envelope glycoprotein and six (6) from matrix protein VP40. All selected epitopes were confirmed by performing docking analysis with Human leukocyte antigen (HLA) molecules, named, respectively, HLA-A*11:01 and HLA-DRB1*04:01.

It was found that the epitope with sequence VQEDDLAAGLSWIPF from envelope glycoprotein was bound by HLA-DRB1*04:01 (-7.8 kcal mol^{-1}), and the epitope with sequence VPAWLPLGIMSNFEY from matrix protein VP40 demonstrated binding energy of -7.0 kcal mol^{-1}) [123]. VP1-epitope 'APIDFDPVP' was found to interact with HLA-A*11:01 most strongly in terms of the least free binding energy (-9.5 kcal mol^{-1}) [123].

These authors focused on MDS to investigate stability and mobility of proteins with selected epitopes in terms of deformability, eigenvalues, B-factors and covariance [123]. Deformability determines the degree to which applying a force can make a given molecule change its shape. Lower deformability means less structural stability. Eigenvalue is related to the notion of stiffness.

The less the eigenvalue, the more readily likely the deformation [132]. The eigenvalue for the complex V1-TLR3 with highly immunogenic epitopes, such as VP1, was $1.02e^{-04}$. This value was considered as significantly low, therefore findings showed deformability for each residue was unlikely to occur. Therefore, it was concluded the filtered epitope was satisfactory for further processing (Figure 2.8). This work by Hasan *et al*. focused on using molecular docking analysis to determine immunogenicity of listed epitopes based on binding affinity. Importantly however, there was a lack of MDS to validate complex stability.

Docking analysis is actually weak in analysing interactions between epitopes and immune receptors. MDS will be more accurate in predicting binding interactions if a dynamic system is considered at the atomic level [133-138]. MDS docking analysis followed by MDS determine both binding interaction and complex stability [124,125,127,139].

Gupta *et al*. investigated applying docking analysis to develop vaccines against carcinoembryonic antigen-related cell adhesion (CEACAM). It has a highly conserved region in cancers [124]. Seven (7) predicted epitopes were docked to HLA allele. The complexes formed were confirmed using ligand interaction study in MDS.

These authors compared the stability of the epitope-HLA complexes to that of original HLA complex structures. The free binding energy was computed to predict immunogenicity of varied epitopes. Findings showed the designed vaccine was highly likely to be efficacious. This was because the: 1) epitopes bound strongly to the major histocompatibility complex (MHC) molecules were screened, and; 2) stability of complex with selected epitope over time has been validated.

2.6.2.2 Parameters to be determined

Lessons from the case studies above have underscored that the common parameters to predict epitope immunogenicity include: 1) molecular mechanical (MM) Poisson-Boltzmann surface area (PBSA) [140]; 2) root-mean-square deviation (RMSD); 3) root-mean-square fluctuation (RMSF), and; 4) radius of gyration (R_g) [124,125,127,139,141-144]. Definitions are given in Table 2.3.

Predictions of epitope immunogenicity are demonstrated through computation of these parameters and have been found to be consistent with experimental immune results. For example, Deng *et al*. developed a multivalent enterovirus subunit vaccine for the prevention of HFMD (hand, foot and mouth disease) [144]. By showing strong binding energy via the PBSA tool and high stability of formed complex via RMSD and RMSF, the designed multivalent enterovirus vaccine was then able to be expressed and purified. Test results *in vivo* demonstrated significant humoral and cell-mediated immune responses.

Recently, additional parameters have been identified through MDS to predict epitope immunogenicity. For example, two (2) constructs designed as potential vaccine for multiple sclerosis disease, namely, Construct 1 and Construct 2, can be considered, Figure 2.10. By showing that the surface area of myelin oligodendrocyte glycoprotein (MOG) antigen in

Construct 1 was similar to antigenic domain in Construct 2, and surface area of myelin basic protein (MBP) in Construct 1 was greater than for Construct 2. Construct 1 was concluded to carry antigenic property for vaccination [145]. This is because the construct with greater antigenic domain area is more likely to induce stronger immune response in the body.

Table 2.3 Common parameters determined in Molecular Dynamics Simulation (MDS) to study protein stability.

Parameter	Role
Root-mean-square (RMSD) deviation	Characterization of conformational change in proteins.
Radius of gyration (R_g)	Basic measurement of overall size of a chain molecule and indication of conformational change in protein rigidity.
Root-mean-square fluctuation (RMSF)	Determination of flexibility differences amongst residues.
Native contacts (Q)	Indication of native contacts preserved under conditions such as hyperthermal or with existence of chemical denaturants (urea).
Hydrogen bonds (HBs)	Maintenance of overall stability of the protein structure.
Solvent-accessible-surface-area (SASA)	Examination of unfolding of proteins exposing buried hydrophobic residues to water.
Secondary structure analysis	Description of conformational change by determination of secondary structure elements change, such as, helical content. Most common element is α-helices.
Ramachandran Plot	A phase diagram of two (2) torsion angles φ and ψ showing if residues are in permitted regions of residue conformations.

MOG: 99-107 Isoform 1	BamHI cleavage site	Enterokinase (EK) linker	MBP: 84-104 Isoform 18.5 KD	BamHI cleavage site	C terminal of IL 16 Isoform 1

Peptide Sequence
MFFRDHSYQEGSGDDDDKGNPVVHFFKNIVTPRTPPPSQGGSMPDLNSSTDSAASASAASDVSV
ESTAEATVCTVTLEKMSAGLGFSLEGGKGSLHGDKPLTINRIFKGAASEQSETVQPGDEILQLGTAMQ
GLTRFEAWNIIKALPDGPDGPVTIVIRRKSLQSKETTAAGDS

MOG: 99-107 Isoform 1	BamHI cleavage site	Enterokinase (EK) linker	MBP: 84-104 Isoform 18.5 KD	BamHI cleavage site	C terminal of IL 16 Isoform 1

Peptide Sequence
MFFRDHSYQEGSGDDDDKGYGSLPQKSHGRTQDENPVVHFGSMPDLNSSTDSAASASAASDVSV
ESTAEATVCTVTLEKMSAGLGFSLEGGKGSLHGDKPLTINRIFKGAASEQSETVQPGDEILQLGTAMQ
GLTRFEAWNIIKALPDGPDGPVTIVIRRKSLQSKETTAAGDS

Figure 2.10 Peptide sequence for Construct 1 and Construct 2 reproduced from Banisharif-Dehkordi *et al*. [144] Copyright 2019, with permission from PMC, 10.4103/1735-5362.251849.

Additionally, a case study on COVID-19 by Kumar *et al*. employed MDS to assess the potency and specificity of the vaccine with target receptors [146-148].

Rather than computing RMSD, RMSF or R_g, they used a Normal Node Algorithm (NNA) to show that there were few atomic fluctuations in the complex system (vaccine and virus specific membrane receptor TLR-2) [132]. This finding implied low system deformations [146].

An Eigen score was computed of 3 x 10^{-6}, Figure 2.11A, that confirmed the rigidity of the motion of the complex [146]. A low Eigen score identifies less stability and easy deformation of the atomics coordinates [149]. However, this is not sufficient to draw any conclusions. This is because Kumar *et al*. did not report comparison of Eigen scores with controls. Through

confirming a low deformation index of 0.1 to 1.0 Å, Figure 2.11B, they concluded that the complex was rigid and stable.

It is concluded that epitope immunogenicity prediction using MDS is therefore critical to understand each parameter, therefore, to predict experimental results.

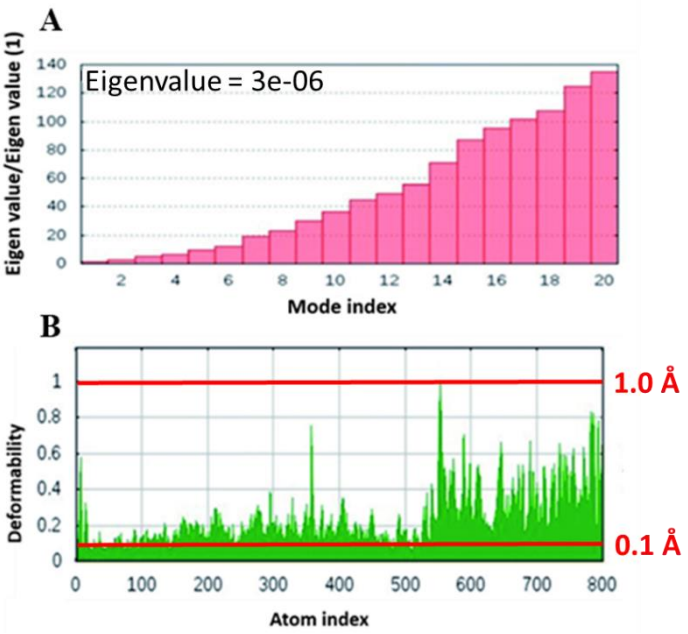


Figure 2.11 MDS analysis of vaccine construct with TLR-2 (A) Eigen value, (B) Deformability reproduced from Kumar et al.32 Copyright 2020, with permission from the Royal Society of Chemistry, 10.1039/D0RA06849G.

2.6.2.3 Computational limitations and experiments

It is not uncommon to observe weak connection, or even no connection, between simulation predictions and later experimental data. This is because of limitations in computing.

These derive from two (2) sources: 1) MDS uses shorter time-scales compared with actual biological processes [150]. This can be obviated at some level with development of parallel algorithms, software and hardware (supercomputer) [150-154], and; 2) selection of empirical force fields limiting approaches to actual biological systems. There are limited selections of force

fields suitable for modelling actual biological system [150,155,156]. As is stated by Flower *et al*., 'There is a tension between what we would wish for and what is available to us' [150]. This statement appears practically true when considering the gap between computational systems and actual biology; biological systems for the present are more complex than computational code. For example, existing methods to predict MHC binding typically rely on large-sets of experimentally-measured binding affinities [157]. The quality of these methods is highly dependent on the available experimental data [158]. In addition, MHC binding data are only available for a limited subset of alleles, whereas there are thousands of different MHC alleles in the human population [157]. Approaches predicting MHC binding are trying to close this gap, but with mixed success [150].

We conclude however that these case studies highlight the significant practical potential of MDS for predicting epitope immunogenicity, compared with direct testing of immunogenicity of a vaccine via experiment. MDS can be reliably used to predict potency and specificity of selected epitopes through addressing complex binding energy and stability.

Despite some limitations, 'MDS predictions should be used before taking action'. This is because development time and costs for new vaccines are reduced, together with boosted success rates for efficacious vaccines.

2.6.3 Prediction of protein structure by MDS

Following careful selection of epitopes with strong immunogenicity, prediction of the structure of epitope-based chimeric vaccine is the MDS following step. Step M2 (Figure 2.7) is considered an essential preparation before step M3, Protein Stability Study. Without an

accurate predicted fusion protein structure, vaccine stability predicted by MDS will not match with experimental data.

2.6.3.1 Prediction methods for protein structure

There are two (2) protein structure prediction methods, namely, 1) template-based modeling (TBM), and; 2) template-free modelling (TFM), as is shown schematically in Figure 2.12.

TBM searches all protein structures determined from prior experiment from the Protein Data Bank (PDB), RCSB [159,160]. Comparative modelling (also called homology modelling) and threading are two (2) usual TBM methods [160,161]. Homology modelling is performed by assuming that two (2) homologous proteins share similar structures. The accuracy of the model is determined by the degree of similarity of the sequences [162-164].

As is seen from Figure 2.12, the black-coloured structure represents existing template with known protein information, and the red-coloured structure represents the predicted part. The red-coloured structure is formed by matching the sequence of black structure and proceeding template alignment. Threading is particularly used for fold recognition. By using certain parts of the target sequence giving high homology, the correct fold can be reliably selected, Figure 2.12 [165,166].

TFM is more computationally challenging compared with TBM. This is because it predicts protein tertiary structure based on primary amino acid sequence without sufficient structural information (Figure 2.12) [119,167,168]. *De novo,* or *ab initio,* is TFM modelling that builds 3D protein models based on first principles [167,169,170]. The more successful are usually based on assembly of known structure fragments with potential energy scoring [119], such as Rosetta[171,172]

and QUARK [173]. Because it requires highly significant computational resources to maintain time-scales, time-length and significant number of interactions for construction of energy function, this predictor is found to work only for small proteins with length up to 120 amino acids [174,175], such as chignolin (10 residues, PDB ID: 1UAO) and Crambin (46 residues; PDB ID: 1CRN) [73.] The design of a fusion protein can be computationally demanding due to greater risks with mis-folding than with single-domain proteins.

Therefore, correct folding, stability and interaction between domains have to be considered [176]. Two (2) cases are presented here to predict fusion protein structure.

Case-1 is applied when the structure of NPC is known. The construction of NPC inserted with epitopes focused on TEM homology modelling of similar protein structures, and introduction of linkers depends on design purposes. For NPCs that are not available in the PDB, Case-2 is applied to predict NPC, followed by prediction of NPC-epitopes structure. Case-2 therefore requires additional steps compared with Case-1.

In the following, there is therefore a selected focus on Case-2 type studies.

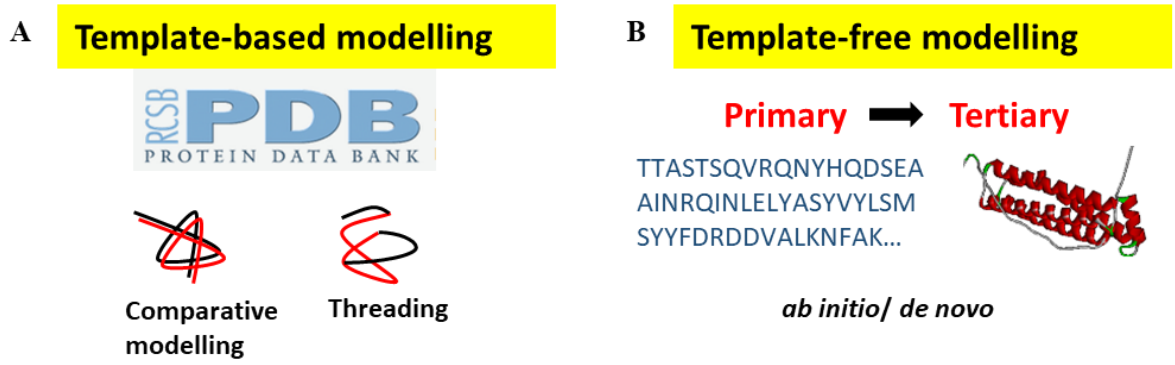


Figure 2.12 Two example methods for prediction of protein structure. (A) Template-based modelling (TBM). For TBM, for comparative modelling, black-colour represents existing structure in Protein Data Bank (PDB), whilst red is predicted structure. For threading, black represents recognised region, and red, predicted region. B) Template-free modelling (TFM). For TFM, protein primary information, amino acid sequence, is the only available information. By performing *ab initio*, or *de novo*, protein tertiary structure (3D) is predicted.

2.6.3.2 Case-2 type protein structure prediction

Shamriz *et al*. confirmed the advantages of MDS to predict initially unknown structures [176]. This study aimed to design and predict structure for a fusion protein consisting of Plasmodium *falciparum* cell-traversal protein (P*f*CelTOS), inserted with human interleukin-2 (IL-2) as adjuvant and M cell-specific peptide ligand (Co1) as antigen. This fusion protein was developed as vaccine for prevention of malaria infection.

Via molecular modelling study of P*f*CelTOS and designed fusion proteins using iterative-threading assembly refinement (I-TASSER), 3D structures were generated. These structures were confirmed in PyMOL by showing majority of residues located on the permitted region of Ramachandran Plot. MDS at usual condition (300 K) were then carried out for analysis of energy (total, kinetic and potential) and structure stability (RMSD, R_g and H bond formation/deformation).

Designed fusion protein structures before and following MDS were critically compared. The desirable energy trend is for constant total energy, constant or decreasing kinetic energy,

together with constant or increasing potential energy. An increasing kinetic energy level reflects breakdown of folded protein structures [176,177], whilst decreasing potential energy instability of designed protein structures [176]. The authors reported stable designed structures based on low values of RMSD and R_g. A constant number for H bonds confirmed the accuracy of designed fusion protein structures.

Sometimes, there exists limited structural information for NP. A combined study is then applied to predict protein structure. For example, Mobini *et al.* designed a VLP-based vaccine using hepatitis B core antigen (HBcAg) as the carrier, with inserted Myrcludex and hepatitis B surface antigen (HbsAg) [178]. Homologous structures for HBcAg and HBsAg were searched for prediction of protein tertiary structures. It was found that HBcAg was 94 % identical to the template (PDB ID: 1QGT) for domain of the protein in the N-terminal (residues 1-140). The template for the nanopeptides (residues 141-149) and C-terminal domain (150-183) was unknown. A combined homology modelling and *ab initio* was applied to predict structure for HBcAg proteins.

A total of 10^4 models were scored based on the discrete optimized potential energy (DOPE) and the top 10 with the lowest DOPE score were readied for conformational analysis. For HBsAg, *ab initio* was performed. The quality of the final model was evaluated by Z-score in comparison with experimentally validated structure of proteins. The conformational behaviour of the designed vaccine was validated via a low value RMSD and R_g.

2.6.3.3 Current challenges and innovative MDS

In practice, using MDS to predict large protein structure can be a practical challenge. At present *de novo* predictors are difficult for large proteins [167,179,180].

A number of *de novo* approaches rely on assembling proteins from short peptide fragments. These are suitable for small proteins containing less than 1000 amino acids [181,182]. A certain time-length is required to perform umbrella-sampling and proceed to protein folding [183]. Current computer resources are limited for microsecond-based simulations [184]. However, protein folding can take microseconds to tens-of-minutes to complete.

Therefore some slow-folding protein simulations are likely to fail in protein correct folding [185]. However, innovations in MDS are expected to improve accuracy and performance on protein structure prediction.

Cheung *et al*. applied a method titled *NiDelta* with ultra-fast MDS to predict 18 large proteins, with more than 100 residues from sequence to tertiary structures [167]. This is shown schematically as Figure 2.13. It is seen in the figure that three (3) steps were involved in this method, Data Processing (Figure 2.13A), Prediction (Figure 2.13B), and Simulation (Figure 2.13C).

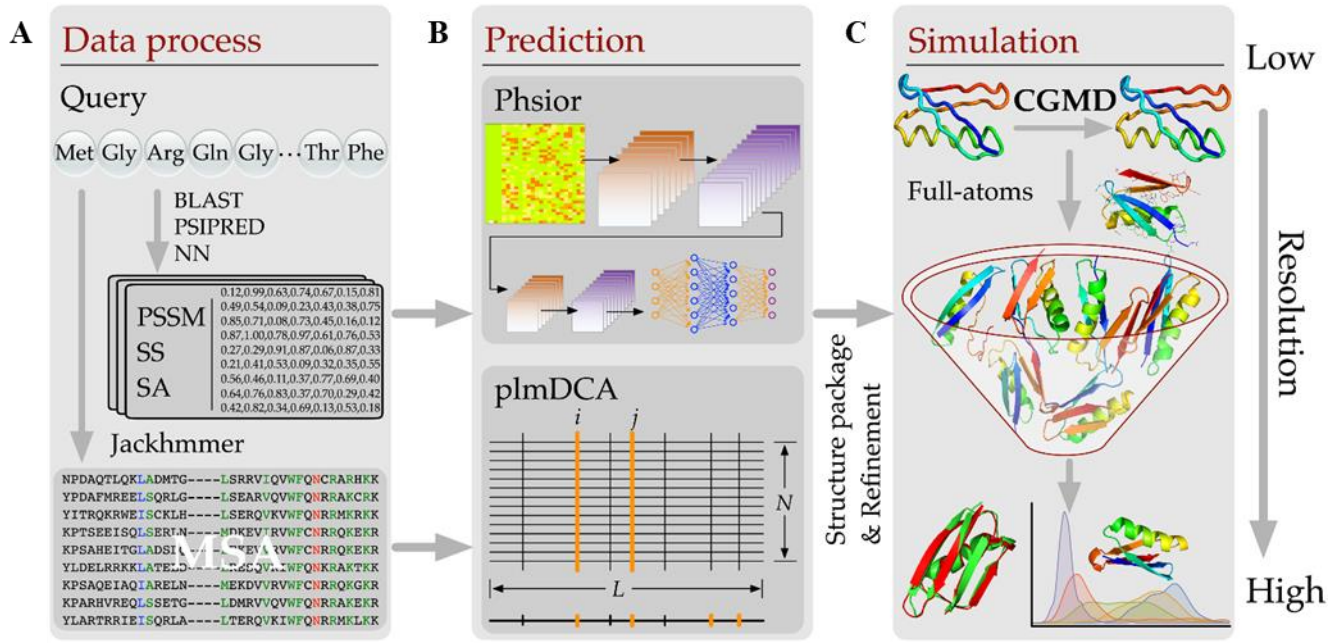


Figure 2.13 Schematic for innovative approach to predict protein tertiary structure following three (3) steps. (A) Step 1, Data Process (sic). MSA stands for multiple sequence alignment. (B) Step 2, Prediction using Phsior and plmDCA. (C) Step 3, Simulation. By performing CGMD, protein tertiary structure with highest resolution is refined and filtered. This figure is reprinted from Cheung *et al.* [167] Copyright 2018, with permission from PlOS ONE, 10.1371/journal.pone.0205819.g001.

NiDelta has two (2) restraints, namely, 1) predicted torsion angle, and; 2) residue-contacts for performing a coarse-grained molecular dynamics (CGMD)-*Upside* [167]. There are two (2) stages involved to process a given target sequence: Stage-1 is to build *Phsior* by constructing a 'non-redundant sequence' set of data from the PDB library and filter it through PISCES [186], and; Stage-2 is to predict residue-contacts via sequence alignments. Additional detail was provided [167].

An ultra-fast MDS simulation was accomplished with *Upside* in which each model was displayed by a reduced chain representation, but with sufficient structural information, such as, backbone, Cα and C atoms. For each of 18 large proteins, 500 *Upside* simulations were carried out beginning with an unfolded structure. Each was folded in several CPU h. Nine (9)

representative residue contacts were compared with native ones by computing Cα-RMSD and TM-score based on experimental reference structures.

It was found that many protein structures were predicted sufficiently, despite 'noises' from incorrect predictions caused by false positive inferences (Figure 2.14 below is an example of noises). For example, protein C-H-RAS P21 (PDB ID: 5P21) was sufficiently predicted with lowest Cα-RMSD of 3 Å and the same fold with TM-score of 0.76.

These findings gave confidence for practical potential of this innovative approach of *Upside* to predict large protein structures.

This study highlighted however a number of important limitations. These included: 1) Residue-residue contacts were not able to be predicted without sufficient multiple sequences in an alignment of a protein family; 2) Despite sufficient sequences, there were failed predictions; false positive prediction in the pairwise contacts caused incorrect building of protein 3D structure, and; 3) It was difficult to accurately predict torsion angles.

It is concluded that overall however this study can be reasonably regarded as a successful demonstration of innovative MDS to predict large protein structures.

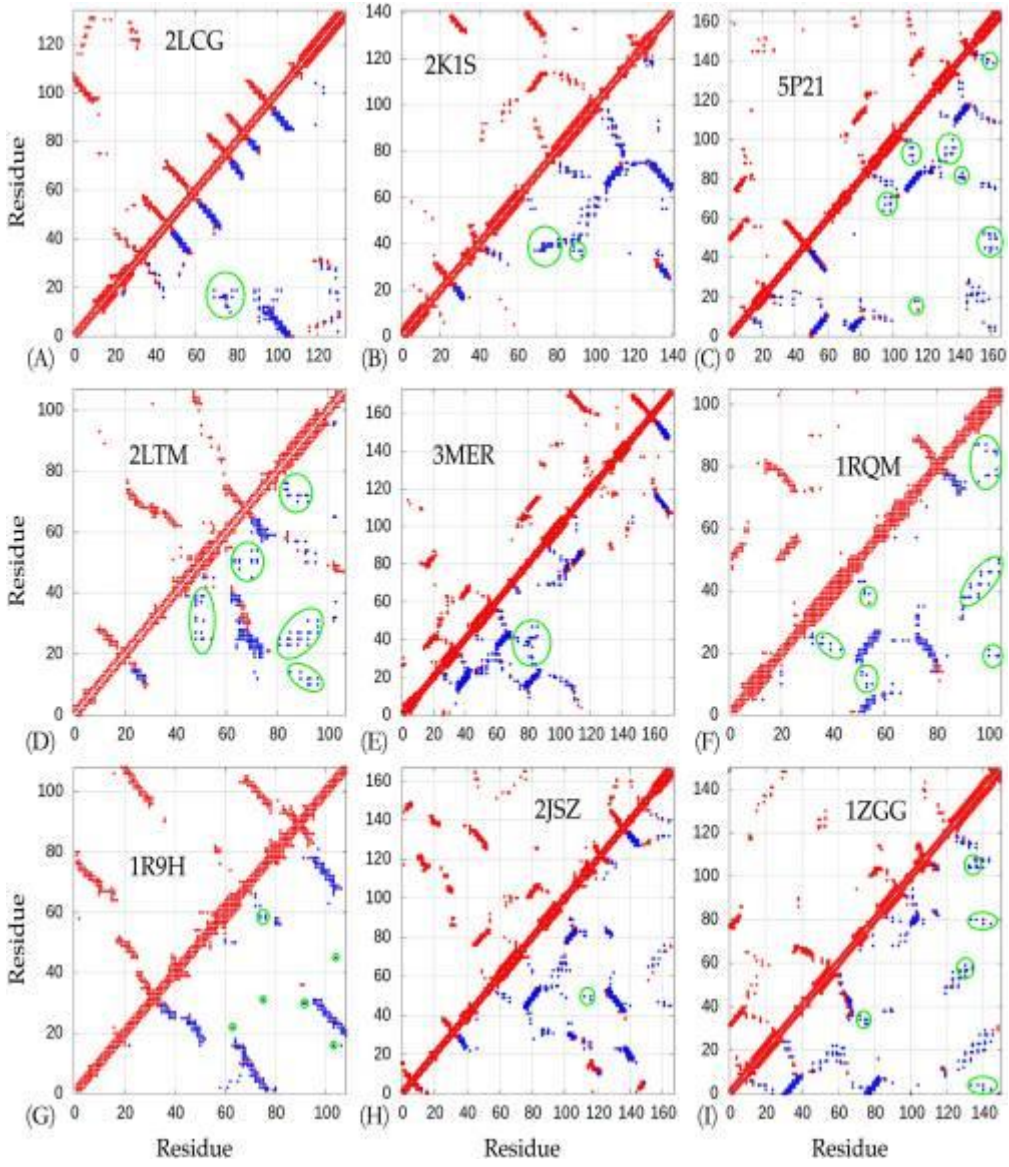


Figure 2.14 Examples of noises from false positive inferences. Noises are denoted as blue-colour dots in green circle. This figure is reprinted from Cheung *et al*. [167]. Copyright 2018, with permission from PlOS ONE, 10.1371/journal.pone.0205819.

Recent innovative MDS considers implications of suitable force fields on studying protein folding or mis-folding. There is a special group of NPs called repeat proteins. The simulation for repeat proteins folding is practically difficult to achieve. Zheng *et al*. applied a coarse-grained protein force field, AWSEM (Associative Memory, Water-mediated, Structure and Energy Model), into MDS to investigate the folding of fused dimers of SH3 domains and Ig domains from human titin [187]. They reported that the frequency of mis-folding was reduced by

decreasing the sequence identity between monomers in the fused dimer. This is supported by the hypothesis that evolution favoured low sequence identity between neighbouring domains in repeat proteins [188].

This finding is practically useful for future epitope-based chimeric vaccine design. For repeating proteins, by altering epitope sequence or length, each monomer can have unequal sequence identity. This is desirable for protein folding and readier downstream processing development.

These studies have shown that overall, MDS has been successfully applied to predict protein structure. A detailed picture of protein folding, which is practically difficult to ascertain from experiment, is gained through simulations. However, contrary to small protein structure prediction, there are additional difficulties involved for large protein application of MDS.

There is a number of recently released, advanced software that apply deep machine learning artificial intelligence (AI) to predict protein structure, with reported high accuracy. DeepMind™ released an open-source version of a deep-learning neural network, AlphaFold2[189]. This was developed for prediction of accurate protein structure when there are no homologous structures available. The key input is protein primary amino acid sequence, and aligned sequences of homologues (multiple sequence alignments and pairwise features). These are processed by the trunk of the network called Evoformer. The detailed procedure is reported by Jumper *et al*. [189]. AlphaFold has demonstrated a near-experimental accuracy in predicting protein 3D structures. For example, accuracy was comparable with other computing methods with a median backbone accuracy confirmed 0.96 Å RMSD for Cα (compared with another 'best method' of RMSD for Cα of 2.8 Å).

Another method called RoseTTA fold was developed by Baek *et al.*, and can obtain accurate models of protein-protein complexes [190]. The authors applied to a three-track network to process sequence, distance and coordinate information concurrently to rapidly generate protein structures which is similar to that from DeepMind. This software advantageously speeds up generation of accurate models for protein structures in MDS.

2.6.4 Prediction of protein stability by MDS

A desirable protein stability is mandatory in epitope-based chimeric vaccine development. Stable proteins ensure the functionality and integrity of vaccines, and avert negative reactions due to aggregates.

Unstable proteins result in protein degradation with loss of function, or aggregation. This reduces biological activity [191]. MDS can be used to predict protein dynamics, structures, and functions at an atomistic level [192-194].

2.6.4.1 Parameters for computation

Parameters commonly computed for determination of protein stability are summarised in Table 2.3. It is required to make connections between these and experimental data. RMSD is widely used to reveal protein structural stability via determining conformational stability changes in proteins [195,196]. RMSD varies significantly in the initial timeframes in simulation, and reaches steady-state when the simulation system is stable. Comparing RMSD values at steady-state is important so as to present an accurate estimation of protein conformation stability. The greater the RMSD, the less stable the protein structure is in MDS [191,194,197-200].

R_g is regarded as the fundamental measurement of protein overall size. This serves to show protein structure change in MDS [195,201]. R_g depends on both original protein size, and protein stability change in time. A low R_g value indicates a compacted protein structure that results in a greater stability [198,199,202,203].

RMSF is a basic parameter commonly computed to specify the flexibility of individual residues. It is associated with RMSD and R_g. The greater RMSF value, the more flexible movement [200,204,205]. RMSF is related to protein stability. Low average RMSF values indicate individual amino acid residues that help preserve protein stability during MDS [206]. High RMSF values confirm flexible regions on proteins. Therefore, RMSF is a practically useful parameter for variant designs with stabilized flexible regions [207,208].

The general use of these parameters appears to work well for natural conditions. However, to ensure a detailed study of protein stability at molecular level, it is necessary to run MDS under conditions including, elevated temperature, range of pH, and/or chemical denaturants. The

reason for doing this is to understand the mechanism by which proteins unfold from native structure.

The fraction of exhibited native contacts, Q, indicates the stability of proteins under these more-extreme conditions, such as hyperthermal, or in the presence of chemical denaturants.[194] Less stable proteins normally unfold rapidly, resulting in a reduction of native contacts compared with stable proteins over the same simulation time.[194,209-211] Hydrophobic interactions between each residue of the protein and water are quantified by native contacts and hydrogen bond (HB). HB contributes favourably to protein stability [212,213], especially, maintaining secondary structure (α-helices).[202] The fewer number of HBs in proteins leads to relatively low stabilities [194,200].

Another parameter closely related to hydrophobic interactions is solvent accessible surface area (SASA). SASA shows the changes in accessibility of protein to solvent [194,200]. It is related to protein stability. Lower SASA generally represents higher thermodynamic stability of protein [170,214-216]. Unfolding of proteins caused by denaturation results in the hydrophobic core being exposed to solution, leading to increasing SASA.

Zhang *et al*. reported that SASA for apoMb, and its variants E109G and E109A, increased in the presence of urea as unfolding proceeded [194]. E109G exhibited a steeper increase in SASA than wild-type apoMb, whilst E109A showed a reduced increase. Based on this finding, it was concluded that E109A improved the stability of wild-type apoMb and E109G had a destabilizing impact.

The connection between SASA and protein stability however may not be significantly obvious without more-extreme conditions. For example, Kumar *et al*. deduced a contrary conclusion [200]. They assessed SASA for wild-type and variant PMP22 structures. They illustrated a greater SASA for the variant protein structure than for the wild-type protein, making the variant structure more stable [200,217]. However based on SASAs, the difference between wild-type and variant was not significant. It is difficult therefore to draw an estimate of protein stability from these SASA data.

Muneeswaran *et al*. computed both RMSD and SASA for wild-type cytochrome (Cyt) c and its variants (Y67E, K72W) [218]. Low RMSD values for both backbone and Cα showed wild-type Cyt c was the more stable compared to its variants. However, the value for SASA was greatest amongst these three. It is not possible to confidently link protein stability with computed SASA from native MDS. This however can be more confidently made for more-extreme conditions.

MDS analysis of secondary structure is gaining momentum in studying protein stability. To address why it is necessary to analyse protein secondary structures, it is important to understand the limitation(s) of MDS.

For large protein molecules there are computational time limitations. Experiments, including heating of proteins until denaturation, requires a significant time (min). This is not ideal given the maximum simulation time is μs. Therefore, changes of protein tertiary structure caused by denaturants without sufficient treatment time are not obvious in MDS, whilst protein secondary structure changes are more visible in given MDS-time length. By analysing key secondary

structure elements (e.g. α helix, β sheet, coil, bend and turn) [120,168,199,219,220], protein stability can be estimated.

For example, Amir *et al*. analysed the changes in secondary structure in wild-type STN1 protein and its variants R13T and D157. It was found that D157Y variant resulted in a significant decrease in secondary structure in comparison with R135T, where there was no significant change.[120] This finding evidenced that D157 mutation impacted conformational stability of STN1. Another study by Schaller *et al*. focused on change in α-helical content in analysing thermal stability of biosurfactant DAMP4 [117]. By comparing initial helix content (%) and end-point helix content (%) of all variants, loss of secondary structure was identified and stability was assessed. A significant loss of helix content underscored an instability of protein.

The Ramachandran plot is closely related to protein secondary structure analyses. Ramachandran plot is a phase diagram, made by two (2) torsion angles φ and ψ for each amino acid residue. It is used to identify whether the conformation of amino acids are in permitted regions [202,221-223]. An example of a Ramachandran plot is presented as Figure 2.15. Most residues marked as black-colour in the figure, were concentrated in the preferred regions (marked in red-colour) [202]. This finding evidenced a stable secondary structure conformation.

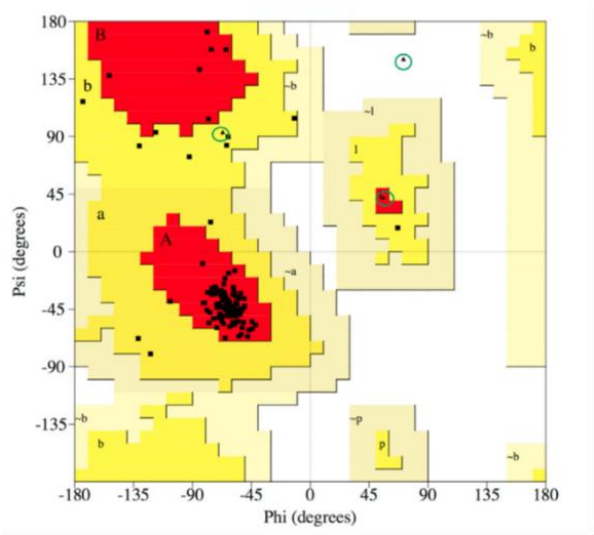


Figure 2.15 Example Ramachandran plot for Rop protein. Red-colour represents most preferred regions; transition of yellow from dark to light represents most to least favourable conformations. This figure is reprinted from Arnittali *et al*. [202]. Copyright 2019, with permission from Elsevier, 10.1016/j.procs.2019.08.181.

2.6.4.2 MDS steps for protein stability

Present application of MDS for determining protein stability involves the three (3) steps: 1) simulation for native/non-native environments, 2) designs via analysing dynamics of unstable residues, and; 3) simulation of designs to determine impact of mutations to improve protein stability. Figure 2.16 shows schematically the connections between these steps.

The focus in the following is on simulation under more-extreme conditions, including elevated temperature, varying pH and the presence of chemical denaturants.

Modelling of protein structure at elevated temperature is used to determine thermal stability. For example, Manjunath *et al*. simulated five (5) functional SAICAR synthetase ligates from different organisms including, mesophilic, thermophilic and hyperthermophilic, at three (3) temperature values of 300, 333 and 363 K [198]. To establish stability differences, they determined RMSD, RMSF, R_g, SASA, HB and salt bridges (SBs). SBs are defined as ion-ion interactions between charged residues, which are important to maintain conformational stability of proteins [224,225].

It was found that mesophilic structures appeared unstable at 363 K. They showed significantly increased RMSD, fluctuating RMSF and R_g at 363 K compared with other groups. There was a gradual collapse of the structure and instability at 363 K for the mesophilic group. Additionally, the hydrophobic SASA more rapidly increased relative to the mesophiles, compared with hyper-thermophiles. In mesophiles, a comparatively higher number of long-lived (with higher percentage existence time) contacts are lost due to increased protein flexibility.

A comparatively greater number of hydrogen bonds are found in the mesophilic proteins at 363 K, compared with that at 300 K, whilst other protein groups remained unchanged. The number of SBs in the mesophilic proteins decreased in comparison with the other groups at 363 K.

These findings led these authors to conclude that mesophilic proteins were not stable at 363 K when compared with other groups.

It is important to note that this research was based on only simulations. Findings from each parameter were supporting each other however. This confirmed the accuracy of the methodology dependent on MDS. This study therefore highlighted the advantages of MDS in investigating protein stability. Experiments are not necessary therefore to be conducted when sufficient self-reinforcing information has been generated from MDS.

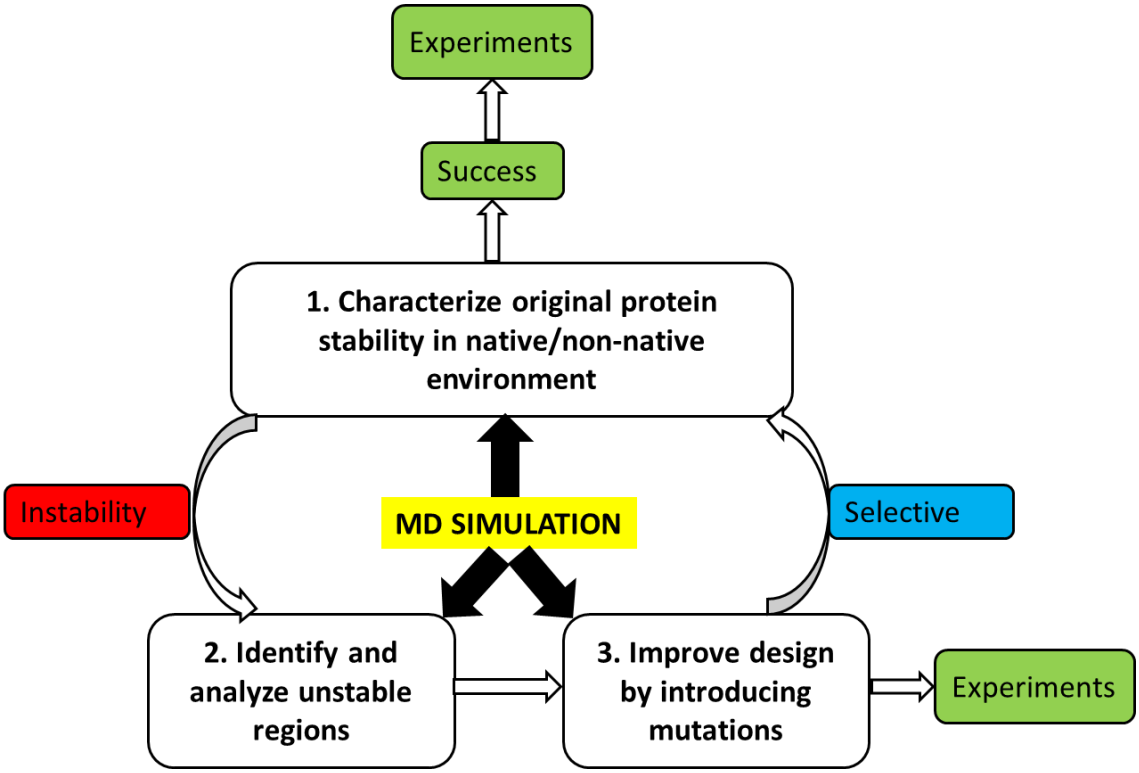


Figure 2.16 MSD flowchart for protein stability.

There is also potential to study stability using both simulation and experiment. This combined technique sometimes is advantageous because it involves validation of simulation.

A case study reported by Bekker *et al*. for example predicted relative thermal stability of single-domain antibodies (sdAbs) using high temperature MDS and validated the results with melting temperature (T_m) obtained from experiments [118]. Melting temperature measures the temperature where half of a protein is folded and the other half is un-folded.

High-temperature MDS at temperature representing 400 and 500 K were performed using 300 K as a control. The methodology was that high temperature MDS which gave valuable information on protein design. However, they reported that RMSD did not correspond well with T_m. There was no significant difference in RMSD between antibodies at the different temperatures. The fraction of native atomic contacts (Q) showed closer estimation to experiment. This Q value determines the fluctuation of the protein corresponding to the atomistic interactions.

It was reported that MDS at 400 K combined with Q gave an accurate measure of thermal stability of sdAbs. This was because there was a good correlation between Q value at 400 K and experimental T_m.

Importantly, this study highlighted a challenge with MDS. This is that the connection between simulation and experimental data needs to be carefully interpreted.

Buffer pH is important to protein stability. Thermal stability is closely related to varied pH of solution. Residue charges are determined based on the solution pH. These are involved in both

intra- and inter-molecular noncovalent interaction. This results in increasing, or decreasing protein stability.

For example, Sviben *et al*. reported that thermal stability of ovalbumin was dependent on pH. This was confirmed both in MDS and with experiment [191]. Ten (10) simulations were conducted for two (2) pH conditions (acidic and alkaline) at five (5) temperatures of 37, 42, 57, 74 and 85 °C. It was found that for both acidic and alkaline environment, at temperatures of 37 and 42 °C more stable systems are achieved than at higher temperatures of 74 and 85 °C.

Additionally, instability in the acidic system was more pronounced than that for the alkaline. These findings were consistent with findings from experiment. The key difference between these two (2) systems was the interaction of arginine 50 and arginine 58, with surrounding residues. The alkaline system was found to have stronger and more persistent interactions for both Arg50 and Arg58 compared with the acidic system [191].

In addition to simulation for different temperature and solution pH, the solvent environment with chemical denaturants can be used to provide more information on differences in protein stability.

Zhang *et al*. applied MDS to investigate the stability of four (4) apomyoglobin (apoMb) variants, namely, wild-type, E109A, E109G and G65A/G73A, in explicit urea solution.[194] The 2 M urea solution at pH 4.2 acted as a denaturant to drive unfolding of the apoMb variants. Through analyses of RMSD, native contacts and SASA, E109A showed greatest stability, and E109G showed least stability in urea solution.

It was found that the greatest change in conformation was related to destabilization of E109G. In the presence of denaturants, less stable proteins unfold more rapidly than others. This results in significant loss of native contacts against simulation time. Based on RMSD, E109G exhibited greatest drop in native contacts, whilst E109A preserved greatest percentage of native contacts. This led to the most stable structure [194].

These findings were highly consistent with experimental results [226]. This study highlighted therefore that MDS can be used to achieve experiment-matched results for proteins in solvent with denaturant present. This application therefore saved significant time and costs involved with experiment.

With desirable stability results from application of MDS, proteins are ready to be tested in experiment to validate stability and, to check vaccine immunogenicity. However, this model situation does not always occur.

The advantage of MDS compared with experiments is that it computes protein stability, and provides molecular level information to explain any resulting stability difference.

The second-related step in using MDS is to find unstable regions of proteins, Figure 2.16. However, the second- and the third-related steps using MDS for protein stability are normally bound together. This is shown in Figure 2.16. Following identification and analyses of flexible regions using MDS, further mutations are investigated with the aim to improve protein stability. Recent studies focus on improving protein thermal stability.

There are a number of methods to improve protein thermal stability, including: introducing disulphide bonds [197,198,227,228]; increasing the number of salt bridges [197,229-232]; optimizing electrostatic interactions [197,228,233,234]; compacting hydrophobic core [197,198,235-237]; enhancing hydrogen bonding [238-242], and; stabilizing α-helices [243]. However it is not necessary to determine all related properties in a given stable protein.

Combination of one or more of these can significantly impact protein thermal stability [198,238]. Relying on experiment to explore possible combinations is practically time consuming to complete, with at most, likely only mixed success. MDS can therefore be used to speed development via judicious simulation.

Pikkemaat *et al*. used haloalkane dehalogenase (DhlA) as a model to explore potential of MDS to identify flexible regions in proteins to boost thermal stability [199]. These authors focused on RMSD of the molecule relative to the X-ray structure. They also determined the effect of thermal denaturants on secondary structure of the molecule. They reported that most of the mean square displacement was concentrated in the region between residues 180 and 210. The secondary structure elements analysis highlighted that this region was highly stable, and that the significant RMSD was not caused by unfolding of the structure. Therefore, they concluded that relatively large motions occurred in the cap domain of DhlA.

It was concluded from the MDS, that DhlA had high mobility in helix-loop-helix region involving residues 185-211.

The authors reported introducing a disulphide cross-link between residue 201 of this flexible region and residue 16 of the main domain. The mutated enzyme exhibited increasing transition temperature from 47.5 to 52.5 °C. Urea denaturation exhibited a similar trend to experiment.

It is concluded that this work demonstrates the advantageous use of MDS to identify mobile protein domains to boost stability.

Gill *et al*. conducted all-atom explicit-solvent MDS for human procarboxypeptidase A2 (AYEwt) with three (3) designed variants, namely, AYEdes, AYEwt-4mut and AYEwt-5mut, at both 25 and 100 °C [170]. MDS findings highlighted potential stabilizing factors at atomic-level in a set of engineered proteins.

Through analyses of RMSD, RMSF and SASA, protein stability was determined and compared. It was predicted that Phe18 and Ala23 from AYEdes would stabilize α1 of AYEwt-3mut. The authors therefore further designed AYEwt-F58M and AYEdes-M58F with the aim to validate this prediction.

They observed that α1 of AYEwt-3mut was stabilized and that phenylalanine had a minor but meaningful stabilizing impact over methionine at position 58.

Importantly, there are a number of examples using MDS to show that mutated structures decrease overall thermal stability.

Meharenna *et al*. investigated the regions of Cytochromes P450 related to thermal stability via high temperature MDS [244]. CYP119 was the most stable P450 compared with P450cam and

P450cin [244]. By comparing trajectories at 500 K, the Cys ligand loop was hypothesized as an important region for improving thermal stability. This region in CYP119 was stabilized by rigid, nonpolar interactions between Tyr26 and Leu308. The authors generated Y26A/L308A CYP119 by replacing these two (2) amino acids in P450cam with, respectively, Gly and Thr. However, the resulting stability was less than that for the wild-type CYP119. The Cys ligand loop unfolded similarly to P450cam.

The melting temperature from experiment exhibited a similar trend for mutated P450cam at 16 °C less than for wild-type CYP119. MDS therefore predicted the Cys ligand loop as a 'hot spot' critically impacting stability.

It is possible to use MDS to predict which variants boost stability, however, it is practically challenging to mimic exact interactions amongst alternated residues. To model similar interaction between Tyr26 and Leu397 in CYP119, larger nonpolar residues are hypothesized as being replaced in P450cam.

Meharenna *et al*. introduced a disulphide bond, however, this variant failed to improve stability. One reason is that disulphide bonds are unlikely to mimic non-bonded intramolecular interactions, and unfavourable packing problems likely occur. In a similar investigation. Bae *et al*. reported that introducing a salt bridge improved stability of mesophilic protein.[245] It is possible therefore to mimic non-bonded intramolecular interactions via introducing salt bridges.

MDS has been used to investigate protein stability and to predict likely vaccine efficacy. It is advantageous to: 1) characterize fusion protein stability as impacted by thermal-, pH- and

chemical denaturant factors, 2) give practical feedback to identify flexible regions, and; 3) investigate the likely impact of mutations designed to improve stability. It is concluded that MDS simulations significantly improve vaccine success rates by filtering and/or building highly stable epitope-based chimeric vaccines.

2.7 Characterization of EBC vaccines

To justify if the designed vaccine candidate is safe and efficacious, physical characterizations and functional analysis by experiments are required. Table 2.4 summarizes some examples of techniques that are commonly applied to analyse EBC vaccines.

Basic physical characterizations aim to generate structural information of nanoparticles, such as molecular weight and size. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates protein subunits by molecular weight, which can also indicate the purity level of eluate pool [246-248]. Size exclusion chromatography (SEC) and dynamic light scattering (DLS) are techniques to analyse molecular weight and particle size of samples [249,250]. Aggregates level can also be determined by a separate peak from the main peak of single-particle proteins. This is extremely important to consider for EBC vaccines, because aggregates can induce immune responses and affect the efficacy of vaccines [251]. Protein stability is also important to analyse in addition to basic physical characterizations. Circular dichroism (CD)[252] and fluorescence spectrometer [253] are other two techniques to analyse protein secondary and tertiary structural changes against thermal or chemical denaturants. Differential scanning calorimetry (DSC) is another efficient technique to determine the denaturing temperature of proteins (T_m). DSC characterizes protein thermal stability, conformational change and domain folding integrity via determination of heat capacity as a function of temperature [59,254]. A decrease in T_m underscores a destabilizing impact on protein structure. Proteins with greater T_m have stronger thermal stability.

After confirming protein structure and stability, functional analyses of EBC vaccine aim to ensure they are efficacious. Impurities such as host cell proteins and endotoxins are expected to be quantified as a part of functional analyses prior to immunology assay for safety concern[255-

[257]. Shown by Table 2.4, other assays such as antigen-specific antibody titers, T-cell proliferation, memory lymphocyte activation and virus neutralisation are commonly used as preliminary immunology-studies of vaccines [258-261]. These assays can be used as a proof-of-concept to suggest if designed vaccines are able to induce potent humoral and cell-mediated immune responses before conducting clinical studies.

Table 2.4 Examples of techniques to analyse EBC vaccines

Technique	Role	References number for studies to analyse vaccines
Physical characterization		
Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)	To separate protein subunits by mass in order to observe purity level	246-248
Western blot	To detect specific proteins in a sample	258
Dynamic light scattering (DLS)	To characterize protein particle size	262
Size exclusion chromatography (SEC)	To characterize levels of aggregates and single particles by separating them by molecular weight	249,250
Transmission electron microscopy (TEM)	To capture images of protein particles	263,264
Differential scanning calorimetry (DSC)	To determine the denaturing temperature of proteins	265
Circular dichroism (CD)	To determine protein secondary structure changes	252
Fluorescence spectrometry	To determine protein tertiary structure changes	253
Functional analysis		
Host cell protein assay	To determine level of host cell proteins in vaccines	255,256
Endotoxin assay	To determine endotoxin level in vaccines	257
Antigen-specific antibodies titers	To determine humoral immune responses induced by vaccines	259

T-cell proliferation assay	To compare the strength of humoral with cell-mediated immune responses induced by vaccines	[260]
Memory lymphocyte T cells differentiation	To test the ability of vaccines to confer long-term protection	[261]
Virus neutralisation	To determine the ability to neutralise viruses	[258]

2.8 Future direction: messenger RNA (mRNA) vaccines

Messenger RNA (mRNA) vaccines are treated as the third-generation nucleic acid vaccine on the basis of first-generation traditional vaccines and second-generation subunit vaccines [266]. Recent COVID pandemic speeds up the progress of mRNA-based vaccines. Several mRNA-based COVID vaccines (Moderna, Pfizer/BioNTech) are currently supplied in Australia.

The mRNA consists of a 5' cap, 5' UTRs, an open reading frame (ORF), 3' UTRs and a poly (A) tail. The mRNA vaccine is produced by inserting the encoded antigen in a DNA template from where the mRNA can be transcribed *in vitro* [267]. The mRNA vaccine works by reaching the cytosol of cells and induces target proteins that can trigger immune responses. mRNA vaccines display a number of advantages compared to traditional vaccines. mRNA vaccines: 1) are considerably safe without involving infectious elements or risks of stable integration to host cell genome and therefore well-tolerated by healthy individuals [268]; 2) induce potent efficacy and immunogenicity [269,270]; and 3) are cost-effective and can be produced considerably rapidly [271]. The generation of mRNA vaccine at clinical scale can be completed within weeks [268].

There are currently two pathways for the delivery of mRNA vaccines. The first pathway is loading mRNA into DCs *ex vivo*, followed by re-infusion of transfected cells [267,272]. This pathway is able to control the cellular target and transfection efficiency. However, it is expensive and time-consuming process. The second pathway is direct injection of mRNA with/without a carrier, which is significantly cost-effective and can be produced fast. However, it is challenging to precise control the efficiency of cell-specific delivery. In addition, naked mRNA is easily degradable, which is difficult for clinical applications [266].

Recent progress of mRNA vaccines focuses on improving stability by a rational molecular

design. The stability of mRNA is closely related to the 5' cap [266]. By specific design of 5' cap, the stability and translation efficiency of mRNA can be improved [273]. In addition, lipid nanoparticles have been successfully introduced to delivery mRNA in clinic [274]. Numerous studies have proven that rapid mixing of lipid nanoparticles and mRNA shows a stable nanostructure [275-277]. For instance, the mRNA can be captured inside the core of lipid nanoparticles via electrostatic interactions with lipids, which effectively protects mRNA from nuclease degradation and enhances the stability of nanoparticles [275,278]. The delivery efficacy of mRNA can also be achieved by modifications of lipid. For example, by adjusting head groups and hydrophobic tails of lipid nanoparticles, cellular uptake of mRNA can be enhanced [274,279,280]. Furthermore, by adjusting the proportions of lipid components, lipid nanoparticle-mRNA can achieve organ selectivity [281,282]. Overall, mRNA vaccines can be another solution to combat infectious diseases.

2.9 Chapter summary

Based on the review of relevant literature in this chapter, key findings are summarised:

1. Human ferritin heavy chain (HFn), as an example of NPC, comprises 24 identical subunits, forming a spherical hollow cage structure of 12 nm in diameter. Every HFn subunit is made up of five (5) helices, A to E.
2. HFn is beneficial to be developed as EBC vaccine because it exhibits: 1) display of antigens and exposure of immunogens in a repetitive and well-organised manner; 2) robust thermal and chemical stability; 3) high safety with low toxicity, great biocompatibility and biodegradability; and 4) strong binding to T cell immunoglobulin and TIM-2 to regulate cellular immunity.
3. There are three (3) promising insertion sites on HFn, N-terminus, C-terminus and flexible loop region. There are controversial claims about comparisons amongst these insertions sites about corresponding protein stability and vaccine immunogenicity.
4. Helix E in HFn plays a significant role in maintaining HFn stability. However, there are lack of studies investigating the potential mechanism on how helix E affecting protein stability.
5. VLP-based EBC vaccine induces strong immunogenicity because of highly mimic structures to virus, however, it introduces four (4) main limitations: 1) difficulty of expression by cost-effective expression systems (*E. coli*); 2) instable protein structures after removal of virus genomes; 3) easy aggregation; and 4) difficulty of removal of impurities (HCPs, DNA and lipid).
6. Flexible linker length is critical to be designed to boost protein stability, however, there is lack of study on investigating this topic.
7. MDS is an emerging alternative for EBC vaccine and has significant advantages. These include prediction of: 1) performance of EBC vaccines 2) fusion protein structure via

considering protein folding and resulting structure stability, and; 3) protein stability. MDS judiciously used, can filter and establish the most stable vaccine for later validation via experiment with reduced time and cost.

To address these current research gaps, in this PhD Thesis, HFn is developed as a vaccine carrier (NPC) by critically designing insertion sites and flexible linker through MDS and experiments. First, MDS is performed to molecularly design HFn-based EBC vaccine by predicting protein stability at two (2) different insertion sites (N- and C-terminus), flexible linker length (3- and 15-residues), and variant study to investigate key residues affecting protein stability. Second, HFn-based vaccine is purified, produced and stability-characterized by experiments to validate simulation results and to prepare for last step. Immunology study is conducted in the last step to compare the immunogenicity of N- and C-insertion site, and to compare the immunogenicity of HFn and HBc with same model epitope and same insertion site.

2.10 References

1 Shin, M. D. *et al.* COVID-19 vaccine development and a potential nanomaterial path forward. *Nature Nanotechnology* **15**, 646-655, doi:10.1038/s41565-020-0737-y (2020).

2 Mak, T. W. & Saunders, M. E. in *The Immune Response* (eds Tak W. Mak & Mary E. Saunders) 695-749 (Academic Press, 2006).

3 Sahdev, P. *et al.* Biomaterials for nanoparticle vaccine delivery systems. *Pharmaceutical Research* **31**, 2563-2582, doi:10.1007/s11095-014-1419-y (2014).

4 Bråve, A. *et al.* Vaccine delivery methods using viral vectors. *Molecular Pharmaceutics* **4**, 18-32, doi:10.1021/mp060098+ (2007).

5 Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature* **455**, 613-619, doi:10.1038/nature07352 (2008).

6 Burrell, C. J. *et al.* in *Fenner and White's Medical Virology (Fifth Edition)* (eds Christopher J. Burrell, Colin R. Howard, & Frederick A. Murphy) 155-167 (Academic Press, 2017).

7 Al-Halifa, S. *et al.* Nanoparticle-based vaccines against respiratory viruses. *Frontiers in Immunology* **10**, doi:10.3389/fimmu.2019.00022 (2019).

8 Bernasconi, V. *et al.* Mucosal vaccine development based on liposome technology. *Journal of Immunology Research* **2016**, 1-16, doi:10.1155/2016/5482087 (2016).

9 Doll, T. A. P. F. *et al.* Nanoscale assemblies and their biomedical applications. *Journal of The Royal Society Interface* **10**, doi:10.1098/rsif.2012.0740 (2013).

10 Schwendener, R. A. Liposomes as vaccine delivery systems: A review of the recent advances. *Therapeutic Advances in Vaccines* **2**, 159-182, doi:10.1177/2051013614541440 (2014).

11 Henriksen-Lacey, M. *et al.* The vesicle size of DDA:TDB liposomal adjuvants plays a role in the cell-mediated immune response but has no significant effect on antibody production. *Journal of Controlled Release* **154**, 131-137, doi:https://doi.org/10.1016/j.jconrel.2011.05.019 (2011).

12 Sharma, A. & Sharma, U. S. Liposomes in drug delivery: Progress and limitations. *International Journal of Pharmaceutics* **154**, 123-140, doi:https://doi.org/10.1016/S0378-5173(97)00135-X (1997).

13 Bose, R. J. *et al.* Biodegradable polymers for modern vaccine development. *Journal of Industrial and Engineering Chemistry* **77**, 12-24, doi:10.1016/j.jiec.2019.04.044 (2019).

14 Han, J. *et al.* Polymer-based nanomaterials and applications for vaccines and drugs. *Polymers (Basel)* **10**, 31, doi:10.3390/polym10010031 (2018).

15 Bachmann, M. F. & Jennings, G. T. Vaccine delivery: A matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **10**, 787-796, doi:10.1038/nri2868 (2010).

16 Peleteiro, M. *et al.* Polymeric Nanocapsules for Vaccine Delivery: Influence of the Polymeric Shell on the Interaction With the Immune System. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.00791 (2018).

17 Lambricht, L. *et al.* in *Micro and Nanotechnology in Vaccine Development* (eds Mariusz Skwarczynski & Istvan Toth) 185-203 (William Andrew Publishing, 2017).

18 Sawant, R. R. & Torchilin, V. P. Liposomes as 'smart' pharmaceutical nanocarriers. *Soft Matter* **6**, 4026-4044, doi:10.1039/B923535N (2010).

19 López-Sagaseta, J. *et al.* Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

20 Scheiblhofer, S. *et al.* Influence of protein fold stability on immunogenicity and its implications for vaccine design. *Expert Review of Vaccines* **16**, 479-489, doi:10.1080/14760584.2017.1306441 (2017).

21 McHugh, C. A. *et al.* Improved stability of a protein vaccine through elimination of a partially unfolded state. *Protein science : a publication of the Protein Society* **13**, 2736-2743, doi:10.1110/ps.04897904 (2004).

22 Structural Genomics, C. *et al.* Protein production and purification. *Nat Methods* **5**, 135-146, doi:10.1038/nmeth.f.202 (2008).

23 Morales-Camacho, J. I. *et al.*Expression, purification and thermal stability evaluation of an engineered amaranth protein expressed in *Escherichia coli*. *Electronic Journal of Biotechnology* **22**, 44-51, doi:https://doi.org/10.1016/j.ejbt.2016.04.001 (2016).

24 Pronker, E. S. *et al.*. Risk in vaccine research and development quantified. *PloS one* **8**, e57755-e57755, doi:10.1371/journal.pone.0057755 (2013).

25 Hospital, A. *et al.* Molecular dynamics simulations: Advances and applications. *Advances and Application in Bioinformatics and Chemistry* **8**, 37-47, doi:10.2147/AABC.S70333 (2015).

26 McCammon, J. A. *et al.* Dynamics of folded proteins. *Nature* **267**, 585-590, doi:10.1038/267585a0 (1977).

27 Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **103**, 227-249, doi:10.1016/0022-2836(76)90311-9 (1976).

28 Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances* **41**, 107547, doi:https://doi.org/10.1016/j.biotechadv.2020.107547 (2020).

29 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine: Nanotechnology, Biology and Medicine* **2**, 137-149, doi:https://doi.org/10.1016/j.nano.2006.07.009 (2006).

30 Mohsen, M. O. *et al.* Major findings and recent advances in virus–like particle (VLP)-based vaccines. *Seminars in Immunology* **34**, 123-132, doi:https://doi.org/10.1016/j.smim.2017.08.014 (2017).

31 He, D. & Marles-Wright, J. Ferritin family proteins and their use in bionanotechnology. *New Biotechnology* **32**, 651-657, doi:https://doi.org/10.1016/j.nbt.2014.12.006 (2015).

32 Truffi, M. *et al.* Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacological Research* **107**, 57-65, doi:https://doi.org/10.1016/j.phrs.2016.03.002 (2016).

33 Jin, Y. *et al.* Ferritin variants: inspirations for rationally designing protein nanocarriers. *Nanoscale* **11**, 12449-12459, doi:10.1039/C9NR03823J (2019).

34 Zang, J. *et al.* Ferritin cage for encapsulation and delivery of bioactive nutrients: From structure, property to applications. *Critical Reviews in Food Science and Nutrition* **57**, 3673-3683, doi:10.1080/10408398.2016.1149690 (2017).

35 Han, J.-A. *et al.* Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine: Nanotechnology, Biology and Medicine* **10**, 561-569, doi:https://doi.org/10.1016/j.nano.2013.11.003 (2014).

36 Harrison, P. M. The structure and function of ferritin. *Biochemical Education* **14**, 154-162, doi:https://doi.org/10.1016/0307-4412(86)90203-7 (1986).

37 Luzzago, A. & Cesareni, G. Isolation of point mutations that affect the folding of the H chain of human ferritin in E.coli. *The EMBO Journal* **8**, 569-576, doi:10.1002/j.1460-2075.1989.tb03411.x (1989).

38 Jappelli, R. *et al.* Loop mutations can cause a substantial conformational change in the carboxy terminus of the ferritin protein. *Journal of Molecular Biology* **227**, 532-543, doi:https://doi.org/10.1016/0022-2836(92)90905-Y (1992).

39 SANTAMBROGIO, P. *et al.* Effects of modifications near the 2-, 3- and 4-fold symmetry axes on human ferritin renaturation. *Biochemical Journal* **322**, 461-468, doi:10.1042/bj3220461 (1997).

40 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

41 Li, C. Q., Soistman, E. & Carter, D. C. Ferritin nanoparticle technology: A new platform for antigen presentation and vaccine development. *Industrial Biotechnology* **2**, 143-147, doi:10.1089/ind.2006.2.143 (2006).

42 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102, doi:10.1038/nature12202 https://www.nature.com/articles/nature12202#supplementary-information (2013).

43 Sliepen, K. *et al.* Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **12**, 82, doi:10.1186/s12977-015-0210-4 (2015).

44 Zhen, Z. *et al.* Ferritins as nanoplatforms for imaging and drug delivery. *Expert opinion on drug delivery* **11**, 1913-1922, doi:10.1517/17425247.2014.941354 (2014).

45 De Souza, A. J. & Kane, L. P. Immune regulation by the TIM gene family. *Immunologic Research* **36**, 147-155, doi:10.1385/IR:36:1:147 (2006).

46 Fargion, S. *et al.* Specific binding sites for H-ferritin on human lymphocytes: modulation during cellular proliferation and potential implication in cell growth control. *Blood* **78**, 1056-1061 (1991).

47 Lee, B.-R. *et al.* Engineered Human Ferritin Nanoparticles for Direct Delivery of Tumor Antigens to Lymph Node and Cancer Immunotherapy. *Scientific reports* **6**, 35182-35182, doi:10.1038/srep35182 (2016).

48 Bachmann, M. F. *et al.* Long-lived memory CD8+ T cells are programmed by prolonged antigen exposure and low levels of cellular activation. *European Journal of Immunology* **36**, 842-854, doi:10.1002/eji.200535730 (2006).

49 Wang, L. *et al.* Structure-based design of ferritin nanoparticle immunogens displaying antigenic loops of Neisseria gonorrhoeae. *FEBS Open Bio* **7**, 1196-1207,

doi:10.1002/2211-5463.12267 (2017).

50 Lee, N. K. *et al.* Ferritin nanocage with intrinsically disordered proteins and affibody: A platform for tumor targeting with extended pharmacokinetics. *Journal of Controlled Release* **267**, 172-180, doi:https://doi.org/10.1016/j.jconrel.2017.08.014 (2017).

51 Guo, J. *et al.* Efficient expression of recombinant human heavy chain ferritin (FTH1) with modified peptides. *Protein Expression and Purification* **131**, 101-108, doi:https://doi.org/10.1016/j.pep.2016.06.003 (2017).

52 Chul, K. K. *et al.* Enhanced In Vivo Tumor Detection by Active Tumor Cell Targeting Using Multiple Tumor Receptor-Binding Peptides Presented on Genetically Engineered Human Ferritin Nanoparticles. *Small* **12**, 4241-4253, doi:10.1002/smll.201600917 (2016).

53 Lee, J.-H. *et al.* Proteinticle Engineering for Accurate 3D Diagnosis. *ACS Nano* **7**, 10879-10886, doi:10.1021/nn404325t (2013).

54 Kang, Y. J. *et al.* Incorporation of Thrombin Cleavage Peptide into a Protein Cage for Constructing a Protease-Responsive Multifunctional Delivery Nanoplatform. *Biomacromolecules* **13**, 4057-4064, doi:10.1021/bm301339s (2012).

55 Kang, H. J. *et al.* Developing an antibody-binding protein cage as a molecular recognition drug modular nanoplatform. *Biomaterials* **33**, 5423-5430, doi:https://doi.org/10.1016/j.biomaterials.2012.03.055 (2012).

56 Jeon, J. O. *et al.* Designed Nanocage Displaying Ligand-Specific Peptide Bunches for High Affinity and Biological Activity. *ACS Nano* **7**, 7462-7471, doi:10.1021/nn403184u (2013).

57 Lee, J.-L. *et al.* Functional expression and production of human H-ferritin in Pichia pastoris. *Biotechnology Letters* **25**, 1019-1023, doi:10.1023/a:1024193104858 (2003).

58 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/c9bm00098d (2019).

59 Nemtseva, E. V. *et al.* Experimental approach to study the effect of mutations on the protein folding pathway. *PLoS One* **14**, e0210361-e0210361, doi:10.1371/journal.pone.0210361 (2019).

60 Choe, S. E. *et al.* Differential stabilization of two hydrophobic cores in the transition state of the villin 14T folding reaction. *Journal of Molecular Biology* **304**, 99-115, doi:10.1006/jmbi.2000.4190 (2000).

61 Lindberg, M. O. *et al*. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proceedings of National Academy of Sciences of the United States of America* **103**, 4083-4088, doi:10.1073/pnas.0508863103 (2006).

62 Matthews, J. M. & Fersht, A. R. Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry* **34**, 6805-6814, doi:10.1021/bi00020a027 (1995).

63 Pilipczuk, J. *et al.* Role of the disulfide bond in stabilizing and folding of the fimbrial protein DraE from uropathogenic Escherichia coli. *Journal of Biological Chemistry* **292**, 16136-16149, doi:10.1074/jbc.M117.785477 (2017).

64 Viguera, A. R. *et al.* The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *Journal of Molecular Biology* **247**, 670-681, doi:10.1006/jmbi.1994.0171 (1995).

65 Ingrassia, R. *et al.* Mutations of ferritin H chain C-terminus produced by nucleotide insertions have altered stability and functional properties. *Journal of Biochemistry* **139**, 881-885, doi:10.1093/jb/mvj101 (2006).

66 Fan, R. *et al.* A Helix Swapping Study of Two Protein Cages. *Biochemistry* **48**, 5623-5630, doi:10.1021/bi900387t (2009).

67 Zhang, Y. & Orner, B. P. Self-assembly in the ferritin nano-cage protein superfamily. *International Journal of Molecular Science* **12**, 5406-5421, doi:10.3390/ijms12085406 (2011).

68 Levi, S. *et al.* Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferro-oxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants. *Journal of Biological Chemistry* **263**, 18086-18092 (1988).

69 Levi, S. *et al.* Mutational analysis of the channel and loop sequences of human ferritin H-chain. *Biochemical Journal* **264**, 381-388, doi:10.1042/bj2640381 (1989).

70 Doll, T. A. P. F. *et al.* Nanoscale assemblies and their biomedical applications. *Journal of The Royal Society Interface* **10**, 20120740, doi:10.1098/rsif.2012.0740 (2013).

71 Frietze, K. M. *et al.* Engineering virus-like particles as vaccine platforms. *Current Opinion in Virology* **18**, 44-49, doi:10.1016/j.coviro.2016.03.001 (2016).

72 Manolova, V. *et al.* Nanoparticles target distinct dendritic cell populations according to their size. *European Journal of Immunology* **38**, 1404-1413, doi:10.1002/eji.200737984

(2008).

73 Zeltins, A. Construction and Characterization of Virus-Like Particles: A Review. *Molecular Biotechnology* **53**, 92-107, doi:10.1007/s12033-012-9598-4 (2013).

74 Gamvrellis, A. *et al.* Vaccines that facilitate antigen entry into dendritic cells. *Immunology & Cell Biology* **82**, 506-516, doi:10.1111/j.0818-9641.2004.01271.x (2004).

75 Slütter, B. & Jiskoot, W. Sizing the optimal dimensions of a vaccine delivery system: a particulate matter. *Expert Opinion on Drug Delivery* **13**, 167-170, doi:10.1517/17425247.2016.1121989 (2016).

76 Zhao, Q. *et al.* Virus-like particle-based human vaccines: quality assessment based on structural and functional properties. *Trends in Biotechnology* **31**, 654-663, doi:https://doi.org/10.1016/j.tibtech.2013.09.002 (2013).

77 Golmohammadi, R. *et al.* The crystal structure of bacteriophage Qβ at 3.5 å resolution. *Structure* **4**, 543-554, doi:https://doi.org/10.1016/S0969-2126(96)00060-3 (1996).

78 Wynne, S. A. *et al.* The crystal structure of the human hepatitis B virus capsid. *Molecular Cell* **3**, 771-780, doi:https://doi.org/10.1016/S1097-2765(01)80009-5 (1999).

79 Baltabekova, A. Z. *et al.* SplitCore technology allows efficient production of virus-like particles presenting a receptor-contacting epitope of human IgE. *Molecular Biotechnology* **57**, 746-755, doi:10.1007/s12033-015-9867-0 (2015).

80 Li, Z. *et al.* Strong hydrophobicity enables efficient purification of HBc VLPs displaying various antigen epitopes through hydrophobic interaction chromatography. *Biochemical Engineering Journal* **140**, 157-167, doi:https://doi.org/10.1016/j.bej.2018.09.020 (2018).

81 Walker, A. *et al.* SplitCore: An exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. *Scientific Reports* **1**, 5, doi:10.1038/srep00005 (2011).

82 Kingsman, S. M. & Kingsman, A. J. Polyvalent recombinant antigens: a new vaccine strategy. *Vaccine* **6**, 304-306, doi:https://doi.org/10.1016/0264-410X(88)90174-0 (1988).

83 Strable, E. & Finn, M. G. Chemical modification of viruses and virus-like particles. *Current Topics in Microbiology and Immunology* **327**, 1-21, doi:10.1007/978-3-540-69379-6_1 (2009).

84 Grgacic, E. V. L. & Anderson, D. A. Virus-like particles: passport to immune recognition. *Methods* **40**, 60-65, doi:https://doi.org/10.1016/j.ymeth.2006.07.018 (2006).

85 Donaldson, B. *et al.* Virus-Like Particles, a versatile subunit vaccine platform. *Subunit Vaccine Delivery*, 159-180, doi:10.1007/978-1-4939-1417-3_9 (2014).

86 Vicente, T. *et al.* Large-scale production and purification of VLP-based vaccines. *Journal of Invertebrate Pathology* **107 Suppl**, S42-48, doi:10.1016/j.jip.2011.05.004 (2011).

87 Joshi, H. *et al.* Epitope engineering and molecular metrics of immunogenicity: A computational approach to VLP-based vaccine design. *Vaccine* **31**, 4841-4847, doi:https://doi.org/10.1016/j.vaccine.2013.07.075 (2013).

88 Chen, X. *et al*. Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

89 Zhao, H. L. *et al.* Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Expression and Purification* **61**, 73-77, doi:10.1016/j.pep.2008.04.013 (2008).

90 Amet, N. *et al.* Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharmaceutical research* **26**, 523-528, doi:10.1007/s11095-008-9767-0 (2009).

91 Bai, Y. *et al.* Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proceedings of National Academy of Sciences of the United States of America* **102**, 7292-7296, doi:10.1073/pnas.0500062102 (2005).

92 Bai, Y. & Shen, W. C. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharmaceutical Research* **23**, 2116-2121, doi:10.1007/s11095-006-9059-5 (2006).

93 Wei, Y. *et al.* Evaluation of lumazine synthase from Bacillus anthracis as a presentation platform for polyvalent antigen display. *Protein Science* **26**, 2059-2072, doi:10.1002/pro.3243 (2017).

94 van Rosmalen, M. *et al.* Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565-6574, doi:10.1021/acs.biochem.7b00902 (2017).

95 Chen, X. *et al.* Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

96 Schuler, B. *et al*. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proceedings of National Academy of Sciences of the United States of America* **102**, 2754-2759, doi:10.1073/pnas.0408164102 (2005).

97 Arai, R. *et al*. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering* **14**, 529-532, doi:10.1093/protein/14.8.529 (2001).

98 George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering, Design and Selection* **15**, 871-879, doi:10.1093/protein/15.11.871 (2002).

99 Klein, J. S. *et al.* Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering Design and Selection* **27**, 325-330, doi:10.1093/protein/gzu043 (2014).

100 Reddy Chichili, V. P. *et al.* Linkers in the structural biology of protein-protein interactions. *Protein science : a publication of the Protein Society* **22**, 153-167, doi:10.1002/pro.2206 (2013)

101 Sabourin, M. *et al.* A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. *Yeast (Chichester, England)* **24**, 39-45, doi:10.1002/yea.1431 (2007).

102 Robinson, C. R. & Sauer, R. T. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences* **95**, 5929, doi:10.1073/pnas.95.11.5929 (1998).

103 Nagi, A. D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Folding and Design* **2**, 67-75, doi:10.1016/s1359-0278(97)00007-2 (1997).

104 Chen, H. *et al.* Effect of linker length and flexibility on the clostridium thermocellum esterase displayed on bacillus subtilis spores. *Applied Biochemistry and Biotechnology* **182**, 168-180, doi:10.1007/s12010-016-2318-y (2017).

105 Farrell, P. J. Epstein–Barr virus and vancer. *Annual Review of Pathology: Mechanisms of Disease* **14**, 29-53, doi:10.1146/annurev-pathmechdis-012418-013023 (2019).

106 Pattle, S. B. & Farrell, P. J. The role of Epstein–Barr virus in cancer. *Expert Opinion on Biological Therapy* **6**, 1193-1205, doi:10.1517/14712598.6.11.1193 (2006).

107 Shi, Y. *et al.* Co-infection of Epstein-Barr virus and human papillomavirus in human tumorigenesis. *Chinese Journal of Cancer* **35**, 16-16, doi:10.1186/s40880-016-0079-1 (2016).

108 Mesri, Enrique A. *et al.* Human viral oncogenesis: a cancer hallmarks analysis. *Cell*

Host & Microbe **15**, 266-282, doi:https://doi.org/10.1016/j.chom.2014.02.011 (2014).

109 Münz, C. *et al.* Human Cd4+ T lymphocytes consistently respond to the latent Epstein-Barr virus nuclear antigen EBNA1. *Journal of Experimental Medicine* **191**, 1649-1660, doi:10.1084/jem.191.10.1649 (2000).

110 Destro, F. *et al.* Proteasome inhibitors induce the presentation of an Epstein-Barr virus nuclear antigen 1-derived cytotoxic T lymphocyte epitope in Burkitt's lymphoma cells. *Immunology* **133**, 105-114, doi:10.1111/j.1365-2567.2011.03416.x (2011).

111 Pati, R. *et al.* Nanoparticle vaccines against infectious diseases. *Frontiers in Immunology* **9**, 2224-2224, doi:10.3389/fimmu.2018.02224 (2018).

112 Lung, P. *et al.* Nanoparticle formulated vaccines: opportunities and challenges. *Nanoscale* **12**, 5746-5763, doi:10.1039/C9NR08958F (2020).

113 Gregory, A. E. *et al.* Vaccine delivery using nanoparticles. *Frontiers in Cellular and Infection Microbiology* **3**, 13-13, doi:10.3389/fcimb.2013.00013 (2013).

114 Ramírez-Salinas, G. L. *et al.* Bioinformatics design and experimental validation of influenza A virus multi-epitopes that induce neutralizing antibodies. *Archives of Virology* **165**, 891-911, doi:10.1007/s00705-020-04537-2 (2020).

115 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).

116 Shamriz, S. & Ofoghi, H. Design, structure prediction and molecular dynamics simulation of a fusion construct containing malaria pre-erythrocytic vaccine candidate, PfCelTOS, and human interleukin 2 as adjuvant. *BMC Bioinformatics* **17**, 71-71, doi:10.1186/s12859-016-0918-8 (2016).

117 Schaller, A. *et al.* Computational study of elements of stability of a four-helix bundle protein biosurfactant. *Journal of Computer-Aided Molecular Design* **29**, 47-58, doi:10.1007/s10822-014-9803-6 (2015).

118 Bekker, G.-J. *et al.* Thermal stability of single-domain antibodies estimated by molecular dynamics simulations. *Protein Science* **28**, 429-438, doi:10.1002/pro.3546 (2019).

119 Geng, H. *et al.* Applications of molecular dynamics simulation in structure prediction of peptides and proteins. *Computational and structural biotechnology journal* **17**, 1162-1170, doi:10.1016/j.csbj.2019.07.010 (2019).

120 Amir, M. *et al.* Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. *Frontiers in Molecular Biosciences* **6**,

doi:10.3389/fmolb.2019.00041 (2019).

121 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).

122 Kelly, S. M. *et al.* How to study proteins by circular dichroism. *Biochimica et Biophysica Acta* **1751**, 119-139, doi:10.1016/j.bbapap.2005.06.005 (2005).

123 Hasan, M. *et al.* Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infection, Genetics and Evolution* **70**, 140-157, doi:https://doi.org/10.1016/j.meegid.2019.03.003 (2019).

124 Gupta, A. *et al.* Vaccine candidate designed against carcinoembryonic antigen-related cell adhesion molecules using immunoinformatics tools. *Journal of Biomolecular Structure and Dynamics*, 1-15, doi:10.1080/07391102.2020.1797539 (2020).

125 Rostamian, M. *et al.* Immunoinformatics and molecular dynamics studies to predict T-cell-specific epitopes of four Klebsiella pneumoniae fimbriae antigens. *Journal of Biomolecular Structure and Dynamics*, 1-11, doi:10.1080/07391102.2020.1810126 (2020).

126 Gupta, N. & Kumar, A. Identification of potent vaccine candidates against campylobacter jejuni using immunoinformatics approach. *International Journal of Peptide Research and Therapeutics* **26**, 1303-1312, doi:10.1007/s10989-019-09933-0 (2020).

127 Samad, A. *et al.* Designing a multi-epitope vaccine against SARS-CoV-2: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1-17, doi:10.1080/07391102.2020.1792347 (2020).

128 Calis, J. J. A. *et al.* Properties of MHC class i presented peptides that enhance immunogenicity. *PLOS Computational Biology* **9**, e1003266, doi:10.1371/journal.pcbi.1003266 (2013).

129 Rammensee, H. *et al.* SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219, doi:10.1007/s002510050595 (1999).

130 Singh, H. & Raghava, G. P. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* **19**, 1009-1014, doi:10.1093/bioinformatics/btg108 (2003).

131 Solanki, V. & Tiwari, V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Scientific Reports* **8**, 9044, doi:10.1038/s41598-018-26689-7 (2018).

132 Lopéz-Blanco, J. R. *et al.* iMod: multipurpose normal mode analysis in internal

coordinates. *Bioinformatics* **27**, 2843-2850, doi:10.1093/bioinformatics/btr497 (2011).

133 Bai, Q. & Yao, X. Investigation of allosteric modulation mechanism of metabotropic glutamate receptor 1 by molecular dynamics simulations, free energy and weak interaction analysis. *Scientific Reports* **6**, 21763, doi:10.1038/srep21763 (2016).

134 Bello, M. & Correa-Basurto, J. Molecular dynamics simulations to provide insights into epitopes coupled to the soluble and membrane-bound MHC-II complexes. *PLOS ONE* **8**, e72575, doi:10.1371/journal.pone.0072575 (2013).

135 Chen, J. *et al.* Molecular mechanism and energy basis of conformational diversity of antibody SPE7 revealed by molecular dynamics simulation and principal component analysis. *Scientific Reports* **6**, 36900, doi:10.1038/srep36900 (2016).

136 Knapp, B. *et al.* Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations. *Briefings in Bioinformatics* **16**, 1035-1044, doi:10.1093/bib/bbv005 (2015).

137 Usman Mirza, M. *et al.* Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports* **6**, 37313, doi:10.1038/srep37313 (2016).

138 Singaravelu, M. *et al.* Molecular dynamics simulations of lectin domain of FimH and immunoinformatics for the design of potential vaccine candidates. *Computational Biology and Chemistry* **52**, 18-24, doi:https://doi.org/10.1016/j.compbiolchem.2014.08.002 (2014).

139 Enayatkhani, M. *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *Journal of Biomolecular Structure and Dynamics*, 1-16, doi:10.1080/07391102.2020.1756411 (2020).

140 Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **10**, 449-461, doi:10.1517/17460441.2015.1032936 (2015).

141 Nain, Z. *et al.* Immunoinformatic and dynamic simulation-based designing of a multi-epitope vaccine against emerging pathogen Elizabethkingia anophelis. *bioRxiv*, 758219, doi:10.1101/758219 (2019).

142 Gupta, N. & Kumar, A. Designing an efficient multi-epitope vaccine against Campylobacter jejuni using immunoinformatics and reverse vaccinology approach. *Microbial Pathogenesis* **147**, 104398,

doi:https://doi.org/10.1016/j.micpath.2020.104398 (2020).

143 Pavitrakar, D. V. *et al.* Design of a multi-epitope peptide vaccine candidate against chandipura virus: an immuno-informatics study. *Journal of Biomolecular Structure and Dynamics*, 1-12, doi:10.1080/07391102.2020.1816493 (2020).

144 Deng, H. *et al.* Development of a multivalent enterovirus subunit vaccine based on immunoinformatic design principles for the prevention of HFMD. *Vaccine* **38**, 3671-3681, doi:https://doi.org/10.1016/j.vaccine.2020.03.023 (2020).

145 Banisharif-Dehkordi, F. *et al.* Design and molecular dynamic simulation of a new double-epitope tolerogenic protein as a potential vaccine for multiple sclerosis disease. *Research in Pharmaceutical Sciences* **14**, 20-26, doi:10.4103/1735-5362.251849 (2019).

146 Kumar, N. *et al.* Design and optimization of a subunit vaccine targeting COVID-19 molecular shreds using an immunoinformatics framework. *RSC Advances* **10**, 35856-35872, doi:10.1039/D0RA06849G (2020).

147 Kumar, N. *et al.* Antimicrobial Peptide Designing and Optimization Employing Large-Scale Flexibility Analysis of Protein-Peptide Fragments. *ACS Omega* **4**, 21370-21380, doi:10.1021/acsomega.9b03035 (2019).

148 Sood, D. *et al.* Mechanistic interaction study of bromo-noscapine with bovine serum albumin employing spectroscopic and chemoinformatics approaches. *Scientific Reports* **8**, 16964, doi:10.1038/s41598-018-35384-6 (2018).

149 López-Blanco, J. R. *et al.* iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Research* **42**, W271-W276, doi:10.1093/nar/gku339 (2014).

150 Flower, D. R. *et al.* T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Research* **6 Suppl 2**, S4-S4, doi:10.1186/1745-7580-6-S2-S4 (2010).

151 Dror, R. O. *et al.* Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology* **135**, 555-562, doi:10.1085/jgp.200910373 (2010).

152 Abhinav, B. *et al.* in *2008 IEEE International Symposium on Parallel and Distributed Processing.* 1-12.

153 Hess, B. *et al.* GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory Computation* **4**, 435-447, doi:10.1021/ct700301q (2008).

154 Klepeis, J. L. *et al.* Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* **19**, 120-127, doi:10.1016/j.sbi.2009.03.004 (2009).

155 Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197, doi:10.1021/ja00124a002 (1995).

156 Guvench, O. & MacKerell, A. D., Jr. Comparison of protein force fields for molecular dynamics simulations. *Methods in Molecular Biology* **443**, 63-88, doi:10.1007/978-1-59745-177-2_4 (2008).

157 Zhang, H. *et al.* Limitations of *ab Initio* predictions of peptide binding to MHC class II molecules. *PLOS ONE* **5**, e9272, doi:10.1371/journal.pone.0009272 (2010).

158 Peters, B. *et al.* A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLOS Computational Biology* **2**, e65, doi:10.1371/journal.pcbi.0020065 (2006).

159 John, B. & Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research* **31**, 3982-3992, doi:10.1093/nar/gkg460 (2003).

160 Qu, X. *et al.* A guide to template based structure prediction. *Current Protein and Peptide Science* **10**, 270-285, doi:10.2174/138920309788452182 (2009).

161 Fiser, A. Template-based protein structure modeling. *Methods in Molecular Biology* **673**, 73-94, doi:10.1007/978-1-60761-842-3_6 (2010).

162 Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

163 Ji, X. *et al.* Homology modeling and molecular dynamics simulation studies of a marine alkaline protease. *Bioinformatics and Biology Insights* **6**, 255-263, doi:10.4137/BBI.S10663 (2012).

164 Yang, H.-C. *et al.* Homology modeling and molecular dynamics simulation combined with x-ray solution scattering defining protein structures of thromboxane and prostacyclin synthases. *The Journal of Physical Chemistry B* **121**, 11229-11240, doi:10.1021/acs.jpcb.7b08299 (2017).

165 Lemer, C. M. *et al.* Protein structure prediction by threading methods: evaluation of current techniques. *Proteins* **23**, 337-355, doi:10.1002/prot.340230308 (1995).

166 Peng, J. & Xu, J. Boosting protein threading accuracy. *Research in Computational Molecular Biology* **5541**, 31-45, doi:10.1007/978-3-642-02008-7_3 (2009).

167 Cheung, N. J. & Yu, W. *De novo* protein structure prediction using ultra-fast molecular dynamics simulation. *PLOS ONE* **13**, e0205819, doi:10.1371/journal.pone.0205819 (2018).

168 Childers, M. C. & Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering* **2**, 9-33, doi:10.1039/C6ME00083E (2017).

169 Li, J. *et al.* An improved integration of template-based and template-free protein structure modeling methods and its assessment in CASP11. *Protein and Peptide Letters* **22**, 586-593, doi:10.2174/0929866522666150520145717 (2015).

170 Gill, M. & McCully, M. E. Molecular dynamics simulations suggest stabilizing mutations in a *de novo* designed α/β protein. *Protein Engineering, Design and Selection* **32**, 317-329, doi:10.1093/protein/gzaa005 (2019).

171 Bradley, P. *et al.* Toward high-resolution *de novo* Structure prediction for small proteins. *Science* **309**, 1868, doi:10.1126/science.1113801 (2005).

172 Rohl, C. A. *et al*. In *Methods in Enzymology* Vol. 383 66-93 (Academic Press, 2004).

173 Xu, D. & Zhang, Y. *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715-1735, doi:10.1002/prot.24065 (2012).

174 Kato, K. *et al.* Validation of molecular dynamics simulations for prediction of three-dimensional structures of small proteins. *Molecules* **22**, doi:10.3390/molecules22101716 (2017).

175 Vallat, B. *et al.* Modularity of protein folds as a tool for template-free modeling of structures. *PLOS Computational Biology* **11**, e1004419, doi:10.1371/journal.pcbi.1004419 (2015).

176 Shamriz, S. & Ofoghi, H. Design, structure prediction and molecular dynamics simulation of a fusion construct containing malaria pre-erythrocytic vaccine candidate, PfCelTOS, and human interleukin 2 as adjuvant. *BMC Bioinformatics* **17**, 71, doi:10.1186/s12859-016-0918-8 (2016).

177 Von Holst, H. & Li, X. Numerical impact simulation of gradually increased kinetic energy transfer has the potential to break up folded protein structures resulting in cytotoxic brain tissue edema. *Journal of Neurotrauma* **30**, 1192-1199, doi:10.1089/neu.2012.2730 (2013).

178 Mobini, S. *et al.* Computational design of a novel VLP-based vaccine for hepatitis B virus. *Frontiers in Immunology* **11**, doi:10.3389/fimmu.2020.02074 (2020).

179 Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annual Review of Biochemistry* **77**, 363-382, doi:10.1146/annurev.biochem.77.062906.171838 (2008).

180 Shen, Y. & Bax, A. Homology modeling of larger proteins guided by chemical shifts. *Nature Methods* **12**, 747-750, doi:10.1038/nmeth.3437 (2015).

181 Kim, D. E. *et al.* Sampling bottlenecks in *de novo* protein structure prediction. *Journal of Molecular Biology* **393**, 249-260, doi:https://doi.org/10.1016/j.jmb.2009.07.063 (2009).

182 Söding, J. Big-data approaches to protein structure prediction. *Science* **355**, 248, doi:10.1126/science.aal4512 (2017).

183 Freddolino, P. L. *et al.* Challenges in protein folding simulations: Timescale, representation, and analysis. *Nature Physics* **6**, 751-758, doi:10.1038/nphys1713 (2010).

184 Schaeffer, R. D. *et al.* Combining experiment and simulation in protein folding: closing the gap for small model systems. *Current opinion in structural biology* **18**, 4-9, doi:10.1016/j.sbi.2007.11.007 (2008).

185 Gershenson, A. *et al.*. Successes and challenges in simulating the folding of large proteins. *The Journal of biological chemistry* **295**, 15-33, doi:10.1074/jbc.REV119.006794 (2020).

186 Wang, G. & Dunbrack, R. L., Jr. PISCES: A protein sequence culling server. *Bioinformatics* **19**, 1589-1591, doi:10.1093/bioinformatics/btg224 (2003).

187 Zheng, W. *et al.* Frustration in the energy landscapes of multidomain protein misfolding. *Proceedings of National Academy of Sciences of the United States of America* **110**, 1680-1685, doi:10.1073/pnas.1222130110 (2013).

188 Han, J. H. *et al.* The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology* **8**, 319-330, doi:10.1038/nrm2144 (2007).

189 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

190 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876, doi:10.1126/science.abj8754 (2021).

191 Sviben, D. *et al.* Investigation of the thermal shift assay and its power to predict protein and virus stabilizing conditions. *Journal of Pharmaceutical and Biomedical Analysis* **161**, 73-82, doi:https://doi.org/10.1016/j.jpba.2018.08.017 (2018).

192 Blundell, T. L. *et al.* Protein engineering and design. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **324**, 447-460,

doi:10.1098/rstb.1989.0059 (1989).

193 Hellinga, H. W. Computational protein engineering. *Nature Structural Biology* **5**, 525-527, doi:10.1038/776 (1998).

194 Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports* **7**, 44651, doi:10.1038/srep44651 (2017).

195 Jiang, Z. *et al.* Effects of an electric field on the conformational transition of the protein: A molecular dynamics simulation study. *Polymers (Basel)* **11**, 282, doi:10.3390/polym11020282 (2019).

196 Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology* **235**, 625-634, doi:10.1006/jmbi.1994.1017 (1994).

197 Chen, Z. *et al.* Molecular dynamics simulation of barnase: Contribution of noncovalent intramolecular interaction to thermostability. *Mathematical Problems in Engineering* **2013**, 504183, doi:10.1155/2013/504183 (2013).

198 Manjunath, K. & Sekar, K. Molecular dynamics perspective on the protein thermal stability: A case study using SAICAR synthetase. *Journal of Chemical Information and Modeling* **53**, 2448-2461, doi:10.1021/ci400306m (2013).

199 Pikkemaat, M. G. *et al.* Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering, Design and Selection* **15**, 185-192, doi:10.1093/protein/15.3.185 (2002).

200 Kumar, C. V. *et al.* Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A. *Advanced Bioinformatics* **2014**, 502618-502618, doi:10.1155/2014/502618 (2014).

201 Lobanov, M. Y. *et al.* Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 623-628, doi:10.1134/S0026893308040195 (2008).

202 Arnittali, M. *et al.* Structure of biomolecules through molecular dynamics simulations. *Procedia Computer Science* **156**, 69-78, doi:https://doi.org/10.1016/j.procs.2019.08.181 (2019).

203 Amir, M. *et al.* Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. *Frontiers in Molecular Biosciences* **6**, 41-41, doi:10.3389/fmolb.2019.00041 (2019).

204 Zhao, Y. *et al.* Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. *PLOS ONE* **10**, e0124377, doi:10.1371/journal.pone.0124377 (2015).

205 Burton, B. *et al.* A computational investigation on the connection between dynamics properties of ribosomal proteins and ribosome assembly. *PLOS Computational Biology* **8**, e1002530, doi:10.1371/journal.pcbi.1002530 (2012).

206 Abdulazeez, S. Molecular simulation studies on B-cell lymphoma/leukaemia 11A (BCL11A). *American Journal of Translational Research* **11**, 3689-3697 (2019).

207 Chapman, A. D. *et al.* Structural basis of substrate specificity in malate dehydrogenases: crystal structure of a ternary complex of porcine cytoplasmic malate dehydrogenase, alpha-ketomalonate and tetrahydoNAD. *Journal of Molecular Biology* **285**, 703-712, doi:10.1006/jmbi.1998.2357 (1999).

208 Dong, Y.-w. *et al.* Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences* **115**, 1274, doi:10.1073/pnas.1718910115 (2018).

209 Best, R. B. *et al.* Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of National Academy of Sciences of the United States of America* **110**, 17874-17879, doi:10.1073/pnas.1311599110 (2013).

210 Xu, Y. *et al.* Stabilizing effect of inherent knots on proteins revealed by molecular dynamics simulations. *Biophysical Journal* **115**, 1681-1689, doi:https://doi.org/10.1016/j.bpj.2018.09.015 (2018).

211 Roy, S. *et al.* The unfolding MD simulations of cyclophilin: analyzed by surface contact networks and their associated metrics. *PLOS ONE* **10**, e0142173, doi:10.1371/journal.pone.0142173 (2015).

212 Pace, C. N. *et al.* Contribution of hydrogen bonds to protein stability. *Protein Science* **23**, 652-661, doi:10.1002/pro.2449 (2014).

213 Chikalov, I. *et al.* Learning probabilistic models of hydrogen bond stability from molecular dynamics simulation trajectories. *BMC Bioinformatics* **12 Suppl 1**, S34-S34, doi:10.1186/1471-2105-12-S1-S34 (2011).

214 Chan, M. K. *et al.* Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463-1469, doi:10.1126/science.7878465 (1995).

215 Lee, Y. *et al.* Dissecting the critical factors for thermodynamic stability of modular proteins using molecular modeling approach. *PLOS ONE* **9**, e98243, doi:10.1371/journal.pone.0098243 (2014).

216 Kamaraj, B. & Purohit, R. *In silico* screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *BioMed Research International* **2013**, 697051, doi:10.1155/2013/697051 (2013).

217 Shy, M. E. *et al.* T118M PMP22 mutation causes partial loss of function and HNPP-like neuropathy. *Annals of Neurology* **59**, 358-364, doi:10.1002/ana.20777 (2006).

218 Muneeswaran, G. *et al.* Molecular dynamics simulation approach to explore atomistic molecular mechanism of peroxidase activity of apoptotic cytochrome c variants. *Informatics in Medicine Unlocked* **11**, 51-60, doi:https://doi.org/10.1016/j.imu.2018.04.003 (2018).

219 Ali, S. K. *et al.* Molecular dynamics-based analyses of the structural instability and secondary structure of the fibrinogen gamma chain protein with the D356V mutation. *Journal of Biomolecular Structure and Dynamics* **35**, 2714-2724, doi:10.1080/07391102.2016.1229634 (2017).

220 Medvedev, K. E. *et al.* Molecular dynamics simulations of the Nip7 proteins from the marine deep- and shallow-water Pyrococcus species. *BMC Structural Biology* **14**, 23, doi:10.1186/s12900-014-0023-z (2014).

221 Janin, J. *et al.* Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* **125**, 357-386, doi:https://doi.org/10.1016/0022-2836(78)90408-4 (1978).

222 Anderson, R. J. *et al.* Main-chain conformational tendencies of amino acids. *Proteins: Structure, Function, and Bioinformatics* **60**, 679-689, doi:10.1002/prot.20530 (2005).

223 Hollingsworth, S. A. & Karplus, P. A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *BioMolecular Concepts* **1**, 271-283, doi:10.1515/bmc.2010.022 (2010).

224 Matsuura, Y. *et al.* Evaluating the strengths of salt bridges in the CutA1 protein using molecular dynamic simulations: a comparison of different force fields. *FEBS Open Bio* **9**, 1939-1956, doi:10.1002/2211-5463.12731 (2019).

225 Pendley, S. S. *et al.* Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. *Proteins* **74**, 612-629, doi:10.1002/prot.22177 (2009).

226 Luo, Y. & Baldwin, R. L. How Ala-->Gly mutations in different helices affect the

stability of the apomyoglobin molten globule. *Biochemistry* **40**, 5283-5289, doi:10.1021/bi010122j (2001).

227 Beeby, M. *et al.* The genomics of disulfide bonding and protein stabilization in thermophiles. *PLOS Biology* **3**, e309, doi:10.1371/journal.pbio.0030309 (2005).

228 Dürrschmidt, P. *et al.* Differentiation between conformational and autoproteolytic stability of the neutral protease from Bacillus stearothermophilus containing an engineered disulfide bond. *European Journal of Biochemistry* **268**, 3612-3618, doi:10.1046/j.1432-1327.2001.02270.x (2001).

229 Chen, J. *et al.* Improving stability of nitrile hydratase by bridging the salt-bridges in specific thermal-sensitive regions. *Journal of Biotechnology* **164**, 354-362, doi:https://doi.org/10.1016/j.jbiotec.2013.01.021 (2013).

230 Makhatadze, G. I. *et al.*. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *Journal of Molecular Biology* **327**, 1135-1148, doi:https://doi.org/10.1016/S0022-2836(03)00233-X (2003).

231 Missimer, J. H. *et al.* Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Science* **16**, 1349-1359, doi:10.1110/ps.062542907 (2007).

232 Thomas, A. S. & Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *Journal of the American Chemical Society* **126**, 2208-2214, doi:10.1021/ja039159c (2004).

233 Papaleo, E. *et al.* Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. *Journal of Molecular Graphics and Modelling* **26**, 93-103, doi:https://doi.org/10.1016/j.jmgm.2006.09.012 (2007).

234 Vinther, J. M. *et al.* Enhanced stability of a protein with increasing temperature. *Journal of the American Chemical Society* **133**, 271-278, doi:10.1021/ja105388k (2011).

235 Dong, H. *et al.* hydrophobic effect on the stability and folding of a hyperthermophilic protein. *Journal of Molecular Biology* **378**, 264-272, doi:https://doi.org/10.1016/j.jmb.2008.02.039 (2008).

236 Pace, C. N. *et al.* Contribution of hydrophobic interactions to protein Stability. *Journal of Molecular Biology* **408**, 514-528, doi:https://doi.org/10.1016/j.jmb.2011.02.053 (2011).

237 Paiardini, A. *et al.* "Hot cores" in proteins: Comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms. *BMC Structural*

Biology **8**, 14, doi:10.1186/1472-6807-8-14 (2008).

238 Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Molecular Biology Reviews* **65**, 1-43, doi:10.1128/MMBR.65.1.1-43.2001 (2001).

239 Jaenicke, R. & Böhm, G. The stability of proteins in extreme environments. *Current Opinion in Structural Biology* **8**, 738-748, doi:https://doi.org/10.1016/S0959-440X(98)80094-8 (1998).

240 Vogt, G. & Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design* **2**, S40-S46, doi:https://doi.org/10.1016/S1359-0278(97)00062-X (1997).

241 Vogt, G. *et al.* Protein thermal stability, hydrogen bonds, and ion pairs11Edited by F. E. Cohen. *Journal of Molecular Biology* **269**, 631-643, doi:https://doi.org/10.1006/jmbi.1997.1042 (1997).

242 Zhou, H.-X. Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophysical Journal* **83**, 3126-3133, doi:10.1016/S0006-3495(02)75316-2 (2002).

243 Li, J. *et al.* A stable α-helix-rich intermediate is formed by a single mutation of the β-sheet protein, src SH3, at pH 3. *Journal of Molecular Biology* **372**, 747-755, doi:https://doi.org/10.1016/j.jmb.2007.07.001 (2007).

244 Meharenna, Y. T. & Poulos, T. L. Using molecular dynamics to probe the structural basis for enhanced stability in thermal stable cytochromes P450. *Biochemistry* **49**, 6680-6686, doi:10.1021/bi100929x (2010).

245 Bae, E. & Phillips, G. N., Jr. Identifying and engineering ion pairs in adenylate kinases. Insights from molecular dynamics simulations of thermophilic and mesophilic homologues. *Journal of Biological Chemistry* **280**, 30943-30948, doi:10.1074/jbc.M504216200 (2005).

246 Jones, D. H. *et al*. Efficient purification and rigorous characterisation of a recombinant gp120 for HIV vaccine studies. *Vaccine* **13**, 991-999, doi:https://doi.org/10.1016/0264-410X(95)00019-W (1995).

247 Santambrogio, P. *et al.* Functional and Immunological Analysis of Recombinant Mouse H- and L-Ferritins from Escherichia coli. *Protein Expression and Purification* **19**, 212-218, doi:https://doi.org/10.1006/prep.2000.1212 (2000).

248 Seo, H.-S. *et al.* Analysis and characterization of hepatitis B vaccine particles

synthesized from Hansenula polymorpha. *Vaccine* **26**, 4138-4144, doi:https://doi.org/10.1016/j.vaccine.2008.05.070 (2008).

249 Liu, J. *et al.* Characterization and Stabilization of Recombinant Human Protein Pentraxin (rhPTX-2). *Journal of Pharmaceutical Sciences* **102**, 827-841, doi:https://doi.org/10.1002/jps.23360 (2013).

250 Heider, S. *et al*. Integrated Method for Purification and Single-Particle Characterization of Lentiviral Vector Systems by Size Exclusion Chromatography and Tunable Resistive Pulse Sensing. *Molecular Biotechnology* **59**, 251-259, doi:10.1007/s12033-017-0009-8 (2017).

251 Agarwal, S. *et al.* Characterizing and Minimizing Aggregation and Particle Formation of Three Recombinant Fusion-Protein Bulk Antigens for Use in a Candidate Trivalent Rotavirus Vaccine. *Journal of Pharmaceutical Sciences* **109**, 394-406, doi:https://doi.org/10.1016/j.xphs.2019.08.001 (2020).

252 Maddux, N. R. *et al.* An Improved Methodology for Multidimensional High-Throughput Preformulation Characterization of Protein Conformational Stability. *Journal of Pharmaceutical Sciences* **101**, 2017-2024, doi:https://doi.org/10.1002/jps.23132 (2012).

253 Hewarathna, A. *et al.* Chemical Stability of the Botanical Drug Substance Crofelemer: A Model System for Comparative Characterization of Complex Mixture Drugs. *Journal of Pharmaceutical Sciences* **106**, 3257-3269, doi:https://doi.org/10.1016/j.xphs.2017.06.022 (2017).

254 Strub, C. *et al.* Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry* **5**, 9, doi:10.1186/1471-2091-5-9 (2004).

255 Zhu, D. *et al.* A quantitative slot blot assay for host cell protein impurities in recombinant proteins expressed in E. coli. *Journal of Immunological Methods* **306**, 40-50, doi:https://doi.org/10.1016/j.jim.2005.07.021 (2005).

256 Dagouassat, N. *et al.* Development of a quantitative assay for residual host cell proteins in a recombinant subunit vaccine against human respiratory syncytial virus. *Journal of Immunological Methods* **251**, 151-159, doi:https://doi.org/10.1016/S0022-1759(01)00321-0 (2001).

257 Trivedi, B. *et al.* Endotoxin content of standardized allergen vaccines. *Journal of Allergy and Clinical Immunology* **111**, 777-783, doi:https://doi.org/10.1067/mai.2003.1338 (2003).

258 Chen, Y. *et al.* A ferritin nanoparticle vaccine for foot-and-mouth disease virus elicited partial protection in mice. *Vaccine* **38**, 5647-5652, doi:https://doi.org/10.1016/j.vaccine.2020.06.063 (2020).

259 Burns, J. M. *et al*. Immunogenicity of a chimeric Plasmodium falciparum merozoite surface protein vaccine in Aotus monkeys. *Malaria Journal* **15**, 159, doi:10.1186/s12936-016-1226-5 (2016).

260 Driesen, J. *et al.* CD25 as an immune regulatory molecule expressed on myeloid dendritic cells. *Immunobiology* **213**, 849-858, doi:https://doi.org/10.1016/j.imbio.2008.07.026 (2008).

261 Krzych, U. *et al.* Memory T cells maintain protracted protection against malaria. *Immunology Letter* **161**, 189-195, doi:10.1016/j.imlet.2014.03.011 (2014).

262 Kissmann, J. *et al.* H1N1 Influenza Virus-Like Particles: Physical Degradation Pathways and Identification of Stabilizers. *Journal of Pharmaceutical Sciences* **100**, 634-645, doi:https://doi.org/10.1002/jps.22304 (2011).

263 Toprani, V. M. *et al.* Structural Characterization and Formulation Development of a Trivalent Equine Encephalitis Virus-Like Particle Vaccine Candidate. *Journal of Pharmaceutical Sciences* **107**, 2544-2558, doi:https://doi.org/10.1016/j.xphs.2018.05.022 (2018).

264 Krueger, S. *et al.* Structural Characterization and Modeling of a Respiratory Syncytial Virus Fusion Glycoprotein Nanoparticle Vaccine in Solution. *Molecular Pharmaceutics* **18**, 359-376, doi:10.1021/acs.molpharmaceut.0c00986 (2021).

265 Vessely, C. *et al.* Stability of a Trivalent Recombinant Protein Vaccine Formulation Against Botulinum Neurotoxin During Storage in Aqueous Solution. *Journal of Pharmaceutical Sciences* **98**, 2970-2993, doi:https://doi.org/10.1002/jps.21498 (2009).

266 Liang, Y. *et al.* Development and Delivery Systems of mRNA Vaccines. *Frontiers in Bioengineering and Biotechnology* **9**, doi:10.3389/fbioe.2021.718753 (2021).

267 Pardi, N. *et al.* mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery* **17**, 261-279, doi:10.1038/nrd.2017.243 (2018).

268 Kowalzik, F. *et al.* mRNA-Based Vaccines. *Vaccines* **9**, doi:10.3390/vaccines9040390 (2021).

269 Geall, A. J. *et al.* Nonviral delivery of self-amplifying RNA vaccines. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14604-14609,

doi:10.1073/pnas.1209367109 (2012).

270 Pardi, N. *et al.* Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. *Nature* **543**, 248-251, doi:10.1038/nature21428 (2017).

271 Chahal, J. S. *et al.* Dendrimer-RNA nanoparticles generate protective immunity against lethal Ebola, H1N1 influenza, and Toxoplasma gondii challenges with a single dose. *Proceedings of the National Academy of Sciences* **113**, E4133-E4142, doi:doi:10.1073/pnas.1600299113 (2016).

272 Benteyn, D.*et al.* mRNA-based dendritic cell vaccines. *Expert Rev Vaccines* **14**, 161-176, doi:10.1586/14760584.2014.957684 (2015).

273 Pardi, N. *et al.* Characterization of HIV-1 Nucleoside-Modified mRNA Vaccines in Rabbits and Rhesus Macaques. *Molecular Therapy - Nucleic Acids* **15**, 36-47, doi:https://doi.org/10.1016/j.omtn.2019.03.003 (2019).

274 Hou, X. *et al.* Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials* **6**, 1078-1094, doi:10.1038/s41578-021-00358-0 (2021).

275 Kim, J. *et al.* Self-assembled mRNA vaccines. *Advanced Drug Delivery Reviews* **170**, 83-112, doi:https://doi.org/10.1016/j.addr.2020.12.014 (2021).

276 Karnik, R. *et al.* Microfluidic Platform for Controlled Synthesis of Polymeric Nanoparticles. *Nano Letters* **8**, 2906-2912, doi:10.1021/nl801736q (2008).

277 Leung, A. K. K. *et al.* Microfluidic Mixing: A General Method for Encapsulating Macromolecules in Lipid Nanoparticle Systems. *The Journal of Physical Chemistry B* **119**, 8698-8706, doi:10.1021/acs.jpcb.5b02891 (2015).

278 Blakney, A. K. *et al.* Inside out: optimization of lipid nanoparticle formulations for exterior complexation and in vivo delivery of saRNA. *Gene Therapy* **26**, 363-372, doi:10.1038/s41434-019-0095-2 (2019).

279 Heyes, J. *et al.* Cationic lipid saturation influences intracellular delivery of encapsulated nucleic acids. *Journal of Controlled Release* **107**, 276-287, doi:https://doi.org/10.1016/j.jconrel.2005.06.014 (2005).

280 Alabi, C. A. *et al.* Multiparametric approach for the evaluation of lipid nanoparticles for siRNA delivery. *Proceedings of the National Academy of Sciences* **110**, 12881-12886, doi:doi:10.1073/pnas.1306529110 (2013).

281 Kranz, L. M. *et al.* Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396-401, doi:10.1038/nature18300 (2016).

282 Cheng, Q. *et al.* Selective organ targeting (SORT) nanoparticles for tissue-specific

mRNA delivery and CRISPR–Cas gene editing. *Nature Nanotechnology* **15**, 313-320, doi:10.1038/s41565-020-0669-6 (2020).

CHAPTER 3 STABILITY OF ENGINEERED FERRITIN NANO VACCINES INVESTIGATED BY COMBINED MOLECULAR SIMULATION AND EXPERIMENT

Statement of Authorship

Title of Paper	Stability of engineered ferritin nano vaccines investigated by combined molecular simulation and experiment
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Qu, Y.; Wang, L.; Yin, S.; Zhang, B.; Jiao, Y.; Sun, Y.; Middelberg, A.; Bi, J. Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments. *The Journal of Physical Chemistry B* 2021, 125, 3830-3842.

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Designed experiment, performed experiments; analysed data; writing manuscript		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30/07/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lijie Wang		
Contribution to the Paper	Designed experiment		
Signature		Date	30/07/2021

Name of Co-Author	Shuang Yin		
Contribution to the Paper	Designed experiment		
Signature		Date	30/07/2021

Name of Co-Author	Bingyang Zhang		
Contribution to the Paper	Performed experiments		
Signature		Date	30/07/2021

Name of Co-Author	Yan Jiao		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	01/08/2021

Name of Co-Author	Yan Sun		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	20/09/2021

Name of Co-Author	Anton Middelberg		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	13/10/21

Name of Co-Author	Jingxiu Bi		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/2021

Stability of engineered ferritin nano vaccines investigated by combined molecular simulation and experiment

Yiran Qu, Lijie Wang, Shuang Yin, Bingyang Zhang, Yan Jiao, Yan Sun, Anton Middelberg, Jingxiu Bi*

Y. Qu, S. Yin, B. Zhang, Dr. Y. Jiao, A/Prof. Dr. J. Bi, Prof. Dr. A. Middelberg

School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide
SA 5005 (Australia)

L. Wang, Prof. Dr. Y. Sun
Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072 (China)

Journal of Physical Chemistry B 2021, 125, 15, 3830–3842 / doi:10.1021/acs.jpcb.1c00276.

Abstract

Human ferritin is regarded as an attractive and promising vaccine platform because of its uniform structure, good plasticity and desirable thermal and chemical stabilities. Besides, it is biocompatible and presumed safe when used as a vaccine carrier. However, there is a lack of knowledge of how different antigen insertion sites on the ferritin nanocage affects the resulting protein stability and performance. To address this question, Epstein-Barr nuclear antigen 1 (EBNA1) was selected as a model epitope and fused it at the DNA level with different insertion sites, namely the N- and C-termini of ferritin, to engineer proteins E1F1 and F1E1, respectively. Protein properties including hydrophobicity, thermal-, pH- and chemical-stability were investigated both by molecular dynamics (MD) simulation and by experiment. Both methods demonstrate that the insertion site plays an important role on protein properties. The C-terminus insertion (F1E1) leads to a less hydrophobic surface and more tolerance to the external influence of high temperature, pH and high concentration of chemical denaturants compared to N-terminus insertion (E1F1). Simulated protein hydrophobicity and thermal stability by MD were in high accordance with experimental results. Therefore, MD simulation can be used as a valuable tool to engineer nano-vaccine candidates, cutting down costs by reducing experimental effort and accelerating vaccine design.

KEYWORDS: ferritin, thermal stability, pH stability, chemical stability, antigen delivery, molecular dynamics simulation

3.1 Introduction

To replace traditional live attenuated strains of pathogens and inactivated killed pathogen vaccines thereby reducing safety risks, there has been a rapid development in advanced vaccine carriers, such as polymers, liposomes and nanoparticle proteins [1-6]. Wild-type heavy-chain ferritin comprises 24 identical subunits, each 21 kDa in size, forming a spherical hollow cage structure 12 nm in diameter [7]. The inner cavity of the spherical hollow cage structure has diameter 8 nm (Figure 3.1A, Figure 3.1B) [8]. Every ferritin subunit is made up of five helices, A to E, with a long BC loop and three short loops (AB, CD and DE) connecting other helices (Figure 3.1A) [9].

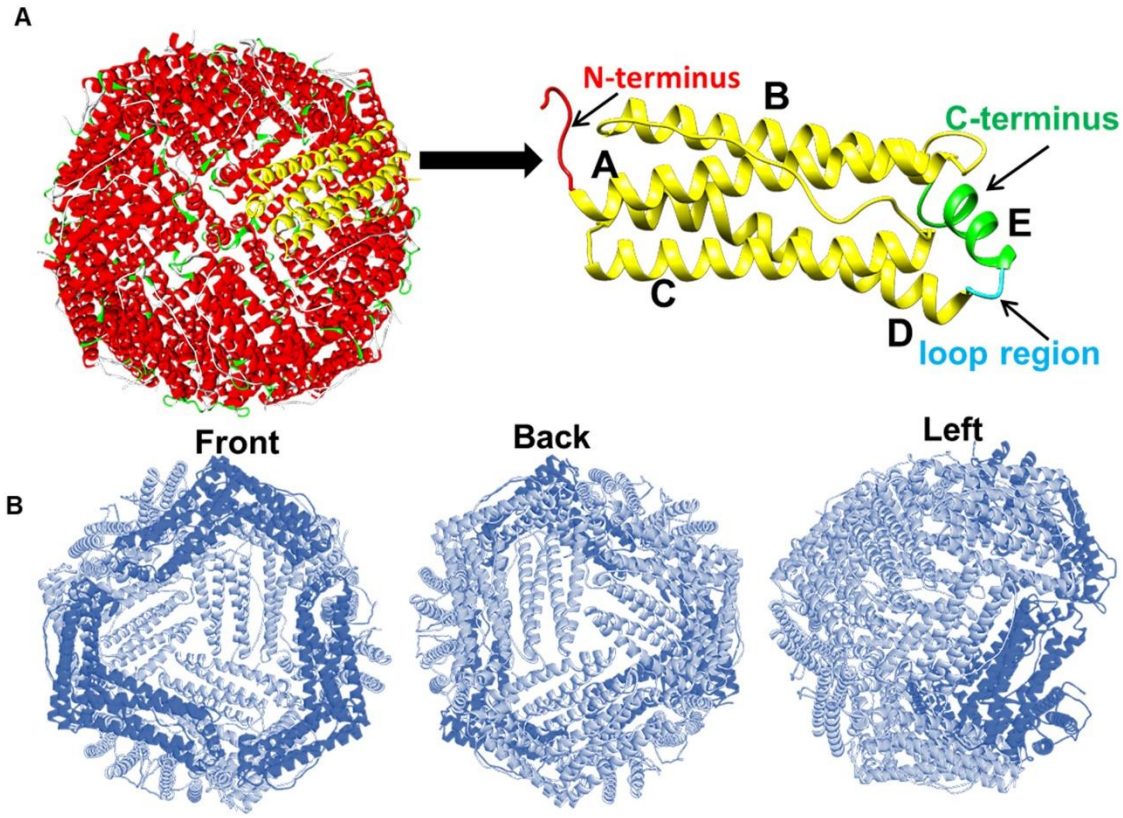


Figure 3.1 Human ferritin heavy chain mimic visualized with Discovery studio 2019 (DS). A) The native ferritin assembly structure and a monomer of native ferritin. Three insertion sites are indicated by different colours (Red: N-terminus; Green: C-terminus; Blue: flexible loop region). B) Ferritin hollow cage viewed by different angles.

Ferritin has considerable potential as a nano vaccine carrier because of its numerous advantages.

First and foremost, by genetic modification, ferritin can carry antigens and expose immunogens in a repetitive and well-organized manner. Ferritin can display exogenous peptides[10,11] and heterologous proteins, such as the influenza virus hemagglutinin (HA)[12] and native-like HIV-1 envelope glycoprotein trimers [13]. Both examples indicate that ferritin-based protein nanoparticles enhance the potency and breadth of virus vaccine immunity [10-13]. Secondly, ferritin is extremely thermally and chemically stable [5], and can be engineered for enhanced physicochemical and biological properties [14]. Lastly, due to its human source, human ferritin is considered safe with low toxicity, high biocompatibility and good biodegradability [15].

As a vaccine carrier displaying epitope candidates, ferritin has been investigated with three potential insertion sites, namely, the N-terminus, C-terminus and the DE loop region. The N-terminus of ferritin is connected to helix A and is located on the outside of the protein shell, whereas the C-terminus is located at the fifth short α-helix E on the inside of the hollow cage (Figure 3.1A) [9]. The native structure of ferritin is defined as a *flip* structure. However, if the exogenous peptide of a heterologous protein inserted at the C-terminus is very large and unable to be packed inside the ferritin cage, the helix E points outside creating a *flop* conformation [16]. It has been found that the volume inside the ferritin cage is approximately 2.1×10^5 Å^3 [16]. By assuming each amino acid has average volume of 150 Å^3, the ferritin could potentially contain a total of 1400 amino acids inside the ferritin hollow cage by inserting them at C-terminus. If there are more than 1400 amino acids are inserted at C-terminus, ferritin tends to become a *flop* structure. In this case, the inserted epitopes at C-terminus are presented outside. Otherwise, if the inserted epitope with less than 1400 amino acids at C-terminus, ferritin intends to be *flip* structure to present epitope inside of the ferritin cage. The *flip* or *flop* structure, may in turn play an important role in defining the quality and extent of the immune response.

There is limited knowledge about the connection between the different insertion sites and the corresponding protein properties. A recent study undertaken by Wang *et al.* displayed different lengths of epitopes from Enterovirus 71 (EV71) at the N- and C-termini and DE loop region (163 position) of ferritin to compare immune responses [10]. They found that peptides inserted into the DE loop region invoked a stronger immune response than at the N-terminus and C-terminus [10]. However, there are still knowledge gaps when comparing protein properties following N-terminus and C-terminus insertion, such as effects of hydrophobicity, thermal, pH- and chemical- stability.

The first important aspect to consider is protein hydrophobicity, which contributes significantly to protein folding and the stability of the structure [17]. A strategy to increase the stability of proteins is to decrease the hydrophobic solvent accessible surface area (SASA) [18]. The interaction of nonpolar residues with water causes the side chain to be more exposed to solvents for proteins with larger hydrophobic SASAs, hence this effect results in reduced stability [18]. Secondly, thermal stability is indispensable when developing a vaccine platform to ensure the bioactivity of the nanoparticle proteins. Most proteins are restricted to an allowed temperature range (4 °C - 60 °C) to maintain structural integrity [19]. One consequence of a temperature excursion out of this allowed range is the loss of protein function. By determining the thermal stability for ferritin and engineered ferritin with different insertion sites, our knowledge of structural diversity is conceivably expanded. Lastly, pH- and chemical stability are two other protein properties important to consider. The isoelectric point (pI), which is related to the pH stability, is the pH value at which the protein's net charge is zero. At the pI, protein solubility and electrical repulsion are lowest, resulting in highest tendency to aggregation and precipitation [20]. pH is also a key factor because it can cause the alterations in the structure of a protein, which affects bioactivity. Protein stability against chemical denaturants, such as

guanidinium chloride (GdnHCL), is helpful to understand protein structure differences [21]. GdnHCL acts as an ionic denaturant and is commonly used to probe the folding properties of proteins [22]. Accordingly, hydrophobicity, thermal-, pH- and chemical stability are deterministic of overall protein properties.

The experimental investigation of protein properties is a time-consuming process, especially when dealing with large numbers of samples during early vaccine development. Recently, molecular dynamics (MD) simulation has been widely applied to investigate important parameters related to protein stability [23-26]. The key factors related to protein stability include the hydrophobic SASA [27,28], root mean square deviation (RMSD) and radius of gyration (R_g). SASA could directly show the changes in the accessibility of protein to solvent [29,30], and it is related to protein stability as well. Lower SASA generally represents higher thermodynamic stability of protein [31-34]. RMSD, R_g and alpha helix content are factors to predict protein structure changes related to thermal change, along with presenting in-depth explanation for the transition [21,35,36]. RMSD is commonly used to reveal protein structural stability [37,38]. The higher RMSD is, the less stable the protein structure is suggested [29,30,39-42]. R_g can be regarded as the basic measurement of protein overall size [37,43]. A smaller R_g value indicates a more compacted protein structure and may result in high stability. Alpha helix content indicates protein secondary structure change at high temperatures. The helical content of proteins increases with increasing thermal stability [44,45]. MD simulation can provide reliable prediction of protein thermal related properties and stability, which is able to be applied in the design of vaccine candidates, to significantly reduce experiment related costs, timeframes and failure.

In this study, Epstein-Barr nuclear antigen 1 (EBNA1) was selected as the model epitope, as it shows high potential as a vaccine candidate against Epstein-Barr Virus (EBV). EBV is

responsible for many human diseases such as lymphoma and gastric cancer, and might enhance the development of several auto immune diseases [46]. EBNA1 is expressed in all EBV-associated tumours and it is an essential protein for EBV maintenance and expression. There are limited therapies for the prevention of EBV-associated diseases and no licensed vaccine is available to combat EBV infection [46]. EBNA1 determines the replication and partitioning of viral genomic DNA at latent viral infection. It is also the only viral protein found in all EBV-related malignancies [47]. Hence, EBNA1-associated vaccines will potentially act as prophylactic vaccines against all EBV strains.

By inserting the EBNA1 epitope at the N-terminus and the C-terminus of ferritin with fifteen residues of linkers between the epitope and the ferritin, this paper aimed to investigate the stability related properties of engineered ferritins. The four main properties we took into account were hydrophobicity, thermal stability, pH stability and chemical stability. We developed a new approach that combined the results from molecular dynamic simulation and experiments to improve stability predictions and the underlying mechanisms of designing ferritin-EBNA1 as an engineered nano vaccine.

3.2 Materials and methods

3.2.1 Protein Data Bank (PDB) file build-up

The PDB files of the Human heavy chain ferritin (PDB ID: 2FHA) and the Epstein-Barr virus nuclear antigen-1 (EBNA1) (PDB ID: 2FZ3) were generated from the RCSB Protein Databank. Discovery Studio (DS) was used to edit the epitope sequence for EBNA1. The model epitope was truncated off from EBNA1 comprising aa 407 to aa 417. A set of ferritin subunit structures without inserted epitope, ferritin monomer (F_1), ferritin trimer (F_3) and ferritin bioassembly structure (F_{24}), were prepared as the reference. By inserting one EBNA1 epitope at the N-terminus of the ferritin monomer structure with the linker, the PDB file named E1F1 monomer was prepared. The length of the linker between the ferritin and EBNA1 was assumed as 3 residues rather than 15 residues, as the longer linkers created complexity in computational simulation and resulted in a highly unstable system. For example, in Figure S3.1 (Supporting Information), E1F1 monomer with 3 residues of linker showed the most stable structure with lowest RMSD compared to 5 and 15 residues of linkers. In the same manner, a set of PDB files for ferritin with inserted EBNA1 epitope at C-terminus with 3 residues of linker were prepared as shown in Table S3.1. Based on theoretical calculation, epitopes and linkers with a total of 624 amino acids, could fit the ferritin cage completely. Hence, F1E1 (C-terminus insertion) was assumed to have ferritin *flip* structure, where the inserted epitopes were contained inside of the protein cage, shown in Figure 3.2.

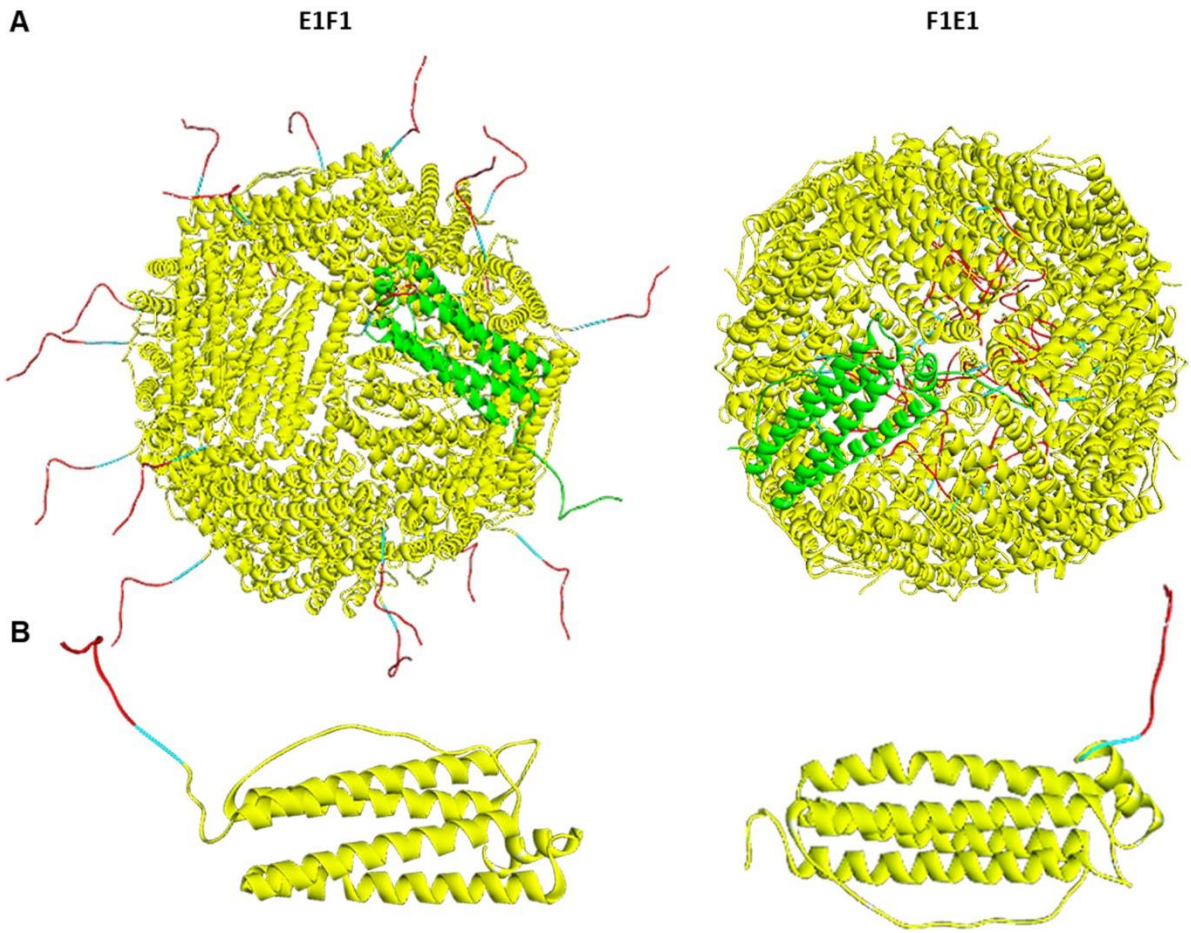


Figure 3.2 Engineered ferritins E1F1 and F1E1 mimic diagrams using DS 2019. Inserted epitopes are highlighted in red and linkers are highlighted in blue. A) Assembly structures of E1F1 and F1E1. A monomer of E1F1 and F1E1 are denoted as green. B) Monomer structure of E1F1 and F1E1. The epitope (red) and linker (blue) are inserted at N-terminus of ferritin monomer for E1F1, while they are inserted at C-terminus of ferritin monomer for F1E1.

3.2.2 Molecular Dynamics (MD) Simulation of ferritin structures and ferritin-EBNA1 structures

GROMACS 2018.3 was used to perform MD simulations of the ferritin subunit structures and ferritin-EBNA1 structures in a water environment [48,49]. The OPLS-AA/L all-atom force field was used to generate topology files [49]. Each protein structure was solvated in a cubic box using the TIP3P water model by specifying spc216.gro in GRAOMCS [50]. The dimensions of the cubic box for each structure are summarised in Table S3.2. The minimum distance between the protein and the wall of the unit cell was set to be 1.0 nm. The protein/ion/water system was

first energy minimized and then equilibrated in a 100-picosecond MD simulation with positional restraints on the heavy atoms of the protein. The equilibration was performed at a temperature of 300 K and a pressure of 1 bar by coupling to an external heat source and an isotropic pressure bath. The position restraints were then released and the MD simulation was produced in 200 nanoseconds and the snapshots were stored per 100 picoseconds for the MD trajectory analysis.

For hydrophobicity study by simulation, bovine serum albumin (BSA) (PDB ID: 3V03) was simulated at the same condition. Hydrophobic SASA was determined by using the *gmx sasa* tool from GROMACS based on the completed MD simulation results for BSA, ferritin, E1F1 (N-terminus insertion) and F1E1 (C-terminus insertion) in water at 25 °C. The Error analysis on hydrophobic SASA was calculated using block averaging (*gmx analyze –ee*).

Thermal stability studies for the monomer and trimer structures of ferritin, EBNA1-ferritin and ferritin-EBNA1 followed similar procedures as above but the MD simulation was produced in 100 nanoseconds over a temperature range from 25 to 90 °C. The snapshots were stored per 50 picosecond for the MD trajectory analysis. RMSD and R_g of C_α were determined using *gmx rmsd* tool and *gmx gyrate* tool from GROMACS. Helix content was calculated using the *gmx dssp* tool from GROMACS. For bio-assembly structures of E1F1 and F1E1, MD simulations were produced in 200 nanoseconds at 150 °C. The snapshots were stored per 100 picoseconds for the MD trajectory analysis. RMSD and helix content were determined.

3.2.3 Plasmid construction of recombinant ferritin complexes

PET 30a vector was used to construct plasmids harbouring genes of two types of Epstein-Barr virus nuclear antigen-1 (EBNA1) fused ferritins (ferritin-EBNA1 and EBNA1-ferritin) and the control ferritin. Target genes were inserted between *BamHI* and *XhoI* restriction sites. The epitope, EBNA1 (HPVGEADYFEY), in two ferritin complexes, is the main immunogenic domain of EBNA.[51] This linear epitope is known to be potent natural targets for cytotoxic T lymphocytes (CTLs) [52,53]. A fifteen-residue linker (GGSGGGGSGGGGSGG) was inserted between the epitope and ferritin in both ferritin complexes. The amino acid sequences information for ferritin and recombinant proteins ferritin-EBNA1 are shown in Table S3.3.

3.2.4 Protein expression and purification

The plasmids were transformed into *E. coli* BL21 (DE3). The expression process of the vector ferritin and the two engineered ferritins were the same. A single bacterial colony expressing each protein, namely, ferritin, EBNA1-ferritin (E1F1) and ferritin-EBNA1 (F1E1), was inoculated in 50 mL LB medium supplemented with kanamycin (100 μg mL^{-1}), and incubated overnight at 37 °C in an orbital mixer incubator (Ratek), shaken at 180 rpm. 10 mL of bacterial suspension was then transferred to 500 mL LB with kanamycin (100 μg/ml) medium. The culture was grown at 37 °C and shaken at 200 rpm until OD_{600} reached 0.8, followed by induction with 1 mM IPTG at 37 °C, 200 rpm for 4 h. Cells were collected after centrifugation at 13,751 xg. Cell pellets were resuspended in lysis buffer and disrupted by sonication (Scientx sonicator at 360 W: 4 seconds on and 6 seconds off for 10 min) with lysis buffer (20 mM PB, 2 mM EDTA, pH 7.0). Bacterial lysate was centrifuged at 13,751 xg for 30 min at 4 °C to remove cell debris and supernatant was harvested for further protein purification.

The protein purification processes for ferritin, EBNA1-ferritin and ferritin-EBNA1 were the

same, except they differed in the used chromatographic resin. First, clarified supernatant was diluted with Milli-Q water to a final protein concentration of 4 mg mL^{-1}. NaCl and NaAc-HAc were added to a final concentration of 1 M and 100mM respectively and the pH was adjusted to 5.0 and mixtures then underwent thermal precipitation at 60 °C for 10 min. Centrifugation (19,802 xg at 4 °C for 10 min) was conducted to remove precipitated impurities. Subsequently, resultant supernatant pH was adjusted to 6.5 by using 1 M Phosphate buffer, pH 8.0, and 1 M Ammonium Sulfate (AS) was added for the following hydrophobic interaction chromatography (HIC). In HIC, a HiTrap Octyl FF column (GE Healthcare, USA) was used for ferritin and E1F1; a HiTrap Butyl FF column (GE Healthcare, USA) was used for F1E1. The equilibrated buffer was 100 mM Phosphate buffer, 1 M AS, pH 6.5. The eluted buffer was 20 mM Phosphate buffer, pH 6.5. HIC was performed on an AKTA pure (GE Healthcare, USA). Briefly, samples were loaded onto columns equilibrated with equilibration buffer. After washing the column for 5 CV with loading buffer, a linear gradient elution of 0 to 100 % elution buffer in 3 column volumes was conducted and eluted peaks were pooled together. Absorbance at 280 and 260 nm were recorded throughout chromatography. Finally, collected fractions from HIC were buffer exchanged to 20 mM PB, pH 7 using a HiTrap Desalting column (GE Healthcare, USA). Protein expression and purities were analysed by SDS-PAGE.

3.2.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein expression and purification were performed with 12 % (w/v) SDS separation gel having a 5 % (w/v) SDS stacking layer. Samples were prepared with 5× loading buffer (containing 6 % (v/v) 1 M Tris-HCL at pH 6.8, 25 % (v/v) glycerol, 1 mg mL^{-1} bromophenol blue, 2 % (v/v) SDS and 5 %v/v β-mercaptoethanol). Treated samples were heated at 100 °C for 10 min and loaded into the loading wells. Separation was conducted at 200V for 60min. The gel was

stained with 0.25 % (w/v) Coomassie R-250 for 60 min and the background was washed using a washing buffer (10 % (v/v) ethanol and 10 % (v/v) acetic acid) for 270 min.

3.2.6 Transmission electron microscopy (TEM)

Purified ferritin and engineered ferritins were diluted to a protein concentration of 0.1 mg mL^{-1}. Carbon-coated copper grids were incubated face down on 10 µL of the sample for 10 min. The grid was washed 3× in droplets of Milli-Q water, and then negatively stained in 2 % (v/v) uranyl acetate for 2 min. Excess liquid was removed with filter paper. Grids were analysed with a Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

3.2.7 Size exclusion chromatography coupled with multiple- angle laser light scattering (SEC-MALS)

SEC-MALS was performed to determine hydrodynamic radius and molecular weight of purified and engineered ferritins. Superose 6 10/300 GL (GE Healthcare, USA) was connected to High Performance Liquid Chromatography (HPLC) (Shimazdu, Japan) coupled with Wyatt Optilab refractive index (RI) and Wyatt DAWN MALLS detector (USA). The equilibrated buffer was 20 mM Phosphate buffer, pH 7.0. The protein concentration was 5 mg mL^{-1}. Loading volume was 50 µL.

3.2.8 Hydrophobicity Study

Hydrophobicity study was performed to compare the hydrophobicity differences amongst ferritin, E1F1 and F1E1. Bovine serum albumin (BSA) was used as a control sample. 1 mL of each protein sample (BSA, ferritin, E1F1 and F1E1) was adjusted to 2 mg mL^{-1}. Each protein sample was mixed with 4 mL of equilibrated buffer (1.0 M ammonium sulfate buffer, 100mM

Phosphate buffer containing Na_2HPO_4 and NaH_2PO_4, pH 6.5). A HiTrap Butyl FF column (GE Healthcare, USA) was used to perform the hydrophobicity study. The elution buffer was 20 mM Phosphate buffer, pH 6.5. After the column was equilibrated, protein samples were loaded to a HiTrap Butyl FF column individually. The column was then eluted with an elution buffer by a linear gradient elution of 0 to 100 % elution buffer in 3 column volumes. The retention time for each protein sample was recorded by AKTA pure (GE Healthcare, USA).

3.2.9 Thermal stability study by differential scanning calorimetry (DSC)

Microcal VP-DSC (Malvern, UK) was used to determine the thermal stability of proteins. Samples were prepared at protein concentration 1 mg mL^{-1} in 20 mM Phosphate buffer (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. For DSC measurement, a scan rate of 90 °C per hour was used for a temperature ramp from 30 to 120 °C, and samples were analysed using the software package Origin 9 (OriginLab Corporation, USA).

3.2.10 Thermal, pH and chemical denaturants stability study by circular dichroism (CD)

CD Chirascan V100 (Applied Photophysics, UK) was used to analyse the secondary structure of the protein samples (0.3 mg mL^{-1} in 20 mM Phosphate buffer, pH 9.0). Each sample was scanned three times. For thermal stability study, samples were equilibrated at 25 °C, 70 °C, 80 or 90 °C for 10 min, respectively. The signal from 190 nm to 260 nm was scanned. CD signals at a wavelength of 222 nm were recorded as a measurement of helix content. For pH stability study, protein samples (1 mg mL^{-1} in 20 mM Phosphate buffer, pH 7.0) were adjusted from pH 7.0 to pH 4.0 by adding acid (1.0 M Acetic acid). Samples were equilibrated at each pH condition for 2 min. The CD signal from 190 nm to 260 nm was scanned. For chemical denaturant stability studies, protein samples (1 mg mL^{-1} in 20 mM Phosphate buffer, pH 7.0)

were mixed with guanidinium chloride (GdnHCl) solution (6 M) and a 20 mM Phosphate buffer (pH 7.0) to reach final concentrations of 0, 0.5, 1.0, 1.5, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 4.0, 4.5 and 5M GdmCl. Mixed solutions were equilibrated for 2 min, 10 min or 20 min to measure the CD of 222 nm and 20 °C for 300 s.

3.3 Results and discussion

3.3.1 Characterization of engineered ferritins

The molecular weight for EBNA1-ferritin (E1F1) and ferritin-EBNA1 (F1E1) subunits measured by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were both approximately 23 kDa as expected based on amino acid sequence (23.35 kDa) *by ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss)*, higher than for ferritin (21 kDa) due to inserted epitope and linker (Figure 3.3A). Hydrodynamic diameters for engineered ferritins E1F1 and F1E1 measured by size exclusion chromatography with multiple-angle laser light scattering (SEC-MALS) were 15.2 nm and 13.3 nm, respectively, larger than native ferritin at 13.2 nm (Figure 3.3B). In SEC experiments, E1F1 showed a smaller elution volume of 12.9 mL compared to F1E1 (13.9 mL) (Figures 3.3C and 3.3D), consistent with the hydrodynamic diameter results by Dynamic light scattering (DLS) (E1F1: 15.3 nm; F1E1: 13.5 nm). These results proved that both E1F1 and F1E1 presented *flip* structures. For N-terminus insertion, epitopes were exposed outside the structure, while for C-terminus insertion, epitopes were located inside the ferritin cage. Transmission electron microscopy (TEM) results did not reveal any differences regarding the particle structures for E1F1 and F1E1. Both were characterized as well-assemble particles having monodispersed structure (Figures 3.3E and 3.3F).

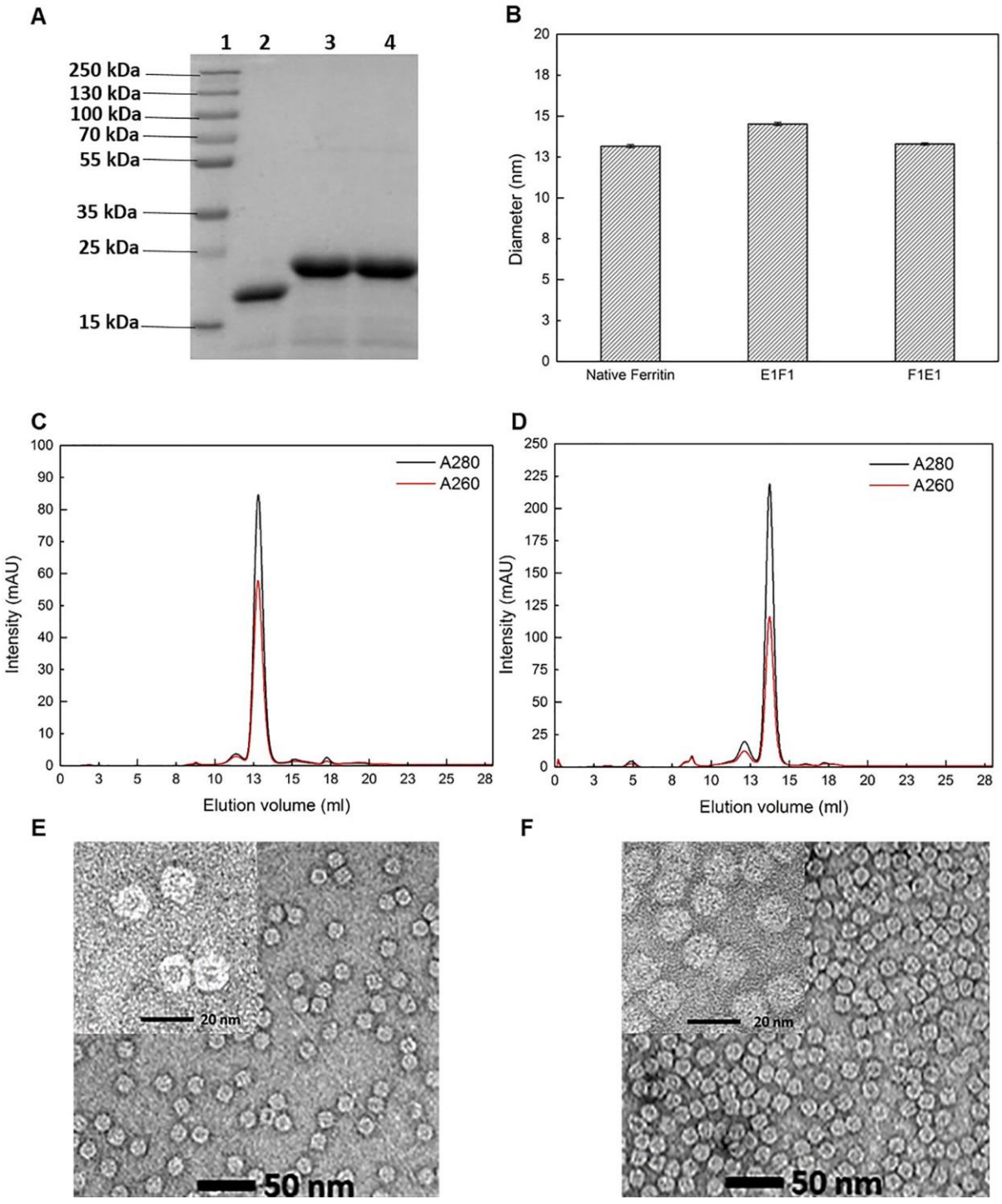


Figure 3.3 Characterization of engineered ferritins. A) Coomassie blue stained sodium dodecyl sulfate PAGE: lane 1, protein marker; lane 2, native ferritin; lane 3, E1F1; lane 4, F1E1. B) MALS evaluation of ferritin, E1F1 and F1E1. C) Size- exclusion (Superose 6) chromatogram for E1F1. D) Size-exclusion (Superose 6) chromatogram for F1E1. E) TEM image of E1F1. F) TEM image of F1E1.

3.3.2 Hydrophobicity study by simulation and experiment

To understand the impact of insertion sites on the properties of the engineered ferritin, hydrophobicity was compared between E1F1 and F1E1. Hydrophobicity was firstly examined by computational simulation. Simulation results for hydrophobic SASA shown in Table 3.1 indicated that hydrophobic SASA was considerably related to protein molecular weight.

Table 3.1 Hydrophobicity- related parameters from experiments and simulations. The PDB files for E1F1 and F1E1 were structured from the Discovery Studio.

Molecule	PDB ID	Molecular weight (kDa)	Retention time by experiment (min)	Average hydrophobic SASA by simulation (nm^2)
BSA	3V03	66.0	20.3	283.9 ± 0.19
Native ferritin	2FHA	504	24.5	681.9 ± 0.37
E1F1	-	560.4	29.6	866.3 ± 0.29
F1E1	-	560.4	26.0	827.5 ± 0.25

BSA showed the least hydrophobic SASA (BSA: 284 nm^2; Ferritin: 682 nm^2; E1F1: 866 nm^2; F1E1: 828 nm^2). E1F1 (N-terminus insertion) was the most hydrophobic of the tested proteins as its hydrophobic SASA was largest (Figure 3.4B). Simulations also showed that the hydrophobic SASA of F1E1 was smaller than for E1F1. Experimentally, four proteins (BSA, ferritin, E1F1 and F1E1) were loaded to a HiTrap Butyl FF column individually, and then eluted under the same conditions. Longer retention time indicates a stronger binding with resin therefore higher protein hydrophobicity. As shown in Figure 3.4A, E1F1 (N-terminus insertion) showed the highest hydrophobicity, with retention time at 29.6 min followed by

F1E1 (C-terminus insertion) at 26.0 min (Table 3.1). The addition of EBNA1 antigens to the termini of ferritin increased total hydrophobicity.

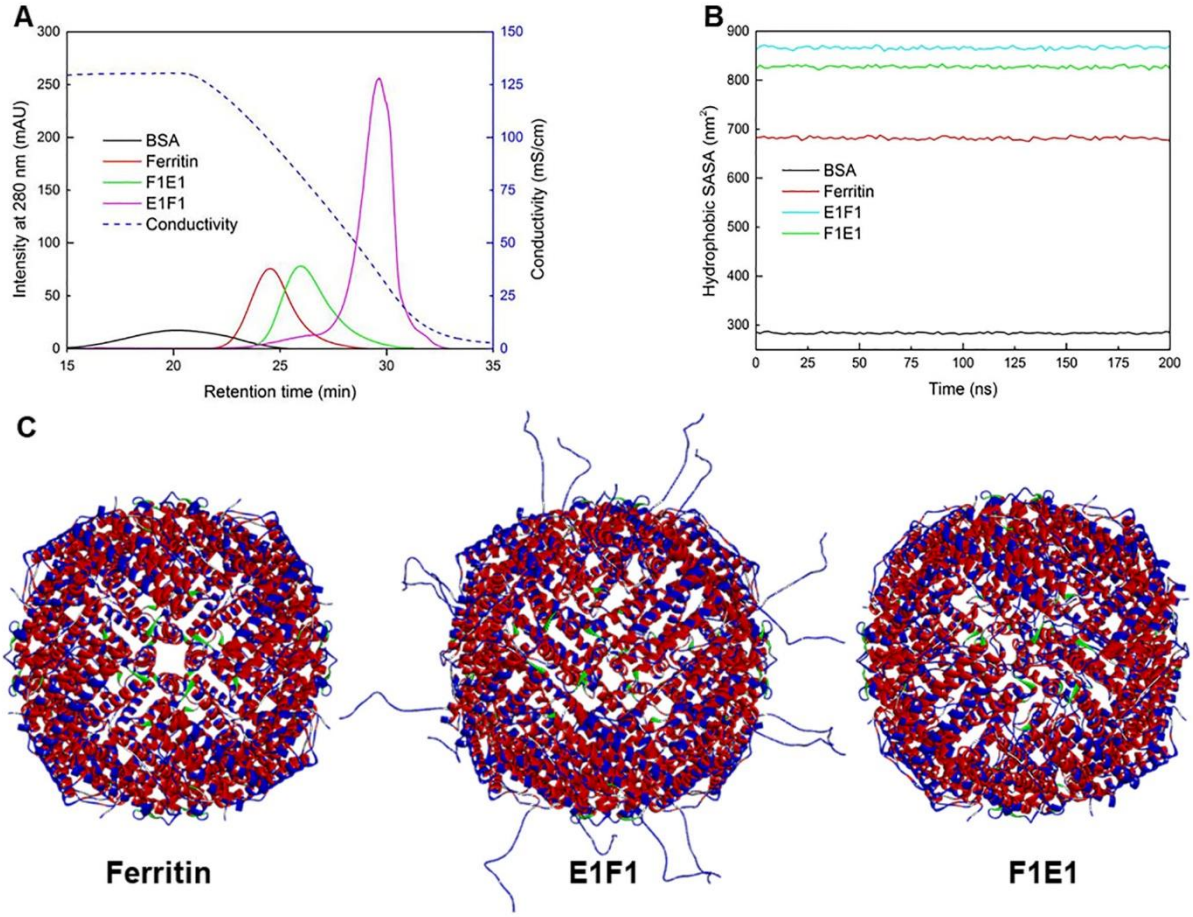


Figure 3.4 Hydrophobicity of engineered and eluted native ferritins. A) Bovine serum albumin (black), native ferritin (red), E1F1 (magenta) and F1E1 (green) from a HiTrap Butyl FF hydrophobic interaction column. B) Computed hydrophobic SASA for BSA (black), ferritin (red), E1F1 (blue) and F1E1 (green) against time at 25 °C using GROMACS. C) Hydrophobicity diagram by Discovery Studio (DS) 2019: hydrophobic (red); hydrophilic (blue).

Simulation results for hydrophobic SASA accorded well with experimental results. It was found that the addition of epitopes increased the hydrophobicity of engineered ferritins compared to native ferritin. This is due to inserted epitopes and linkers consisting of hydrophobic amino acids Proline (P), Valine (V), Glycine (G), Alanine (A), and Phenylalanine (F). Both simulation and experiment revealed that E1F1 (N-terminus) was the most hydrophobic of the tested proteins. Although the theoretical molecular weights for E1F1 and

F1E1 were same, N and C-terminus insertions impact differently with regard to hydrophobicity, as the space allocation of same epitopes are quite differently. As it has been mentioned above, both E1F1 and F1E1 presented in ferritin *flip* structures, where N-terminus insertion locates epitopes outside the ferritin cage and C-terminus insertion locates epitopes inside the ferritin cage. On this account, N-terminus insertion was more hydrophobic than C-terminus insertion. These results also confirmed that computational determination of protein hydrophobic SASA predicted well the practical protein hydrophobicity.

3.3.3 Thermal stability study by simulation and experiment

Thermally stable proteins are expected to have low alpha helix content loss [21], less fluctuating regions [24,42,54] and are expected to be more compact [24]. The amount of alpha helix content loss reveals protein secondary structure changes. Less alpha helix content loss suggests a more stable secondary structure [21,25]. Lower RMSD indicates a more thermally stable protein structure [24,54]. Lower R_g shows a more compact protein structure and higher stability [24]. Simulations of the monomer of ferritin, E1F1 and F1E1 at defined temperatures (25 °C, 70 °C, 80 and 90 °C) were run to investigate the difference in thermal stability by analysis of alpha helix content, RMSD and R_g of the central carbon atom (Cα). Ferritin monomer presented the highest stability over all the tested temperatures. After 100 ns, the helical content reduction at 90 °C was minimal compared with that at 25 °C (Figure 3.5A). High temperature impacted less on the E1F1 monomer (N-terminus insertion) in terms of helical content change; E1F1 monomer only lost 4 % helical content at 90 °C compared to F1E1 monomer (23 % helical content loss) (Table S3.5, Supporting Information). The RMSD results highlight that the stability of F1E1 monomer fluctuated more wildly with increasing temperature than E1F1 monomer (Figure S3.4A2 and Figure S3.4A3, Supporting Information). The R_g results for

E1F1 and F1E1 monomers were similar except at 90 °C, where F1E1 monomer started to be less compact than E1F1 monomer (Figure S3.4B2 and Figure S3.4B3, Supporting Information). Simulation results suggest that N-terminus insertion of monomer was more stable than C-terminus insertion at 90 °C.

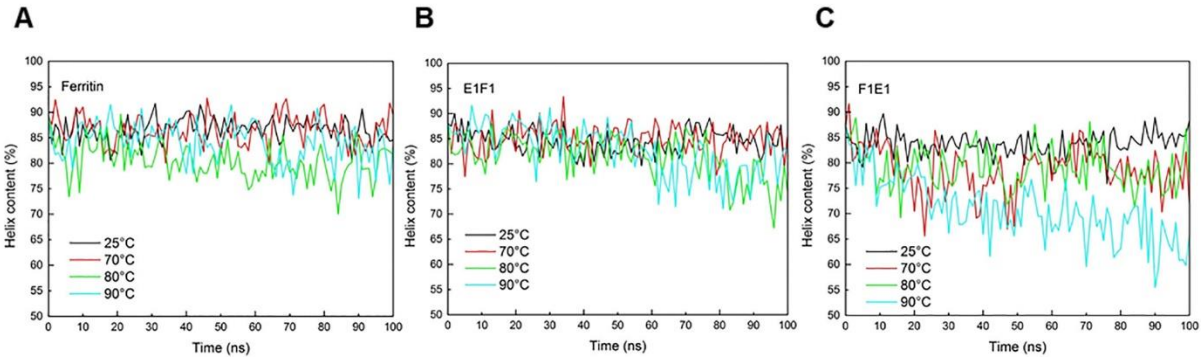


Figure 3.5 Alpha helix content for monomers of ferritin, E1F1 and F1E1 using GROMACS over a temperature range from 25 to 90°C. A) One subunit of ferritin. B) One subunit of E1F1. C) One subunit of F1E1.

To consider interactions amongst each monomer, trimer structures for ferritin-associated proteins were also simulated. Helical content decreased due to increasing temperature and displayed a similar trend for all three trimer structures, ferritin, E1F1 and F1E1 (Figure 3.6). Unlike the trend for the monomer structure of F1E1, the trimer structure enhanced the protein secondary structure's ability to resist high temperatures. There was minimal helical content decrease for the F1E1 trimer compared with the F1E1 monomer at high temperatures up to 90°C. It was also observed that, at 90 °C, F1E1 trimer had less helical content loss than the E1F1 trimer, 3 % compared to 10 % (Table S3.6, Supporting Information). This result indicated that high temperature (90 °C) affected less on secondary structure of the F1E1 trimer (C-terminus insertion) compared to the E1F1 trimer (N-terminus insertion). It was also found that, except for ferritin, RMSD variance for trimer structures was similar to monomer structures (Figure S3.6 and Figure S3.2). That suggests, under all tested temperatures for 100 ns, there were measurable thermal effects on secondary structures of E1F1 and F1E1. However,

regarding to structural stability in terms of RMSD, the comparisons between E1F1 and F1E1 were not obvious.

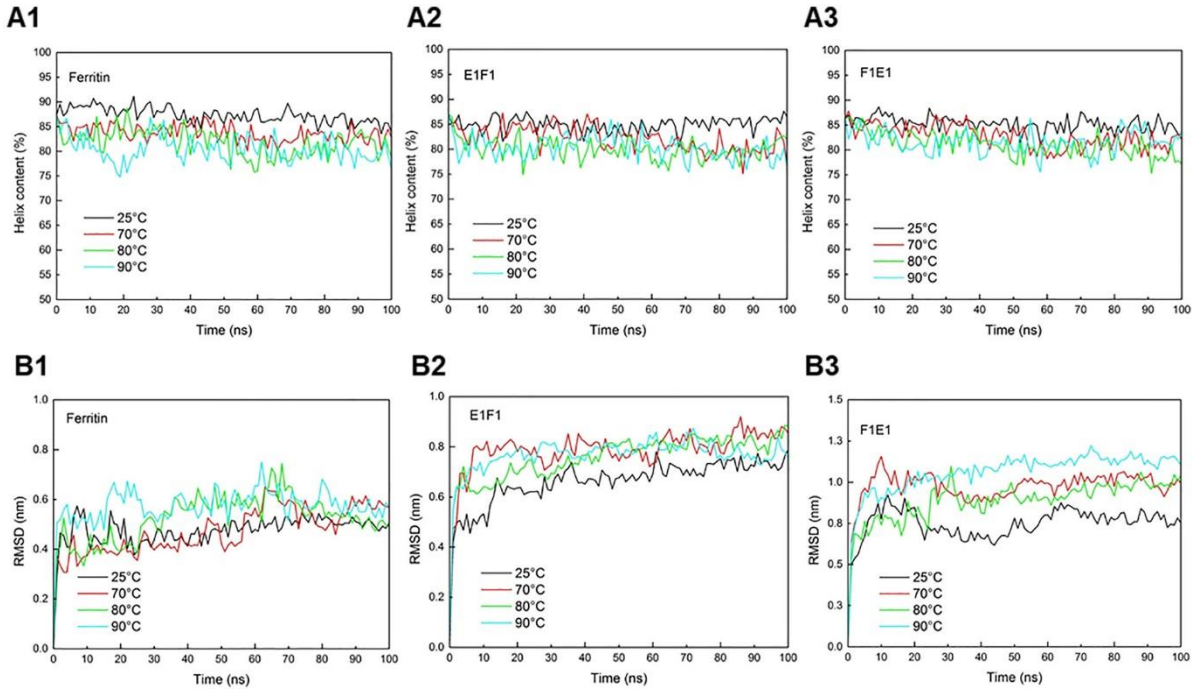


Figure 3.6 Alpha helix content and RMSD for trimer structures of ferritin, E1F1 and F1E1 using GROMACS over a temperature range of 25 to 90°C. A1) Ferritin alpha helix content. A2) E1F1 alpha helix content. A3) F1E1 alpha helix content. B1) Ferritin RMSD. B2) E1F1 RMSD. B3) RMSD.

To support a better understanding of thermal effect on protein assembly structures and a more observable thermal stability comparisons, 24 subunits of ferritin, E1F1 and F1E1 were simulated at a relatively high temperature of 150 °C, well above the reported denaturation temperature of ferritin at 102 °C [55]. Due to computation capacity limitations, an extreme condition was proposed as 150 °C for 200 ns [21,56]. Helical content losses for 24 subunits of ferritin, E1F1 and F1E1 were 21, 16, and 21 % respectively (Figure 3.7A). The simulation results also indicated that the three assembly proteins were still integrated after 200 ns. Comparing with trimer structures, thermal stabilities for all three assembly proteins were improved, as the secondary bonds in all three proteins survived the thermal effect to preserve protein tertiary structures. The assembly of protein subunits is driven by secondary bonds, such

as noncovalent interactions, including hydrogen bonding, electrostatic association, hydrophobic effects and van der Walls forces [57,58]. It was also observed that thermal stabilities for each protein were different at 150 °C regarding to their backbone RMSD. Figure 3.7B reveals that ferritin was the most thermally stable over 200 ns with lowest RMSD, and F1E1 was found to be more thermally stable than E1F1, indicated by a lower RMSD obtained. RMSD related to ferritin core part without consideration of linkers and epitopes was also calculated to determine how fusion proteins E1F1 and F1E1 were perturbed compared to wild type ferritin (Figure S3.3, Supporting Information). It showed that RMSD for core parts of E1F1 and F1E1 were similar but higher than that of type ferritin. It demonstrates the less stable core parts of E1F1 and F1E1 than wild type ferritin. R_g results (Figure 3.7C) also indicated F1E1 had a more compact structure than E1F1 at 150 °C. Computed hydrophobic SASA for each protein decreased against time (Figure 3.7D). That indicates, since simulation started, less hydrophobic surface area was exposed to water. Hydrophobic SASA for E1F1 was found to be lost most as 37 %. To conclude, MD simulation predicted native ferritin presents the best thermal stability, while C-terminus insertion to ferritin showed better thermal stability against high temperature up to 150 °C than N-terminus insertion.

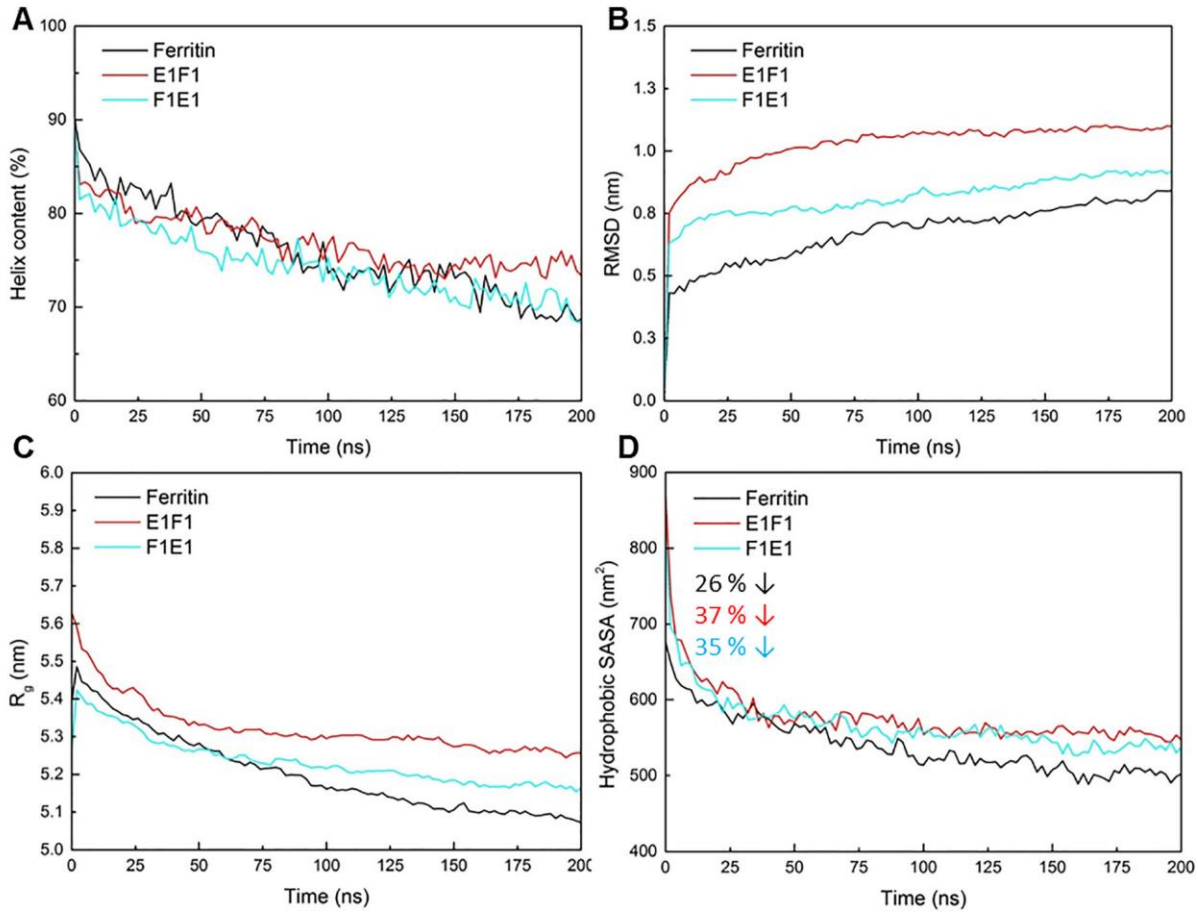
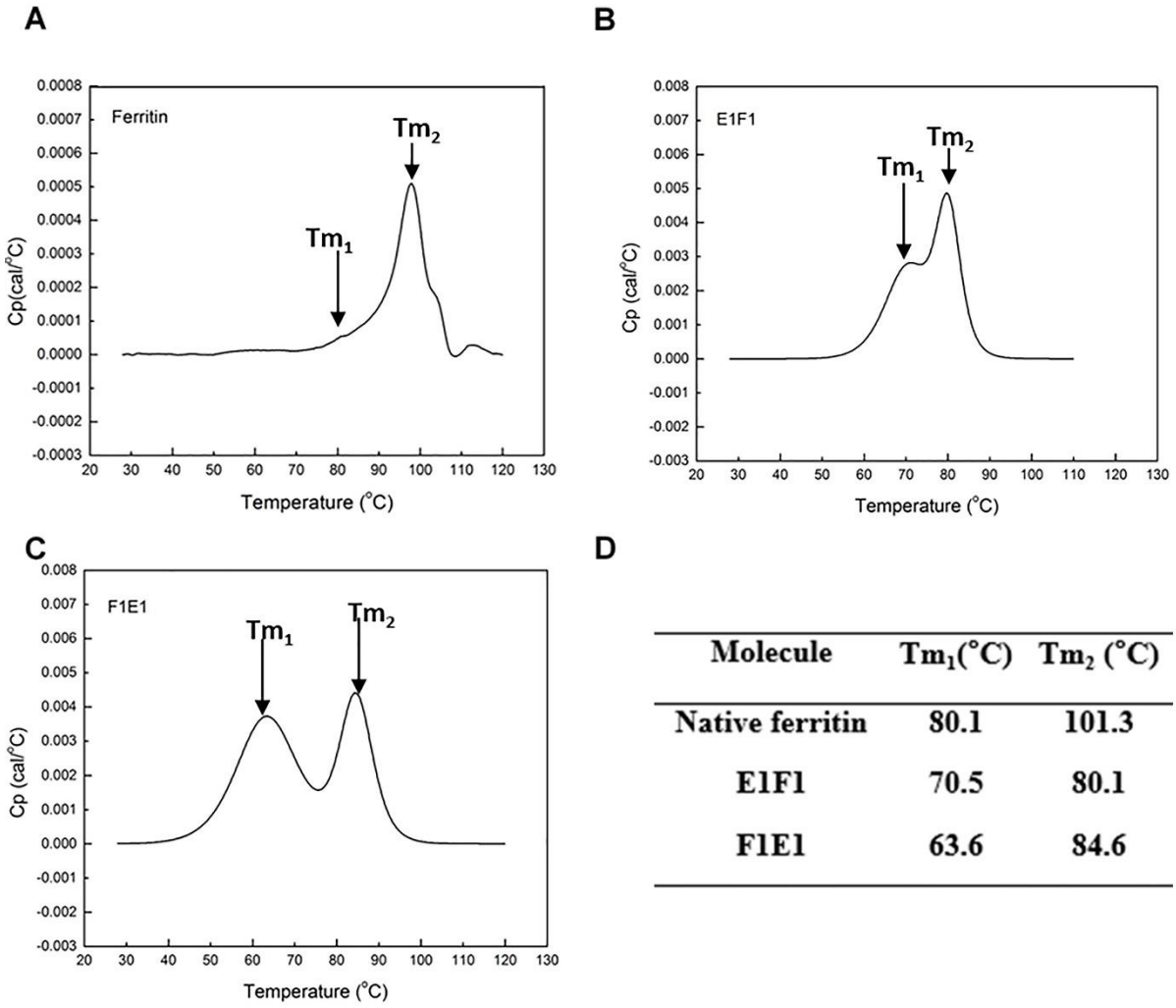


Figure 3.7 Simulation data for assembled structures of ferritin, E1F1 and F1E1 using GROMACS at 150°C for 200 ns. A) Alpha helix content. B) RMSD. C) Rg. D) Hydrophobic SASA; Ferritin dropped 26 %, E1F1 dropped 37 % and F1E1 dropped 35 %, compared with initial structure.

To validate this simulation prediction, thermal stability was characterized by experiment using differential scanning calorimetry (DSC). As shown in Figure 3.8A, the first peak (T_{m1}) suggested where the protein structure began to partially change but will integrate structure remaining. After the second peak (T_{m2}), protein structures were completely destroyed. *T*m1 range for three proteins was around 63.6 to 80.1 °C (Figures 3.8A, 3.8B and 3.8C). Native ferritin had robust thermal stability, and the structure was completely denatured above 100 °C (Figure 3.8A). In comparison with native ferritin, adding epitopes decreased the maximum tolerated temperature. For both E1F1 and F1E1, the proteins were completely disassembled below 90 °C (Figure 3.8B, Figure 3.8C). The denaturation temperature for C-terminus insertion

was slightly higher than for N-terminus insertion (T_{m2} values 85 and 80 °C, respectively). The experimental results are well aligned with the MS simulation result; C-terminus insertion showed better thermal stability than N-terminus insertion.



Molecule	Tm_1(°C)	Tm_2 (°C)
Native ferritin	80.1	101.3
E1F1	70.5	80.1
F1E1	63.6	84.6

Figure 3.8 Differential scanning calorimetry (DSC) results for ferritin, E1F1 and F1E1 (Protein concentration: 1 mg mL^{-1}; Sample buffer: 20 mM Phosphate buffer pH 7.0). A) Native ferritin. B) E1F1. C) F1E1. D) Tm1 and Tm2 values for native ferritin, E1F1 and F1E1 from DSC.

For detailed investigation regarding protein secondary structure changes, four temperatures (25, 70, 80 and 90 °C) were selected based on DSC results. CD showed the helical content of proteins was connected to the secondary structural changes of ferritin, E1F1 and F1E1, at defined temperatures (Figure S3.4, Supporting Information). Both E1F1 and F1E1 showed less

tolerant thermal stabilities than native ferritin. The overall decrease of the secondary structure of helix content for ferritin is 15 % at 90 °C, while E1F1 and F1E1 lost 25 % and 29 % helix content, respectively (Table S3.4, Supporting Information). N-terminus insertion displayed preserved helicity until 80 °C. After 80 °C, there was a dramatic helical content reduction. In contrast, C-terminus insertion appeared to have weaker helicity above 70 °C. All of these results from CD are consistent with the results from DSC and indicate the engineered ferritin is slightly less thermally tolerant than native ferritin for temperatures up to 70°C.

Overall, both molecular simulation and experimental results clearly confirmed a change of insertion site on an engineered ferritin and affected significantly on the thermal stability of the protein. Based on the simulation results, E1F1 monomer showed similar thermal property to native ferritin monomer in terms of helical content and RMSD value up to 90 °C. In comparison, F1E1 monomer showed less thermal stability and major helix content loss. However, the thermal stability for the trimer structure of F1E1 was enhanced. At temperatures in the range 25 to 90 °C, the trimer structures of the native ferritin, E1F1 and F1E1 all presented similar trends regarding helix content change. It was found that the assembly structure of protein could significantly increase resistance to high temperature. Engineered ferritins and native ferritin could still retain their structures at 150 °C for 200 ns in simulation. Considering only terminus insertions, E1F1 showed slightly stronger helicity than F1E1 at 150 °C for 200 ns, which was consistent with CD results. On the other hand, F1E1 showed lower fluctuated regions and better structure stability than E1F1. This could explain tertiary structure of F1E1 presented better thermal stability in DSC results. By comparing the hydrophobic SASA for these two proteins, E1F1 had a larger hydrophobic SASA than F1E1, which means more hydrophobic side chains are exposed to water and an unfolding process is favoured therefore

resulting in weaker thermal stability. Hence, C-terminus insertion was more thermally stable than N-terminus insertion when considering the assembled structure**.**

3.3.4 Effect of pH change on engineered ferritins

To investigate the effect of pH on engineered ferritins, equal amounts of purified E1F1 and F1E1 were adjusted to pH 4, 5, 6 and 7, and all samples were inculcated in CD for 5 min. By observing the band at wavelengths 210 to 230 nm, F1E1 showed obvious shift only at pH 5, whereas E1F1 has two obvious curve shifts at pH 4 and pH 5 (Figure 3.9A and Figure 3.9B). The theoretical pI for E1F1 and F1E1 is approximately 5.13, therefore, at pH 5, both E1F1 and F1E1 secondary structures shifted because of the protein isoelectric point (pI) effect. At pH equal to pI, purified proteins have zero net charge and the attraction force becomes dominant, resulting in proteins that are more likely to aggregate and precipitate [59]. Nevertheless, F1E1 was more tolerant to a lower pH than E1F1 as the secondary structure of E1F1 was completely changed from pH 5, while F1E1 was able to remain the secondary structure at pH 4. For N-terminus insertion, epitopes are located on the surface of ferritin particle. The aggregation of epitopes is possible for E1F1 at pH 4. C-terminus insertion buries the epitopes inside the ferritin cage and therefore protects them from outside low pH conditions. Consequently, C-terminus insertion is expected to be more tolerant to a lower pH than N-terminus insertion.

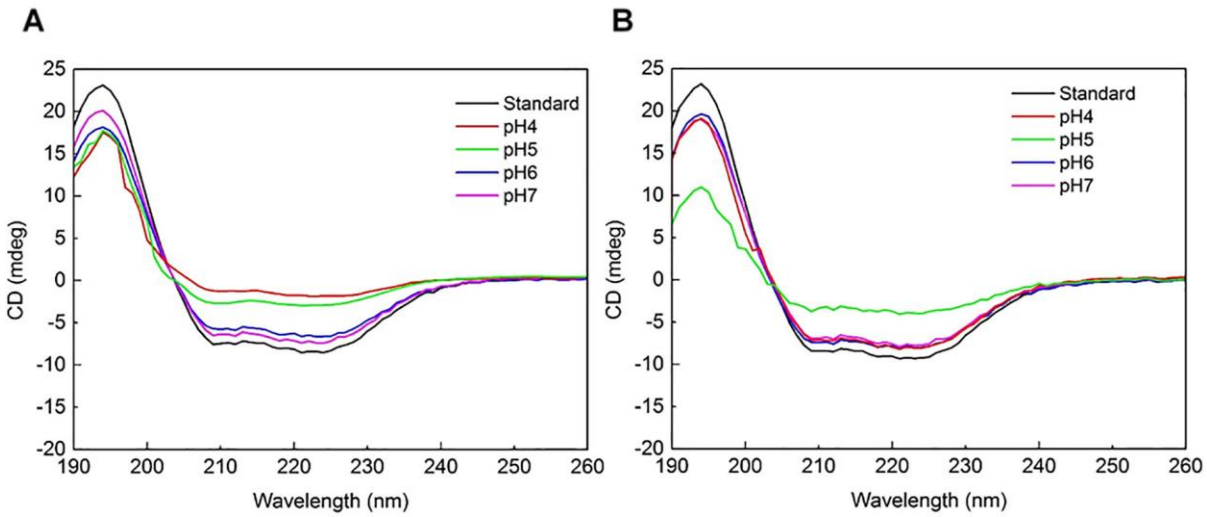


Figure 3.9 Circular dichroism spectra of E1F1 (A) and F1E1 (B) (Sample concentration: 1 mg mL^{-1}; Sample buffer: 20 mM Phosphate buffer, pH 7.0). The spectra range shows the secondary structures of E1F1 are completely changed at pH 5. F1E1 is stable over the tested pH, except at pH 5. Standard curve represents the signal measurement for tested proteins without pH adjustments.

3.3.5 Effect of chemical denaturants on engineered ferritins

To test the stability of E1F1 and F1E1 against chemical denaturation, CD spectra at wavelength 222 nm were measured with varying guanidinium chloride (GdnHCl)-concentrations. The samples of all 3 proteins were equilibrated with GdnHCl solution for 20 min. Native ferritin was used as the reference. Native ferritin was highly chemically stable. Ferritin lost 50 % of its native structure at 5 M GdnHCl (Figure 3.10).

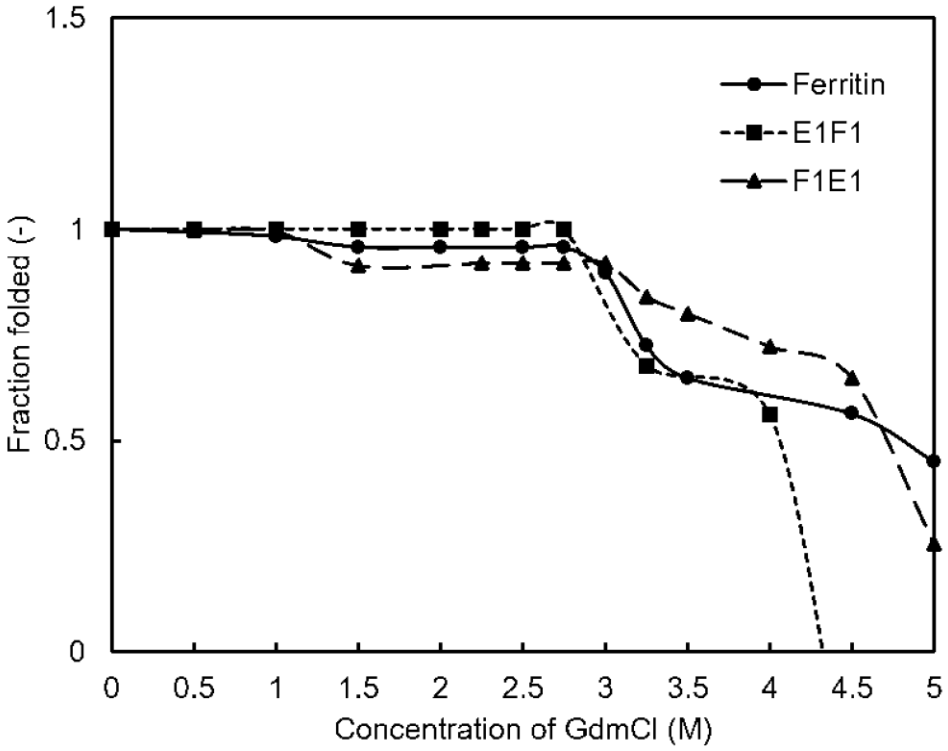


Figure 3.10 Circular dichroism spectra of ferritin, E1F1 and F1E1 with varied guanidinium chloride concentration from 0 M to 5 M at 20 °C for 20 min.

Compared with ferritin, both E1F1 and F1E1 were less tolerant to chemical denaturants. Transition of E1F1 occurred from 3 M GdnHCl. Fifty percent of the native structures in E1F1 and F1E1 were lost at 4.25 M and 4.60 M GdnHCl, respectively. It could be clearly observed that F1E1 showed better chemical stability than E1F1. E1F1 was completely unfolded at 4.5 M GdnHCl, whereas, for F1E1 and ferritin, folded structures still existed at pH 4.5. There is still limited knowledge about the mechanism of how chemical denaturant GdnHCl interacting with ferritin, F1E1 and E1F1 by binding to the specific sites or by altering the interactions between protein and water molecules. One possible explanation for the different behaviours towards chemical denaturants by E1F1 and F1E1 is the hydrophobic effect [60]. The smaller hydrophobic patch in F1E1 (Shown in Table 3.1) and also the simulation result of hydrophobic SASA encourage the protein to interact with water molecules, which moves towards a low entropy folded state. In contrast, for protein E1F1, as the EBNA1 epitope and soft linkers, are mainly composed of hydrophobic amino acids (Pro, Val, Gly, Ala and Phe), they were

completely exposed outside the ferritin cage, which contributed a larger amount of hydrophobic side chains to the surrounding solvent than ferritin and E1F1. The contribution of denaturant co-solvents to this hydrophobic effect was even stronger. It has been found from both MD simulation and experiment by England *et al.* [61], that guanidinium in particular interacts with the hydrophobic regions of a protein through its flat and nonpolar surface to reduce the unfavourable of its exposure to solvent. Hence, larger hydrophobic regions drive a stronger interaction between guanidinium and proteins. In the end, C-terminus insertion is more stable than N-terminus insertion against chemical denaturant GdnHCl.

3.4 Conclusions

The insertion site of epitopes in ferritin as an engineered nano vaccine plays an important role in the modified protein's stability and performance. In this study, two insertion sites in ferritin were considered. E1F1 was produced by inserting epitope and soft linkers at N-terminus of ferritin where epitopes were on the surface of nanoparticles, while F1E1 was inserting epitopes at C-terminus located inside the ferritin cage. Via a combined study of simulation and experiments, we demonstrated that N-terminus insertion (E1F1) and C-terminus insertion (F1E1) have different protein properties. Both simulation and experimental data suggested that engineered ferritin with N-terminus insertion was more hydrophobic and less stable than that of C-terminus insertion. The less hydrophobic SASA provided F1E1 less interactions between non-polar residues with water molecules. Therefore, a more tolerant pH and chemical denaturant stability was obtained. This work improves understanding on ferritin nanoparticle as vaccine carrier. It also provides a method by combining molecular simulation and experiment to predict engineered protein stability with epitope insertion, which potentially shortens the development process for nanoparticle protein based vaccine design.

3.5 References

1 Angelov, B. *et al.* DNA/Fusogenic lipid nanocarrier assembly: Millisecond structural dynamics. *The Journal of Physical Chemistry Letters* **4**, 1959-1964, doi:10.1021/jz400857z (2013).

2 Doll, T. A. P. F. *et al.* Nanoscale assemblies and their biomedical applications. *Journal of The Royal Society Interface* **10**, doi:10.1098/rsif.2012.0740 (2013).

3 Feger, G. *et al.* Prediction of amphiphilic cell-penetrating peptide building blocks from protein-derived amino acid sequences for engineering of drug delivery nanoassemblies. *The Journal of Physical Chemistry B* **124**, 4069-4078, doi:10.1021/acs.jpcb.0c01618 (2020).

4 Joshi, H. *et al.* Epitope engineering and molecular metrics of immunogenicity: A computational approach to VLP-based vaccine design. *Vaccine* **31**, 4841-4847, doi:https://doi.org/10.1016/j.vaccine.2013.07.075 (2013).

5 López-Sagaseta, J. *et al.* Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

6 Lua, L. H. L. *et al.* Bioengineering virus-like particles as vaccines. *Biotechnology and Bioengineering* **111**, 425-440, doi:10.1002/bit.25159 (2014).

7 Han, J.-A. *et al.* Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine: Nanotechnology, Biology and Medicine* **10**, 561-569, doi:https://doi.org/10.1016/j.nano.2013.11.003 (2014).

8 He, D. & Marles-Wright, J. Ferritin family proteins and their use in bionanotechnology. *New Biotechnology* **32**, 651-657, doi:https://doi.org/10.1016/j.nbt.2014.12.006 (2015).

9 Harrison, P. M. The structure and function of ferritin. *Biochemical Education* **14**, 154-162, doi:https://doi.org/10.1016/0307-4412(86)90203-7 (1986).

10 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

11 Li, C. Q. *et al.* Ferritin nanoparticle technology... A new platform for antigen presentation and vaccine development. *Industrial Biotechnology* **2**, 143-147, doi:10.1089/ind.2006.2.143 (2006).

12 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly

neutralizing H1N1 antibodies. *Nature* **499**, 102, doi:10.1038/nature12202 https://www.nature.com/articles/nature12202#supplementary-information (2013).

13 Sliepen, K. *et al.* Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **12**, 82, doi:10.1186/s12977-015-0210-4 (2015).

14 Deshpande, S. *et al.* Thermostable exoshells fold and stabilize recombinant proteins. *Nature Communications* **8**, 1442, doi:10.1038/s41467-017-01585-2 (2017).

15 Zhen, Z. *et al.* Ferritins as nanoplatforms for imaging and drug delivery. *Expert Opinion on Drug Delivery* **11**, 1913-1922, doi:10.1517/17425247.2014.941354 (2014).

16 Luzzago, A. & Cesareni, G. Isolation of point mutations that affect the folding of the H chain of human ferritin in *E.coli*. *The EMBO Journal* **8**, 569-576, doi:10.1002/j.1460-2075.1989.tb03411.x (1989).

17 Malleshappa Gowder, S. *et al.* Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *The Scientific World Journal* **2014**, 7, doi:10.1155/2014/971258 (2014).

18 Strub, C. *et al.* Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry* **5**, 9, doi:10.1186/1471-2091-5-9 (2004).

19 Argos, P. *et al.* Thermal stability and protein structure. *Biochemistry* **18**, 5698-5703, doi:10.1021/bi00592a028 (1979).

20 Novák, P. & Havlíček, V. in *Proteomic Profiling and Analytical Chemistry (Second Edition)* (eds P. Ciborowski & J. Silberring) 51-62 (Elsevier, 2016).

21 Schaller, A. *et al.* Computational study of elements of stability of a four-helix bundle protein biosurfactant. *Journal of Computer-Aided Molecular Design* **29**, 47-58, doi:10.1007/s10822-014-9803-6 (2015).

22 Camilloni, C. *et al.* Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophysical journal* **94**, 4654-4661, doi:10.1529/biophysj.107.125799 (2008).

23 Alizadeh-Rahrovi, J. *et al.* Structural stability of myoglobin and glycomyoglobin: a comparative molecular dynamics simulation study. *Journal of Biological Physics* **41**, 349-366, doi:10.1007/s10867-015-9383-2 (2015).

24 Du, J. *et al.* Understanding thermostability factors of barley limit dextrinase by molecular dynamics simulations. *Frontiers in Molecular Biosciences* **7**, doi:10.3389/fmolb.2020.00051 (2020).

25 Gu, J. *et al.* Molecular dynamics perspective on the thermal stability of mandelate

racemase. *Journal of Biomolecular Structure and Dynamics* **37**, 383-393, doi:10.1080/07391102.2018.1427631 (2019).

26 Jiang, X. *et al.* Structural and dynamic evolution of the amphipathic N-terminus diversifies enzyme thermostability in the glycoside hydrolase family 12. *Physical Chemistry Chemical Physics* **18**, 21340-21350, doi:10.1039/C6CP02998A (2016).

27 Bhadra, P. & Siu, S. W. I. Refined empirical force field to model protein–self-assembled monolayer interactions based on AMBER14 and GAFF. *Langmuir* **35**, 9622-9633, doi:10.1021/acs.langmuir.9b01367 (2019).

28 Tazikeh-Lemeski, E. Binding free energy and the structural changes determination in hGH protein with different concentrations of copper ions (A molecular dynamics simulation study). *Journal of Theoretical and Computational Chemistry* **15**, 1650045, doi:10.1142/S0219633616500450 (2016).

29 Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports* **7**, 44651, doi:10.1038/srep44651 (2017).

30 Kumar, C. V. *et al.* Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A. *Advanced Bioinformatics* **2014**, 502618-502618, doi:10.1155/2014/502618 (2014).

31 Gill, M. & McCully, M. E. Molecular dynamics simulations suggest stabilizing mutations in a *de novo* designed α/β protein. *Protein Engineering, Design and Selection* **32**, 317-329, doi:10.1093/protein/gzaa005 (2019).

32 Chan, M. K. *et al.* Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463-1469, doi:10.1126/science.7878465 (1995).

33 Lee, Y. *et al.* Dissecting the critical factors for thermodynamic stability of modular proteins using molecular modeling approach. *PLOS ONE* **9**, e98243, doi:10.1371/journal.pone.0098243 (2014).

34 Kamaraj, B. & Purohit, R. *In silico* screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *BioMed Research International* **2013**, 697051, doi:10.1155/2013/697051 (2013).

35 Stănciuc, N. *et al.* pH and heat induced structural changes of chicken ovalbumin in relation with antigenic properties. *International Journal of Biological Macromolecules* **93**, 572-581, doi:https://doi.org/10.1016/j.ijbiomac.2016.09.025 (2016).

36 Zheng, F. *et al.* Activity and thermostability of GH5 endoglucanase chimeras from

mesophilic and thermophilic parents. *Applied and Environmental Microbiology* **85**, e02079-02018, doi:10.1128/aem.02079-18 (2019).

37 Jiang, Z. *et al.* Effects of an electric field on the conformational transition of the protein: A molecular dynamics simulation study. *Polymers (Basel)* **11**, 282, doi:10.3390/polym11020282 (2019).

38 Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology* **235**, 625-634, doi:10.1006/jmbi.1994.1017 (1994).

39 Sviben, D. *et al.* Investigation of the thermal shift assay and its power to predict protein and virus stabilizing conditions. *Journal of Pharmaceutical and Biomedical Analysis* **161**, 73-82, doi:https://doi.org/10.1016/j.jpba.2018.08.017 (2018).

40 Chen, Z. *et al.* Molecular dynamics simulation of barnase: Contribution of noncovalent intramolecular interaction to thermostability. *Mathematical Problems in Engineering* **2013**, 504183, doi:10.1155/2013/504183 (2013).

41 Manjunath, K. & Sekar, K. Molecular dynamics perspective on the protein thermal stability: A case study using SAICAR synthetase. *Journal of Chemical Information and Modeling* **53**, 2448-2461, doi:10.1021/ci400306m (2013).

42 Pikkemaat, M. G. *et al.* Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering, Design and Selection* **15**, 185-192, doi:10.1093/protein/15.3.185 (2002).

43 Lobanov, M. Y. *et al.* Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 623-628, doi:10.1134/S0026893308040195 (2008).

44 Chakravarty, S. & Varadarajan, R. Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study. *Biochemistry* **41**, 8152-8161, doi:10.1021/bi025523t (2002).

45 Ding, Y. *et al.* Application of principal component analysis to determine the key structural features contributing to iron superoxide dismutase thermostability. *Biopolymers* **97**, 864-872, doi:10.1002/bip.22093 (2012).

46 Farrell, P. J. Epstein–Barr virus and cancer. *Annual Review of Pathology: Mechanisms of Disease* **14**, 29-53, doi:10.1146/annurev-pathmechdis-012418-013023 (2019).

47 Duellman, S. J. *et al.* Phosphorylation sites of Epstein-Barr virus EBNA1 regulate its function. *Journal of General Virology* **90**, 2251-2259, doi:10.1099/vir.0.012260-0 (2009).

48 He, L. & Zhu, J. Computational tools for epitope vaccine design and evaluation. *Current Opinion in Virology* **11**, 103-112, doi:https://doi.org/10.1016/j.coviro.2015.03.013 (2015).

49 Lindahl, E. *et al.* GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modeling Annual* **7**, 306-317, doi:10.1007/s008940100045 (2001).

50 Berhanu, W. M. & Hansmann, U. H. E. Side-chain hydrophobicity and the stability of $A\beta_{16-22}$ aggregates. *Protein Science* **21**, 1837-1848, doi:10.1002/pro.2164 (2012).

51 Erdur, H. *et al.* EBNA1 antigen-specific CD8+ T cells in cerebrospinal fluid of patients with multiple sclerosis. *Journal of Neuroimmunology* **294**, 14-17, doi:https://doi.org/10.1016/j.jneuroim.2016.03.010 (2016).

52 Burrows, S. R. *et al.* Have we cut ourselves too short in mapping CTL epitopes? *Trends in Immunology* **27**, 11-16, doi:10.1016/j.it.2005.11.001 (2006).

53 Miles, J. J. *et al.* TCR alpha genes direct MHC restriction in the potent human T cell response to a class I-bound viral epitope. *Journal of Immunology* **177**, 6804-6814, doi:10.4049/jimmunol.177.10.6804 (2006).

54 Sharma, R. *et al.* Elucidating the preference of dimeric over monomeric form for thermal stability of Thermus thermophilus isopropylmalate dehydrogenase: A molecular dynamics perspective. *Journal of Molecular Graphics and Modelling* **96**, 107530, doi:https://doi.org/10.1016/j.jmgm.2020.107530 (2020).

55 McNally, J. R. *et al.* Variant L-chain ferritins that cause neuroferritinopathy alter ferritin functionality and iron permeability. *Metallomics* **11**, 1635-1647, doi:10.1039/C9MT00154A (2019).

56 Day, R. *et al.* Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding. *Journal of Molecular Biology* **322**, 189-203, doi:https://doi.org/10.1016/S0022-2836(02)00672-1 (2002).

57 Chen, H. *et al.* Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions. *Chemical Communications* **52**, 7402-7405, doi:10.1039/C6CC03108K (2016).

58 Stoffelen, C. & Huskens, J. Soft supramolecular nanoparticles by noncovalent and host–guest interactions. *Small* **12**, 96-119, doi:10.1002/smll.201501348 (2016).

59 Zellner, M. *et al.* Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. *ELECTROPHORESIS* **26**, 2481-2489, doi:10.1002/elps.200410262 (2005).

60 Pace, C. N. in *Methods in Enzymology* Vol. 131 266-280 (Academic Press, 1986).

61 England, J. L. *et al.* Chemical denaturants inhibit the onset of dewetting. *Journal of the American Chemical Society* **130**, 11854-11855, doi:10.1021/ja803972g (2008).

3.6 Supporting Information

3.6.1 Supporting figures

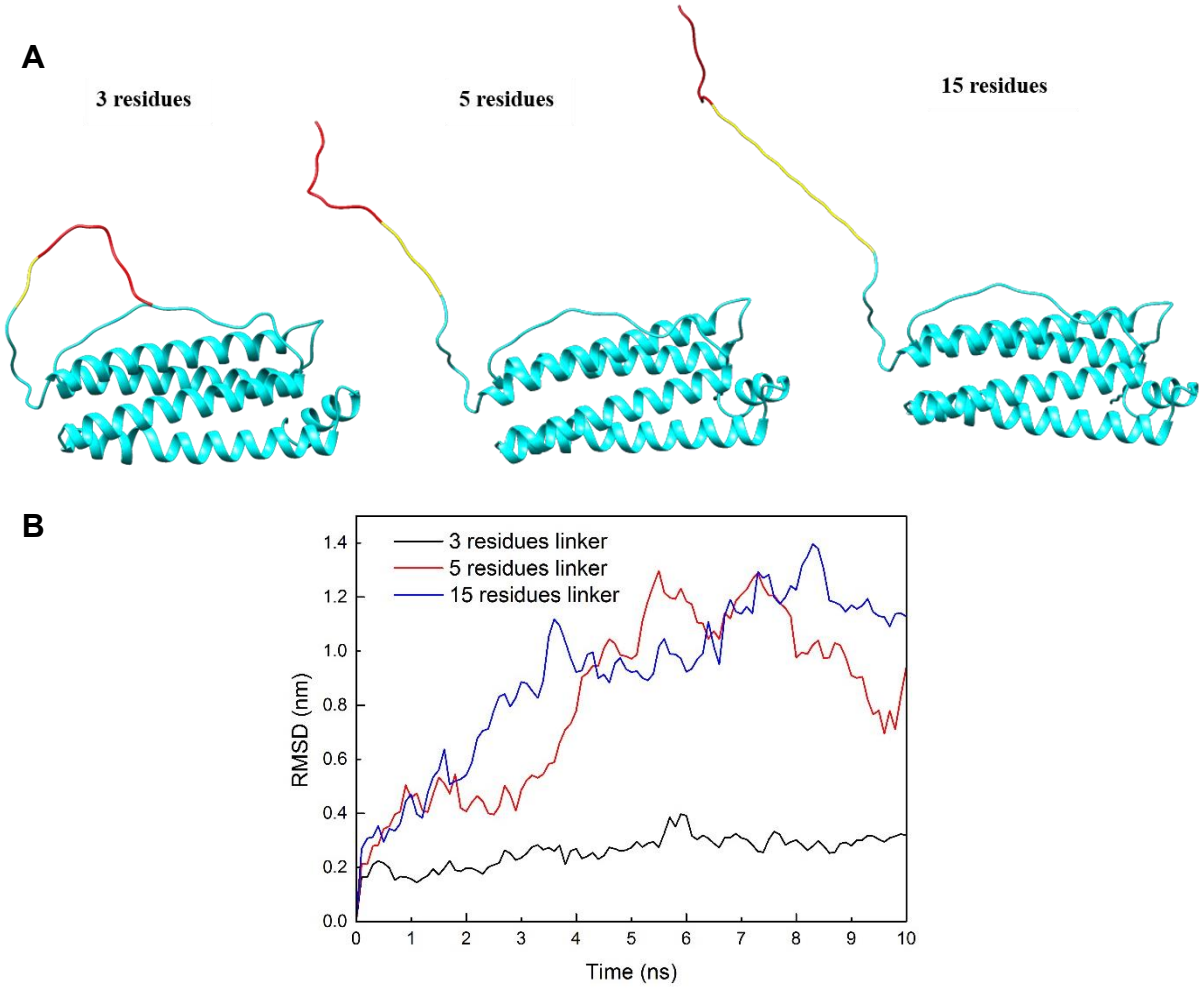


Figure S3.1 Stability comparisons for monomer of E1F1 (N-terminus insertion) with different lengths of linkers (3 residues, 5 residues and 15 residues). A) Mimic diagrams of E1F1 monomers with different lengths of linkers (3 residues, 5 residues and 15 residues) by Chimera. B) RMSD of E1F1 monomers with different lengths of linkers (3 residues, 5 residues and 15 residues).

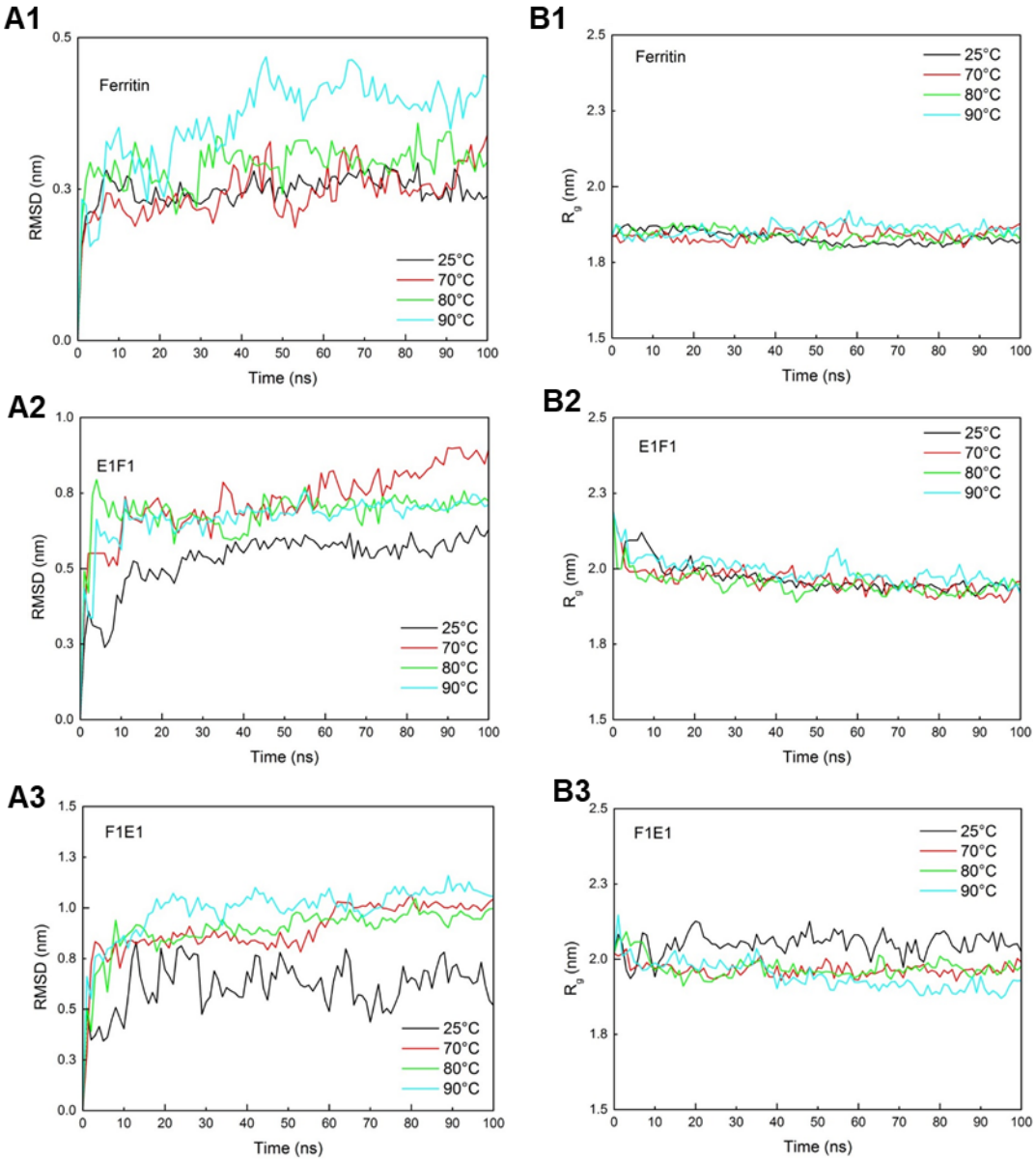


Figure S3.2 Simulation results from GROMACS for monomers of ferritin, E1F1, and F1E1 at 25, 70, 80 and 90 °C. A1) RMSD of ferritin. A2) RMSD of E1F1. A3) RMSD of F1E1. B1) R_g of ferritin. B2) R_g of E1F1. B3) R_g of F1E1.

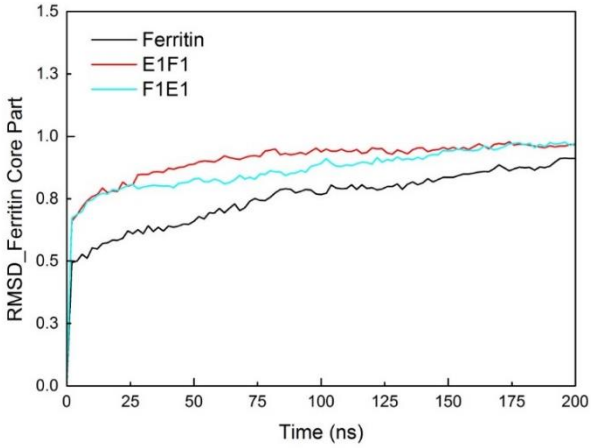


Figure S3.3 Simulation RMSD data for assembled structures of ferritin, E1F1 and F1E1 using GROMACS at 150°C for 200 ns by comparing ferritin core part only.

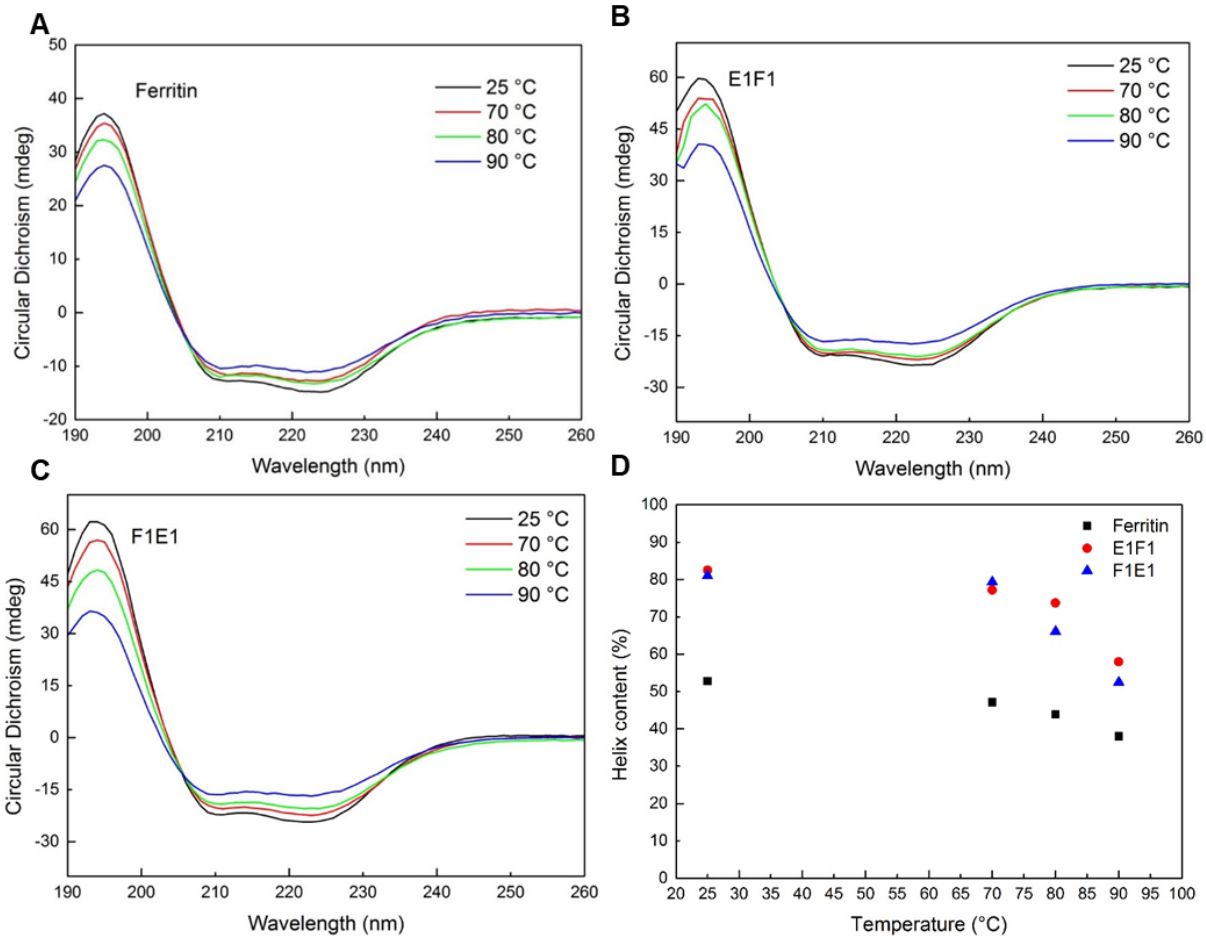


Figure S3.4 Circular dichroism spectra of ferritin, E1F1 and F1E1 at 25, 70, 80 and 90 °C. A) Ferritin. B) E1F1. C) F1E1. D) Helix content of ferritin, E1F1 and F1E1 at 25, 70, 80 and 90 °C.

3.6.2 Supporting tables

Table S3.1 Constructed PDB files by Discovery Studio.

PDB file name	Represented structure name
F_1	Ferritin monomer subunit
F_3	Ferritin trimer subunit
F_{24}	Ferritin bio assembly
E_1F_1 monomer	Monomer of ferritin with inserted EBNA1 at the N-terminal with linkers (GGS)
E_1F_1 trimer	Trimer of ferritin with inserted EBNA1 at the N-terminal with linkers (GGS)
E_1F_1 bio assembly	Bio assembly of ferritin with inserted EBNA1 at the N-terminal with linkers (GGS)
F_1E_1 monomer	Monomer of ferritin with inserted EBNA1 at the C-terminal without linkers
F_1E_1 trimer	Trimer of ferritin with inserted EBNA1 at the C-terminal without linkers
F_1E_1 bio assembly	Bio assembly of ferritin with inserted EBNA1 at the C-terminal without linkers

Table S3.2 Cubic box for each structure generated from GROMACS.

PDB file name	Cubic box dimensions (nm*nm*nm)
F_1	10*10*10
F_3	12*12*12
F_{24}	22*22*22
E_1F_1 monomer	10*10*10
E_1F_1 trimer	12*12*12
E_1F_1 bio assembly	22*22*22
F_1E_1 monomer	10*10*10
F_1E_1 trimer	12*12*12
F_1E_1 bio assembly	22*22*22

Table S3.3 Amino acid sequence of ferritin, E1F1 and F1E1.

Name	Amino acid Sequence
Ferritin	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSH EEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES
EBNA1-Ferritin **(Ferritin with inserted EBNA1 epitope at the N-terminus)**	HPVGEADYFEYGGSGGGGSGGGGSGGTTASTSQVRQNYHQDSEAAINRQINLELYASYV YLSMSYYFDRDDVALKNFAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES
Ferritin-EBNA1 **(Ferritin with inserted EBNA1 epitope at the C-terminus)**	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSH EEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNESGGSGGGGSGGGGSGG HPVGEADYFEY

Table S3.4 CD result of helix content of ferritin, E1F1 and F1E1 over temperature range of 25, 70, 80 and 90°C.

Temperature (°C)	Ferritin helix content (%)	E1F1 helix content (%)	F1E1 helix content (%)
25	52.8	82.5	81.1
70	47.1	77.2	79.4
80	43.9	73.7	66.1
90	38	58	52.5

Table S3.5 Simulation result for helix content of one subunit of ferritin, E1F1 and F1E1 at defined temperatures.

Temperature (°C)	Ferritin helix content (%)	E1F1 helix content (%)	F1E1 helix content (%)
25	84.6	85.7	88.4
70	89.6	85.7	71.7
80	77.3	74.4	76.6
90	83.6	82.2	65.9

Table S3.6 Simulation result for helix content of trimer of ferritin, E1F1 and F1E1 at defined temperatures.

Temperature (°C)	Ferritin helix content (%)	E1F1 helix content (%)	F1E1 helix content (%)
25	83.8	86.5	86.4
70	82.8	77.2	85.5
80	78.8	82.0	82.1
90	76.5	76.4	83.7

3.6.3 Other supporting information not included in online publication

Figure S3.5 shows that E1F1 (lane 4) and F1E1 (lane 8) both achieved soluble expression by using expression condition 37 °C, 200 rpm, 4 h. There were minor inclusion bodies expression (lane 5: E1F1; lane 9: F1E1).

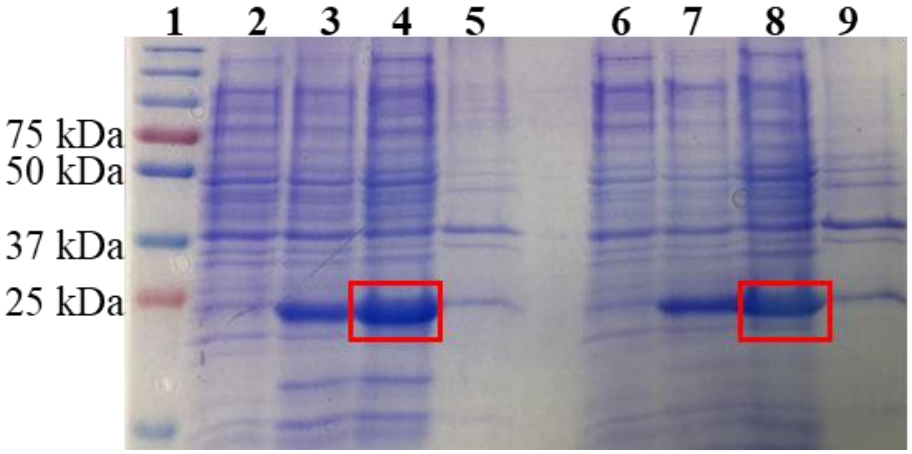


Figure S3.5 SDS-PAGE image of engineered ferritins (E1F1 and F1E1) expression. Theoretical MW for E1F1 and F1E1 were 23.35 kDa. Lane 1: Protein marker; lane 2: E1F1 before adding IPTG; lane 3: E1F1 after adding IPTG; lane 4: E1F1 soluble expression; lane 5: E1F1 inclusion bodies expression; lane 6: F1E1 before adding IPTG; lane 7: F1E1 after adding IPTG; lane 8: F1E1 soluble expression; lane 9: F1E1 inclusion bodies expression. Red box indicated target proteins E1F1/F1E1.

Figure S3.6 and S3.7 showed the first step of E1F1 purification, acid and heat precipitation. It was found that the condition (pH 5, 60 °C, 10 min, and total protein concentration 4 mg mL^{-1}) demonstrated the most purified protein.

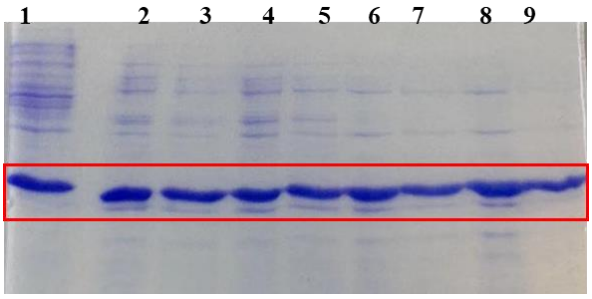


Figure S3.6 SDS PAGE image of E1F1 heat precipitation at pH 5 for various conditions. Lane 1: E1F1 before purification; lane 2: 50 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 3: 50 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 4: 50 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 5: 50 °C, 10 min, total protein concentration 4 mg mL^{-1}; lane 6: 60 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 7: 60 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 8: 60 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 9: 60 °C, 10 min, total protein concentration 4 mg mL^{-1}. Red box indicated target protein E1F1.

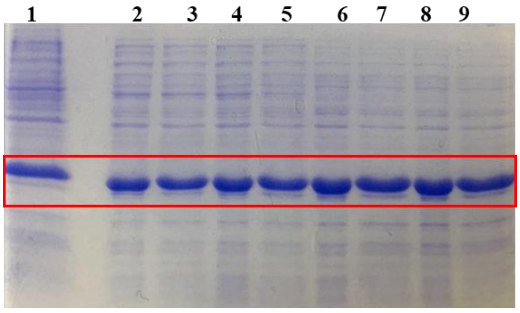


Figure S3.7 SDS PAGE image of E1F1 heat precipitation at pH 5.5 for various conditions. Lane 1: E1F1 before purification; lane 2: 50 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 3: 50 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 4: 50 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 5: 50 °C, 10 min, total protein concentration 4 mg mL^{-1}; lane 6: 60 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 7: 60 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 8: 60 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 9: 60 °C, 10 min, total protein concentration 4 mg mL^{-1}. Red box indicated target protein E1F1.

Figure S3.8 indicated that the optimal acid and heat precipitation condition for F1E1 was pH 4.5, 60 °C, 10 min, total protein concentration 4 mg mL^{-1}.

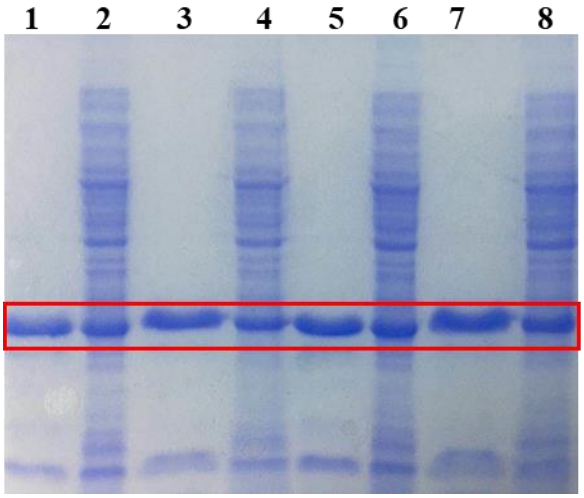


Figure S3.8 SDS PAGE image of F1E1 heat precipitation at pH 4.5 for various conditions. Lane 1: F1E1 supernatant after heating precipitation at 60 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 2: F1E1 precipitation after heating precipitation 60 °C, 5 min, total protein concentration 8 mg mL^{-1}; lane 3: F1E1 supernatant after heating precipitation at 60 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 4: F1E1 precipitation after heating precipitation 60 °C, 5 min, total protein concentration 4 mg mL^{-1}; lane 5: F1E1 supernatant after heating precipitation at 60 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 6: F1E1 precipitation after heating precipitation 60 °C, 10 min, total protein concentration 8 mg mL^{-1}; lane 7: F1E1 supernatant after heating precipitation at 60 °C, 10 min, total protein concentration 4 mg mL^{-1}; lane 8: F1E1 precipitation after heating precipitation 60 °C, 10 min, total protein concentration 4 mg mL^{-1}. Red box indicated target protein F1E1.

Figure S3.9 showed the second step of E1F1 and F1E1, using HIC to remove nucleic acid. The major elution peak for both two proteins achieved more than 90 % purity.

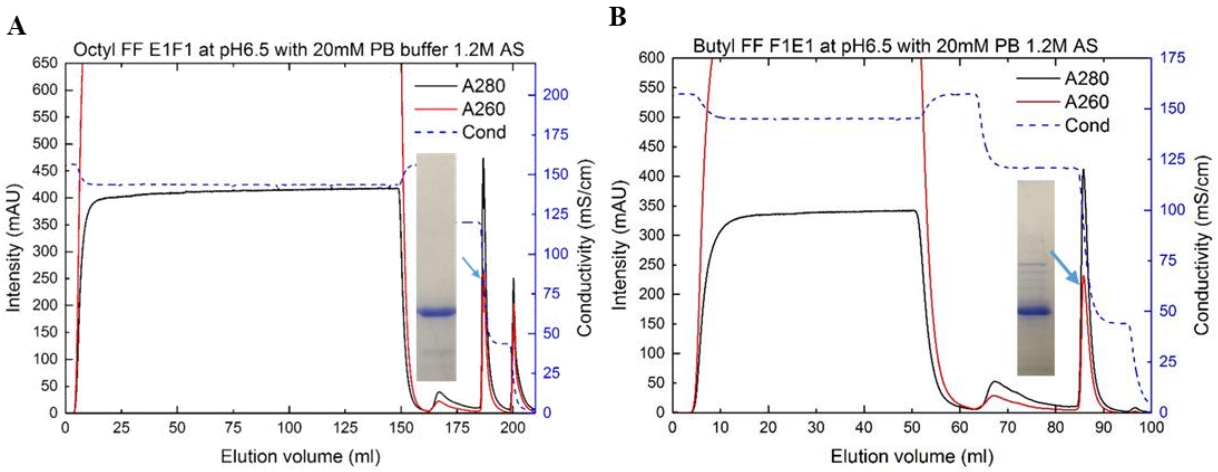


Figure S3.9 Chromatogram of E1F1 and F1E1 purification by HIC. A) E1F1 purification by Octyl FF at pH 6.5, 1.2M AS. B) F1E1 purification by Butyl FF at pH 6.5, 1.2M AS.

Figure S3.10 showed the purified E1F1 and F1E1 a good diversity of nanoparticles in TEM.

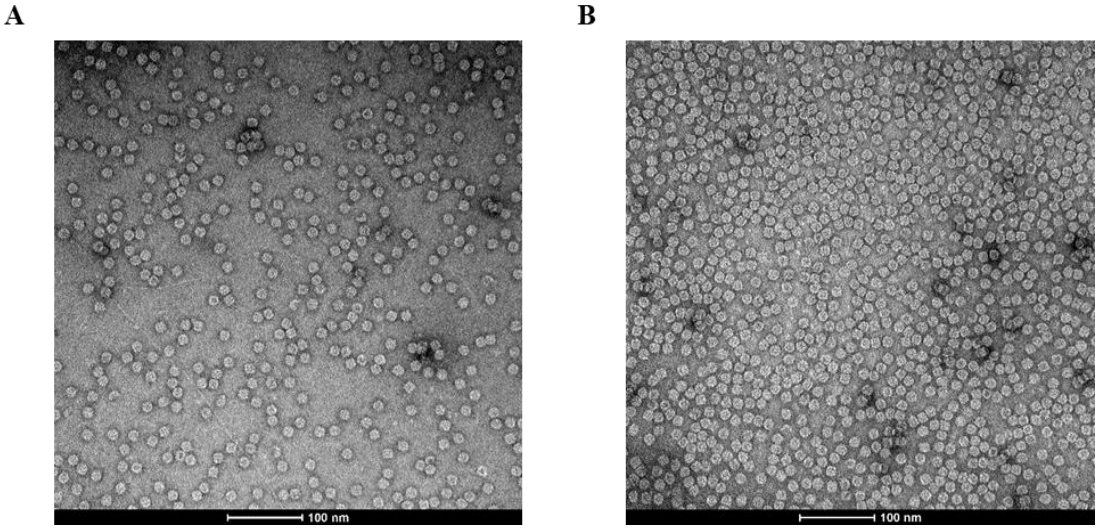


Figure S3.10 TEM images of purified E1F1 and F1E1. A) E1F1. B) F1E1.

Publication included in Chapter 3

Qu, Y. *et al.* Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments. *The Journal of Physical Chemistry B* **125**, 3830-3842, doi:10.1021/acs.jpcb.1c00276 (2021).

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments

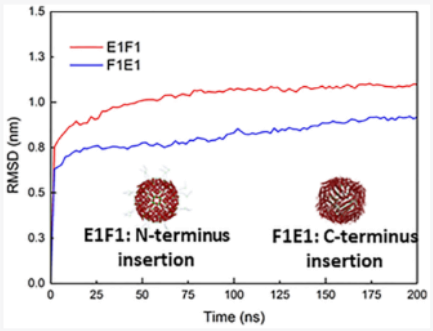
Yiran Qu, Lijie Wang, Shuang Yin, Bingyang Zhang, Yan Jiao, Yan Sun, Anton Middelberg, and Jingxiu Bi*

Cite This: *J. Phys. Chem. B* 2021, 125, 3830−3842

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Human ferritin is regarded as an attractive and promising vaccine platform because of its uniform structure, good plasticity, and desirable thermal and chemical stabilities. Besides, it is biocompatible and presumed safe when used as a vaccine carrier. However, there is a lack of knowledge of how different antigen insertion sites on the ferritin nanocage impact the resulting protein stability and performance. To address this question, we selected Epstein−Barr nuclear antigen 1 as a model epitope and fused it at the DNA level with different insertion sites, namely, the N- and C-termini of ferritin, to engineer proteins E1F1 and F1E1, respectively. Protein properties including hydrophobicity and thermal, pH, and chemical stability were investigated both by molecular dynamics (MD) simulation and by experiments. Both methods demonstrate that the insertion site plays an important role in protein properties. The C-terminus insertion (F1E1) leads to a less hydrophobic surface and more tolerance to the external influence of high temperature, pH, and high concentration of chemical denaturants compared to N-terminus insertion (E1F1). Simulated protein hydrophobicity and thermal stability by MD were in high accordance with experimental results. Thus, MD simulation can be used as a valuable tool to engineer nanovaccine candidates, cutting down costs by reducing the experimental effort and accelerating vaccine design.



INTRODUCTION

To replace traditional live attenuated strains of pathogens and inactivated killed pathogen vaccines, thereby reducing safety risks, there has been a rapid development in advanced vaccine carriers, such as polymers, liposomes, and nanoparticle proteins.[1−6] Wild-type heavy-chain ferritin comprises 24 identical subunits, each 21 kDa in size, forming a spherical hollow cage structure 12 nm in diameter.[7] The inner cavity of the spherical hollow cage structure has a diameter of 8 nm (Figure 1A,B).[8] Every ferritin subunit is made up of five helices, A to E, with a long BC loop and three short loops (AB, CD, and DE) connecting other helices (Figure 1A).[9]

Ferritin has considerable potential as a nanovaccine carrier because of its numerous advantages. First and foremost, by genetic modification, ferritin can carry antigens and expose immunogens in a repetitive and well-organized manner. Ferritin can display exogenous peptides[10,11] and heterologous proteins, such as the influenza virus hemagglutinin[12] and native-like HIV-1 envelope glycoprotein trimers.[13] Both examples indicate that ferritin-based protein nanoparticles enhance the potency and breadth of virus vaccine immunity.[10−13] Second, ferritin is extremely thermally and chemically stable[5] and can be engineered for enhanced physicochemical and biological properties.[14] Last, due to its human source, human ferritin is considered safe with low toxicity, high biocompatibility, and good biodegradability.[15]

As a vaccine carrier displaying epitope candidates, ferritin has been investigated with three potential insertion sites, namely, the N-terminus, C-terminus, and DE loop region. The N-terminus of ferritin is connected to helix A and is located on the outside of the protein shell, whereas the C-terminus is located at the fifth short α-helix E on the inside of the hollow cage (Figure 1A).[9] The native structure of ferritin is defined as a *flip* structure. However, if the exogenous peptide of a heterologous protein inserted at the C-terminus is very large and unable to be packed inside the ferritin cage, helix E points outside creating a flop conformation.[16] It has been found that the volume inside the ferritin cage is approximately 2.1×10^5 Å^3.[16] By assuming that each amino acid has an average volume of 150 Å^3, the ferritin could potentially contain a total of 1400 amino acids inside the ferritin hollow cage by inserting them at the C-terminus. If there are more than 1400 amino acids

Received: January 11, 2021
Revised: March 14, 2021
Published: April 7, 2021

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

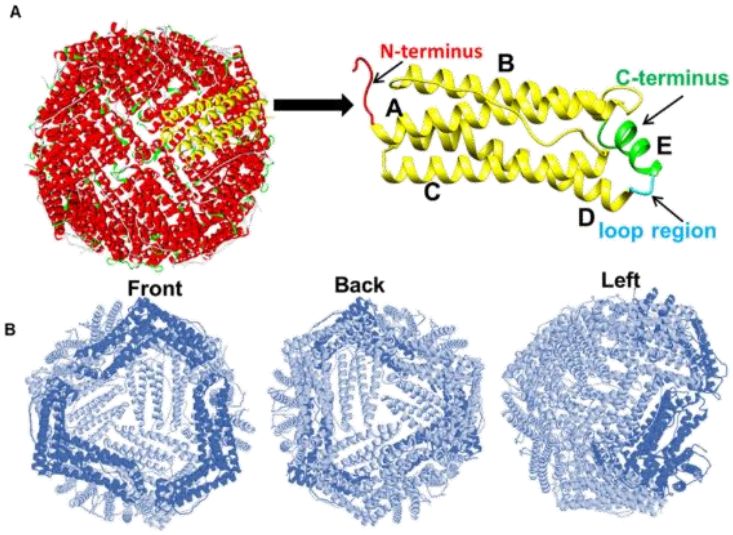


Figure 1. Human ferritin heavy-chain mimic visualized with DS 2019. (A) Native ferritin assembly structure and a monomer of native ferritin. Three insertion sites are indicated by different colors (red: N-terminus; green: C-terminus; blue: flexible loop region). (B) Ferritin hollow cage viewed by different angles.

inserted at the C-terminus, ferritin tends to become a *flop* structure. In this case, the inserted epitopes at the C-terminus are presented outside. Otherwise, if the inserted epitope is with less than 1400 amino acids at the C-terminus, ferritin intends to be a *flip* structure to present the epitope inside of the ferritin cage. The *flip* or *flop* structure may in turn play an important role in defining the quality and extent of the immune response.

There is limited knowledge about the connection between the different insertion sites and the corresponding protein properties. A recent study undertaken by Wang et al. displayed different lengths of epitopes from Enterovirus 71 at the N- and C-termini and the DE loop region (163 position) of ferritin to compare immune responses.[10] They found that peptides inserted into the DE loop region invoked a stronger immune response than at the N-terminus and C-terminus.[10] However, there are still knowledge gaps when comparing protein properties following N-terminus and C-terminus insertion, such as effects of hydrophobicity and thermal, pH, and chemical stability.

The first important aspect to consider is protein hydrophobicity, which contributes significantly to protein folding and the stability of the structure.[17] A strategy to increase the stability of proteins is to decrease the hydrophobic solvent accessible surface area (SASA).[18] The interaction of nonpolar residues with water causes the side chain to be more exposed to solvents for proteins with larger hydrophobic SASAs; hence, this effect results in reduced stability.[18] Second, thermal stability is indispensable when developing a vaccine platform to ensure the bioactivity of the nanoparticle proteins. Most proteins are restricted to an allowed temperature range (4−60 °C) to maintain structural integrity.[19] One consequence of a temperature excursion out of this allowed range is the loss of protein functions. By determining the thermal stability for ferritin and engineered ferritin with different insertion sites, our knowledge of structural diversity is conceivably expanded. Last, pH and chemical stability are two other important protein properties to consider. The isoelectric point (pI), which is related to the pH stability, is the pH value at which the protein's net charge is 0. At the pI, protein solubility and electrical repulsion are the lowest, resulting in the highest tendency for aggregation and precipitation.[20] pH is also a key factor because it can cause the alterations in the structure of a protein, which affects bioactivity. Protein stability against chemical denaturants, such as guanidinium chloride (GdnHCL), is helpful to understand protein structure differences.[21] GdnHCL acts as an ionic denaturant and is commonly used to probe the folding properties of proteins.[22] Accordingly, hydrophobicity and thermal, pH, and chemical stability are deterministic of overall protein properties.

The experimental investigation of protein properties is a time-consuming process, especially when dealing with large numbers of samples during early vaccine development. Recently, molecular dynamics (MD) simulation has been widely applied to investigate important parameters related to protein stability.[23−26] The key factors related to protein stability include the hydrophobic SASA,[27,28] root-mean-square deviation (RMSD), and radius of gyration (R_g). SASA could directly show the changes in the accessibility of the protein to the solvent,[29,30] and it is related to protein stability as well. A lower SASA generally represents higher thermodynamic stability of the protein.[31−34] RMSD, R_g, and α-helix content are factors to predict protein structure changes related to thermal change, along with presenting in-depth explanation for the transition.[21,35,36] RMSD is commonly used to reveal protein structural stability.[37,38] The higher the RMSD is, the less stable the protein structure is suggested.[29,30,39−42] R_g can be regarded as the basic measurement of protein overall size.[37,43] A smaller R_g value indicates a more compacted protein structure and may result in high stability. α-Helix content indicates the protein secondary structure change at high temperatures. The helical content of proteins increases with increasing thermal stability.[44,45] MD simulation can provide reliable prediction of protein thermal-related properties and stability, which is able to be applied in the design of vaccine candidates, to significantly reduce experiment related costs, time frames, and failure.

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

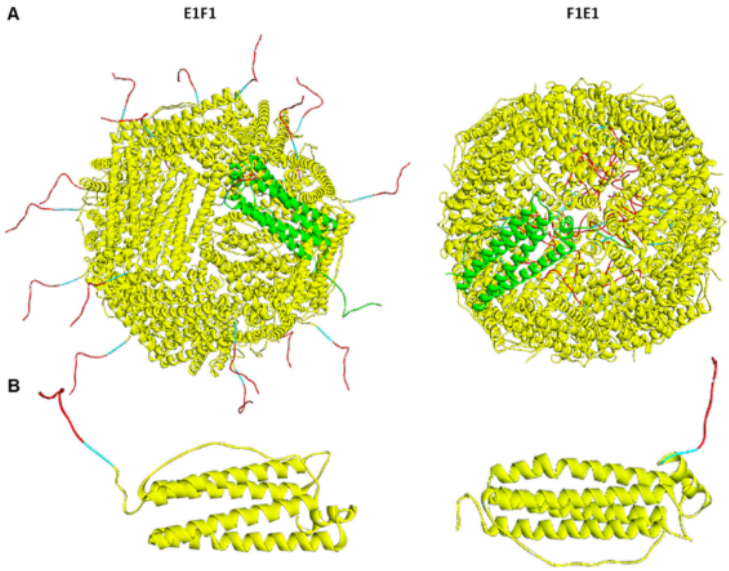


Figure 2. Engineered ferritins E1F1 and F1E1 mimic diagrams using DS 2019. Inserted epitopes are highlighted in red, and linkers are highlighted in blue. (A) Assembly structures of E1F1 and F1E1. A monomer of E1F1 and F1E1 are denoted in green. (B) Monomer structure of E1F1 and F1E1. The epitope (red) and linker (blue) are inserted at the N-terminus of the ferritin monomer for E1F1, while they are inserted at the C-terminus of the ferritin monomer for F1E1.

In this study, Epstein−Barr nuclear antigen 1 (EBNA1) was selected as the model epitope as it shows high potential as a vaccine candidate against the Epstein−Barr virus (EBV). EBV is responsible for many human diseases such as lymphoma and gastric cancer and might enhance the development of several auto immune diseases.[46] EBNA1 is expressed in all EBV-associated tumors, and it is an essential protein for EBV maintenance and expression. There are limited therapies for the prevention of EBV-associated diseases, and no licensed vaccine is available to combat EBV infection.[46] EBNA1 determines the replication and partitioning of viral genomic DNA at latent viral infection. It is also the only viral protein found in all EBV-related malignancies.[47] Hence, EBNA1-associated vaccines will potentially act as prophylactic vaccines against all EBV strains.

By inserting the EBNA1 epitope at the N-terminus and the C-terminus of ferritin with 15 residues of linkers between the epitope and the ferritin, we aimed to investigate the stability-related properties of engineered ferritins. The four main properties we took into account were hydrophobicity, thermal stability, pH stability, and chemical stability. We developed a new approach that combined the results from MD simulation and experiments to improve stability predictions and the underlying mechanisms of designing ferritin−EBNA1 as an engineered nanovaccine.

MATERIALS AND METHODS

Protein Data Bank File Buildup. The Protein Data Bank (PDB) files of the human heavy-chain ferritin (PDB ID: 2FHA) and EBNA1 (PDB ID: 2FZ3) were generated from the RCSB PDB. Discovery Studio (DS) was used to edit the epitope sequence for EBNA1. The model epitope was truncated off from EBNA1 comprising aa 407 to aa 417. A set of ferritin subunit structures without the inserted epitope, ferritin monomer (F_1), ferritin trimer (F_3), and ferritin bioassembly structure (F_{24}) were prepared as the reference. By inserting one EBNA1 epitope at the N-terminus of the ferritin monomer structure with the linker, the PDB file named E1F1 monomer was prepared. The length of the linker between the ferritin and EBNA1 was assumed as 3 residues rather than 15 residues as the longer linkers created complexity in computational simulation and resulted in a highly unstable system. For example, in Figure S1 (Supporting Information), the E1F1 monomer with 3 residues of the linker showed the most stable structure with the lowest RMSD compared to 5 and 15 residues of linkers. In the same manner, a set of PDB files for ferritin with the inserted EBNA1 epitope at the C-terminus with three residues of the linker were prepared, as shown in Table S1. Based on theoretical calculations, epitopes and linkers with a total of 624 amino acids could fit the ferritin cage completely. Hence, F1E1 (C-terminus insertion) was assumed to have a ferritin *flip* structure, where the inserted epitopes were contained inside of the protein cage, shown in Figure 2.

MD Simulation of Ferritin Structures and Ferritin−EBNA1 Structures. GROMACS 2018.3 was used to perform MD simulations of the ferritin subunit structures and ferritin−EBNA1 structures in a water environment.[48,49] The OPLS-AA/L all-atom force field was used to generate topology files.[49] Each protein structure was solvated in a cubic box using the TIP3P water model by specifying spc216.gro in GRAOMCS.[50] The dimensions of the cubic box for each structure are summarized in Table S2. The minimum distance between the protein and the wall of the unit cell was set to be 1.0 nm. The protein/ion/water system was first energy-minimized and then equilibrated in a 100 ps MD simulation with positional restraints on the heavy atoms of the protein. The equilibration was performed at a temperature of 300 K and a pressure of 1 bar by coupling to an external heat source and an isotropic pressure bath. The position restraints were then released, and the MD simulation was produced in 200 ns and the snapshots were stored per 100 ps for the MD trajectory analysis.

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

For the hydrophobicity study by simulation, bovine serum albumin (BSA) (PDB ID: 3V03) was simulated under the same conditions. Hydrophobic SASA was determined by using the gmx sasa tool from GROMACS based on the completed MD simulation results for BSA, ferritin, E1F1 (N-terminus insertion), and F1E1 (C-terminus insertion) in water at 25 °C. The error analysis on hydrophobic SASA was calculated using block averaging (gmx analyze -ee).

Thermal stability studies for the monomer and trimer structures of ferritin, EBNA1−ferritin, and ferritin−EBNA1 followed similar procedures to above, but the MD simulation was produced in 100 ns over a temperature range from 25 to 90 °C. The snapshots were stored per 50 ps for the MD trajectory analysis. RMSD and R_g of C_α were determined using the gmx rmsd tool and gmx gyrate tool from GROMACS. The helix content was calculated using the gmx dssp tool from GROMACS. For bioassembly structures of E1F1 and F1E1, MD simulations were produced in 200 ns at 150 °C. The snapshots were stored per 100 ps for the MD trajectory analysis. RMSD and helix content were determined.

Plasmid Construction of Recombinant Ferritin Complexes. The PET 30a vector was used to construct plasmids harboring genes of two types of EBNA1-fused ferritins (ferritin−EBNA1 and EBNA1−ferritin) and the control ferritin. Target genes were inserted between *BamHI* and *XhoI* restriction sites. The epitope, EBNA1 (HPVGEADYFEY), in two ferritin complexes, is the main immunogenic domain of EBNA.[51] This linear epitope is known to be a potent natural target for cytotoxic T-lymphocytes.[52,53] A 15-residue linker (GGSGGGGSGGGGSGG) was inserted between the epitope and ferritin in both ferritin complexes. The amino acid sequence information for ferritin and recombinant protein ferritin−EBNA1 is shown in Table S3.

Protein Expression and Purification. The plasmids were transformed into *Escherichia coli* (*E. coli*) BL21 (DE3). The expression process of the vector ferritin and the two engineered ferritins was the same. A single bacterial colony expressing each protein, namely, ferritin, EBNA1−ferritin (E1F1), and ferritin−EBNA1 (F1E1), was inoculated in 50 mL of the LB medium supplemented with kanamycin (100 μg/mL) and incubated overnight at 37 °C in an orbital mixer incubator (Ratek), shaken at 180 rpm. 10 mL of the bacterial suspension was then transferred to 500 mL of the LB with kanamycin (100 μg/mL) medium. The culture was grown at 37 °C and shaken at 200 rpm until OD_{600} reached 0.8, followed by induction with 1 mM isopropyl β-d-1-thiogalactopyranoside at 37 °C, 200 rpm for 4 h. Cells were collected after centrifugation at 13,751*g*. Cell pellets were resuspended in the lysis buffer and disrupted by sonication (a Scientx sonicator at 360 W: 4 s on and 6 s off for 10 min) with the lysis buffer [20 mM phosphate buffer (PB), 2 mM ethylenediaminetetraacetic acid, pH 7.0]. Bacterial lysate was centrifuged at 13,751*g* for 30 min at 4 °C to remove cell debris, and the supernatant was harvested for further protein purification.

The protein purification processes for ferritin, EBNA1−ferritin, and ferritin−EBNA1 were the same, except that they differed in the used chromatographic resin. First, the clarified supernatant was diluted with Milli-Q water to a final protein concentration of 4 mg/mL. NaCl and NaAc−HAc were added to a final concentration of 1 M and 100 mM, respectively; the pH was adjusted to 5.0, and mixtures then underwent thermal precipitation at 60 °C for 10 min. Centrifugation (19,802 *g* at 4 °C for 10 min) was conducted to remove precipitated impurities. Subsequently, the resultant supernatant pH was adjusted to 6.5 by using 1 M PB, pH 8.0, and 1 M ammonium sulfate (AS) was added for the following hydrophobic interaction chromatography (HIC). In HIC, a HiTrap Octyl FF column (GE Healthcare, USA) was used for ferritin and E1F1; a HiTrap Butyl FF column (GE Healthcare, USA) was used for F1E1. The equilibrated buffer was 100 mM PB, 1 M AS, pH 6.5. The eluted buffer was 20 mM PB, pH 6.5. HIC was performed on an AKTA pure (GE Healthcare, USA). Briefly, samples were loaded onto columns equilibrated with the equilibration buffer. After washing the column for 5 column volumes with the loading buffer, a linear gradient elution of 0−100% elution buffer in 3 column volumes was conducted and eluted peaks were pooled together. Absorbance at 280 and 260 nm was recorded throughout chromatography. Finally, collected fractions from HIC were buffer-exchanged to 20 mM PB, pH 7 using a HiTrap Desalting column (GE Healthcare, USA). Protein expression and purities were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. Protein expression and purification were performed with 12% (w/v) SDS separation gel having a 5% (w/v) SDS stacking layer. Samples were prepared with 5× loading buffer [containing 6% (v/v) 1 M Tris-HCL at pH 6.8, 25% (v/v) glycerol, 1 mg/mL bromophenol blue, 2% (v/v) SDS, and 5% v/v β-mercaptoethanol]. Treated samples were heated at 100 °C for 10 min and loaded into the loading wells. Separation was conducted at 200 V for 60 min. The gel was stained with 0.25% (w/v) Coomassie R-250 for 60 min, and the background was washed using a washing buffer [10% (v/v) ethanol and 10% (v/v) acetic acid] for 270 min.

Transmission Electron Microscopy. Purified ferritin and engineered ferritins were diluted to a protein concentration of 0.1 mg/mL. Carbon-coated copper grids were incubated facedown on 10 μL of the sample for 10 min. The grid was washed 3× in droplets of Milli-Q water and then negatively stained in 2% (v/v) uranyl acetate for 2 min. The excess liquid was removed with a filter paper. Grids were analyzed with a Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

Size Exclusion Chromatography Coupled with Multiple-Angle Laser Light Scattering. Size exclusion chromatography with multiple-angle laser light scattering (SEC-MALS) was performed to determine the hydrodynamic radius and molecular weight of purified and engineered ferritins. Superose 6 10/300 GL (GE Healthcare, USA) was connected to a high-performance liquid chromatography system (Shimadzu, Japan) coupled with a Wyatt Optilab refractive index and Wyatt DAWN MALLS detector (USA). The equilibrated buffer was 20 mM PB, pH 7.0. The protein concentration was 5 mg/mL. The loading volume was 50 μL.

Hydrophobicity Study. A hydrophobicity study was performed to compare the hydrophobicity differences among ferritin, E1F1, and F1E1. BSA was used as a control sample. 1 mL of each protein sample (BSA, ferritin, E1F1, and F1E1) was adjusted to 2 mg/mL. Each protein sample was mixed with 4 mL of the equilibrated buffer (1.0 M AS buffer, 100 mM PB containing Na_2HPO_4 and NaH_2PO_4, pH 6.5). A HiTrap Butyl FF column (GE Healthcare, USA) was used to perform the hydrophobicity study. The elution buffer was 20 mM PB, pH 6.5. After the column was equilibrated, protein samples were loaded to a HiTrap Butyl FF column individually. The column

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

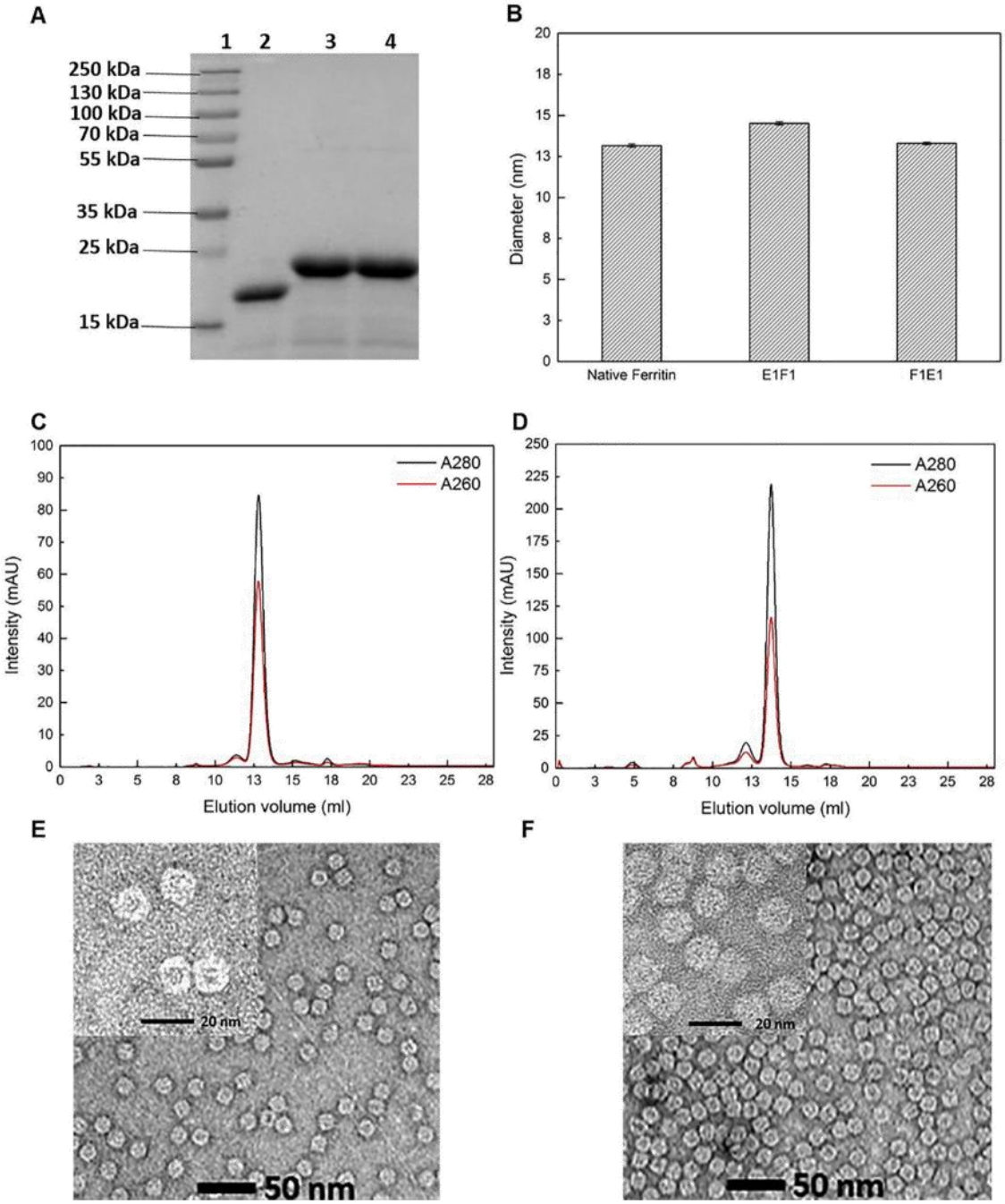


Figure 3. Characterization of engineered ferritins. (A) Coomassie blue-stained SDS-PAGE: lane 1, protein marker; lane 2, native ferritin; lane 3, E1F1; lane 4, F1E1. (B) MALS evaluation of ferritin, E1F1, and F1E1. (C) Size-exclusion (Superose 6) chromatogram for E1F1. (D) Size-exclusion (Superose 6) chromatogram for F1E1. (E) TEM image of E1F1. (F) TEM image of F1E1.

was then eluted with an elution buffer by a linear gradient elution of 0−100% elution buffer in 3 column volumes. The retention time for each protein sample was recorded using an AKTA pure (GE Healthcare, USA).

Thermal Stability Study by Differential Scanning Calorimetry. A MicroCal VP-DSC (Malvern, UK) was used to determine the thermal stability of proteins. Samples were prepared at a protein concentration of 1 mg/mL in 20 mM PB (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. For the differential scanning calorimetry (DSC) measurement, a scan rate of 90 °C per hour was used for a temperature ramp from 30 to 120 °C, and samples were analyzed using the software package Origin 9 (OriginLab Corporation, USA).

Thermal, pH, and Chemical Denaturant Stability Study by Circular Dichroism. A circular dichroism (CD) Chirascan V100 (Applied Photophysics, UK) was used to

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

analyze the secondary structure of the protein samples (0.3 mg/mL in 20 mM PB, pH 9.0). Each sample was scanned three times. For the thermal stability study, samples were equilibrated at 25, 70, 80, or 90 °C for 10 min. The signal from 190 to 260 nm was scanned. CD signals at a wavelength of 222 nm were recorded as a measurement of helix content. For the pH stability study, protein samples (1 mg/mL in 20 mM PB, pH 7.0) were adjusted from pH 7.0 to pH 4.0 by adding an acid (1.0 M acetic acid). Samples were equilibrated at each pH condition for 2 min. The CD signal from 190 to 260 nm was scanned. For chemical denaturant stability studies, protein samples (1 mg/mL in 20 mM PB, pH 7.0) were mixed with guanidinium chloride (GdmCl) solution (6 M) and 20 mM PB (pH 7.0) to reach final concentrations of 0, 0.5, 1.0, 1.5, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 4.0, 4.5, and 5 M GdmCl. Mixed solutions were equilibrated for 2, 10, or 20 min to measure the CD of 222 nm and 20 °C for 300 s.

RESULTS AND DISCUSSION

Characterization of Engineered Ferritins. The molecular weights for EBNA1−ferritin (E1F1) and ferritin−EBNA1 (F1E1) subunits measured by SDS-PAGE were both approximately 23 kDa, as expected based on the amino acid sequence (23.35 kDa) by ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss), higher than that for ferritin (21 kDa) due to inserted epitope and linker (Figure 3A). The hydrodynamic diameters for engineered ferritins E1F1 and F1E1 measured by SEC-MALS were 15.2 and 13.3 nm, respectively, larger than that of native ferritin at 13.2 nm (Figure 3B). In SEC experiments, E1F1 showed a smaller elution volume of 12.9 mL compared to F1E1 (13.9 mL) (Figure 3C,D), consistent with the hydrodynamic diameter results by dynamic light scattering (E1F1: 15.3 nm; F1E1: 13.5 nm). These results proved that both E1F1 and F1E1 presented *flip* structures. For N-terminus insertion, epitopes were exposed outside the structure, while for C-terminus insertion, epitopes were located inside the ferritin cage. Transmission electron microscopy (TEM) results did not reveal any differences regarding the particle structures for E1F1 and F1E1. Both were characterized as well-assembled particles having a monodispersed structure (Figure 3E,F).

Hydrophobicity Study by Simulation and Experiments. To understand the impact of insertion sites on the properties of the engineered ferritin, hydrophobicity was compared between E1F1 and F1E1. Hydrophobicity was first examined by computational simulation. Simulation results for hydrophobic SASA shown in Table 1 indicated that hydrophobic SASA was considerably related to protein molecular weight.

Table 1. Hydrophobicity-Related Parameters from Experiments and Simulations[a]

molecule	PDB ID	molecular weight (kDa)	retention time by experiments (min)	average hydrophobic SASA by simulation (nm^2)	
BSA	3V03	66.0	20.3	283.9	0.19
native ferritin	2FHA	504	24.5	681.9	0.37
E1F1	N/A	560.4	29.6	866.3	0.29
F1E1	N/A	560.4	26.0	827.5	0.25

[a]The PDB files for E1F1 and F1E1 were structured from the DS.

BSA showed the least hydrophobic SASA (BSA: 284 nm^2; ferritin: 682 nm^2; E1F1: 866 nm^2; F1E1: 828 nm^2). E1F1 (N-terminus insertion) was the most hydrophobic of the tested proteins as its hydrophobic SASA was the largest (Figure 4B). Simulations also showed that the hydrophobic SASA of F1E1 was smaller than that of E1F1. Experimentally, four proteins (BSA, ferritin, E1F1, and F1E1) were loaded to a HiTrap Butyl FF column individually and then eluted under the same conditions. A longer retention time indicates a stronger binding with the resin and thus higher protein hydrophobicity. As shown in Figure 4A, E1F1 (N-terminus insertion) showed the highest hydrophobicity, with the retention time at 29.6 min, followed by F1E1 (C-terminus insertion) at 26.0 min (Table 1). The addition of EBNA1 antigens to the termini of ferritin increased total hydrophobicity.

Simulation results for hydrophobic SASA accorded well with experimental results. It was found that the addition of epitopes increased the hydrophobicity of engineered ferritins compared to native ferritin. This is due to inserted epitopes and linkers consisting of the hydrophobic amino acids proline (P), valine (V), glycine (G), alanine (A), and phenylalanine (F). Both simulation and experiments revealed that E1F1 (N-terminus) was the most hydrophobic of the tested proteins. Although the theoretical molecular weights for E1F1 and F1E1 were the same, N- and C-terminus insertions impact differently with regard to hydrophobicity as the space allocation of the same epitopes is quite different. As has been mentioned above, both E1F1 and F1E1 presented in ferritin *flip* structures, where N-terminus insertion locates epitopes outside the ferritin cage and C-terminus insertion locates epitopes inside the ferritin cage. On this account, N-terminus insertion was more hydrophobic than C-terminus insertion. These results also confirmed that computational determination of protein hydrophobic SASA predicted well the practical protein hydrophobicity.

Thermal Stability Study by Simulation and Experiments. Thermally stable proteins are expected to have low α-helix content loss[21] and less fluctuating regions[24,42,54] and are expected to be more compact.[24] The amount of α-helix content loss reveals protein secondary structure changes. Less α-helix content loss suggests a more stable secondary structure.[21,25] A lower RMSD indicates a more thermally stable protein structure.[24,54] A lower R_g shows a more compact protein structure and higher stability.[24] Simulations of the monomer of ferritin, E1F1, and F1E1 at defined temperatures (25, 70, 80, and 90 °C) were run to investigate the difference in thermal stability by analysis of α-helix content, RMSD, and R_g of the central carbon atom (C_α). The ferritin monomer presented the highest stability over all the tested temperatures. After 100 ns, the helical content reduction at 90 °C was minimal compared with that at 25 °C (Figure 5A). High temperature impacted less on the E1F1 monomer (N-terminus insertion) in terms of helical content change; the E1F1 monomer only lost 4% helical content at 90 °C compared to the F1E1 monomer (23% helical content loss) (Table S5, Supporting Information). The RMSD results highlight that the stability of the F1E1 monomer fluctuated more wildly with increasing temperature than that of the E1F1 monomer (Figure S4A2,A3, Supporting Information). The R_g results for E1F1 and F1E1 monomers were similar except at 90 °C, where the F1E1 monomer started to be less compact than the E1F1 monomer (Figure S4B2,B3, Supporting Information). Simu-

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

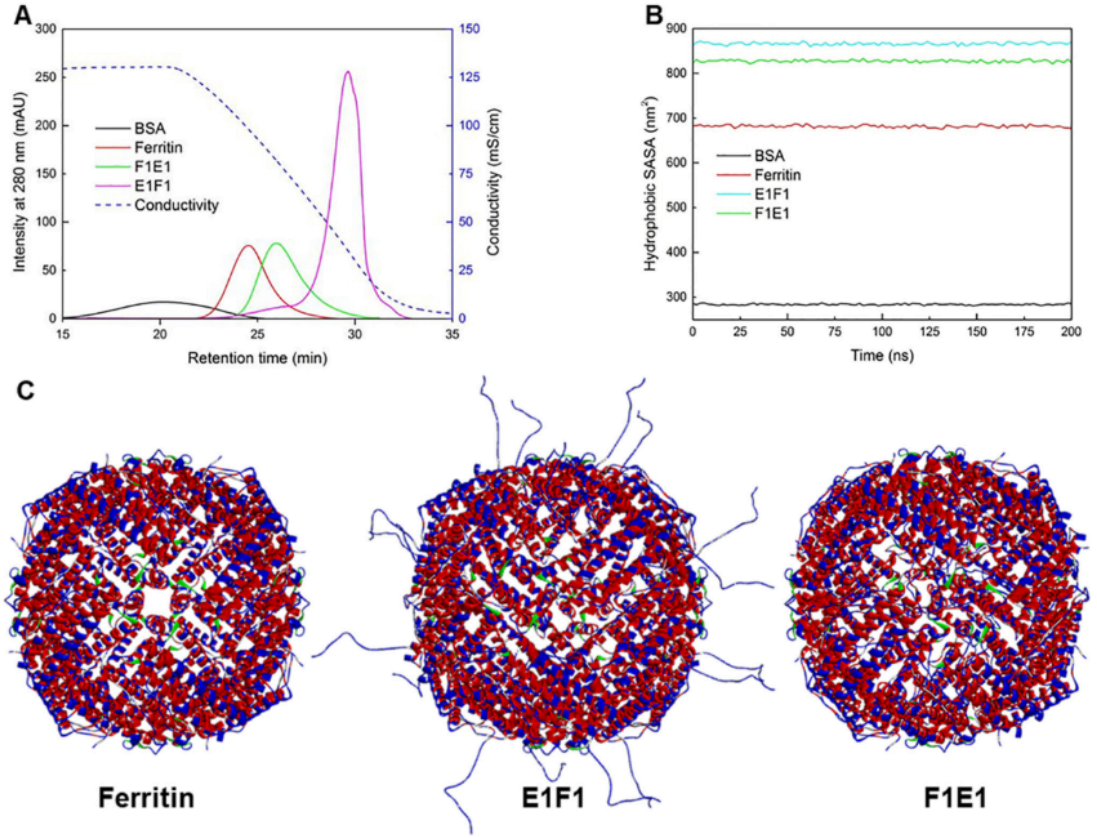


Figure 4. Hydrophobicity of engineered and eluted native ferritins. (A) BSA (black), native ferritin (red), E1F1 (magenta), and F1E1 (green) from a HiTrap Butyl FF hydrophobic interaction column. (B) Computed hydrophobic SASA for BSA (black), ferritin (red), E1F1 (blue), and F1E1 (green) against time at 25 °C using GROMACS. (C) Hydrophobicity diagram by DS 2019: hydrophobic (red); hydrophilic (blue).

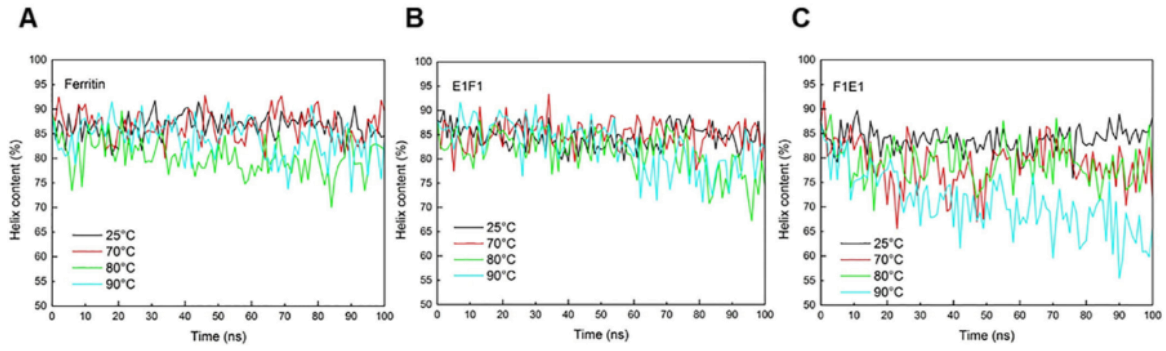


Figure 5. α-Helix content for monomers of ferritin, E1F1, and F1E1 using GROMACS over a temperature range from 25 to 90 °C. (A) One subunit of ferritin. (B) One subunit of E1F1. (C) One subunit of F1E1.

lation results suggest that N-terminus insertion of the monomer was more stable than C-terminus insertion at 90 °C.

To consider interactions among each monomer, trimer structures for ferritin-associated proteins were also simulated. Helical content decreased due to increasing temperature and displayed a similar trend for all three trimer structures, ferritin, E1F1, and F1E1 (Figure 6). Unlike the trend for the monomer structure of F1E1, the trimer structure enhanced the protein secondary structure's ability to resist high temperatures. There was a minimal helical content decrease for the F1E1 trimer compared with the F1E1 monomer at high temperatures up to 90 °C. It was also observed that at 90 °C, the F1E1 trimer had less helical content loss than the E1F1 trimer, 3% compared to 10% (Table S6, Supporting Information). This result indicated that the high temperature (90 °C) impacted less on the secondary structure of the F1E1 trimer (C-terminus insertion) compared to the E1F1 trimer (N-terminus insertion). It was also found that except for ferritin, RMSD variance for trimer structures was similar to that of monomer structures (Figures 6 and S2). This suggests that under all tested temperatures for

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

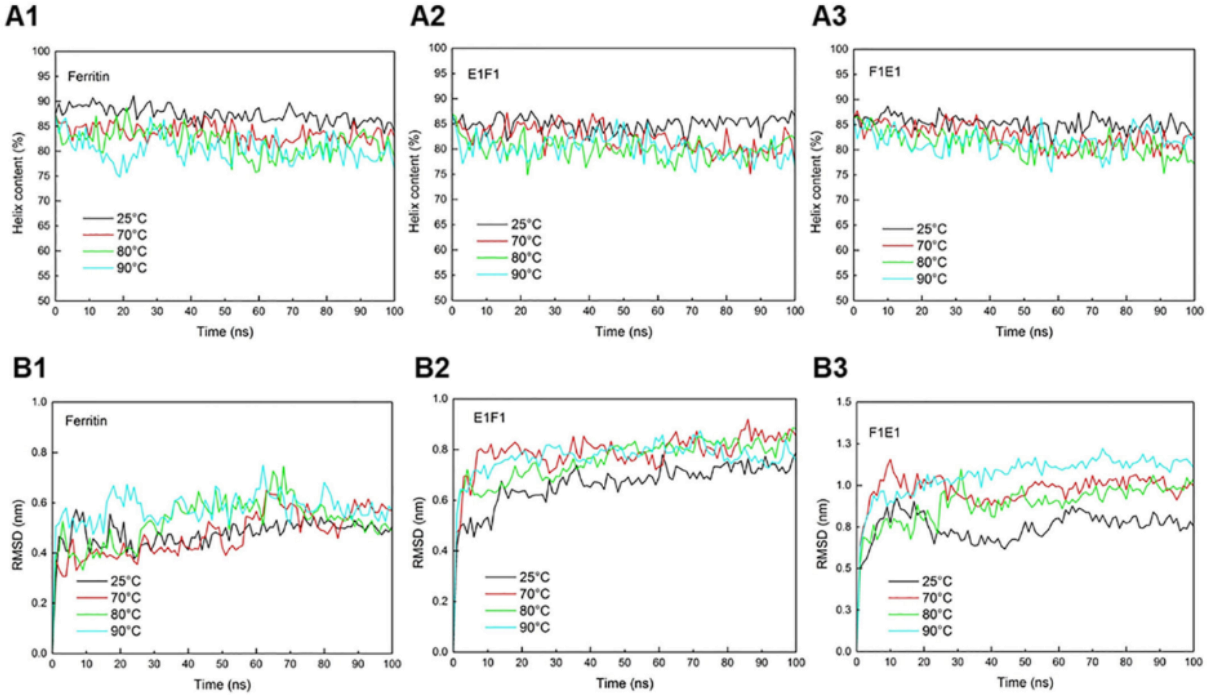


Figure 6. α-Helix content and RMSD for trimer structures of ferritin, E1F1, and F1E1 using GROMACS over a temperature range of 25−90 °C. (A1) Ferritin α-helix content. (A2) E1F1 α-helix content. (A3) F1E1 α-helix content. (B1) Ferritin RMSD. (B2) E1F1 RMSD. (B3) F1E1 RMSD.

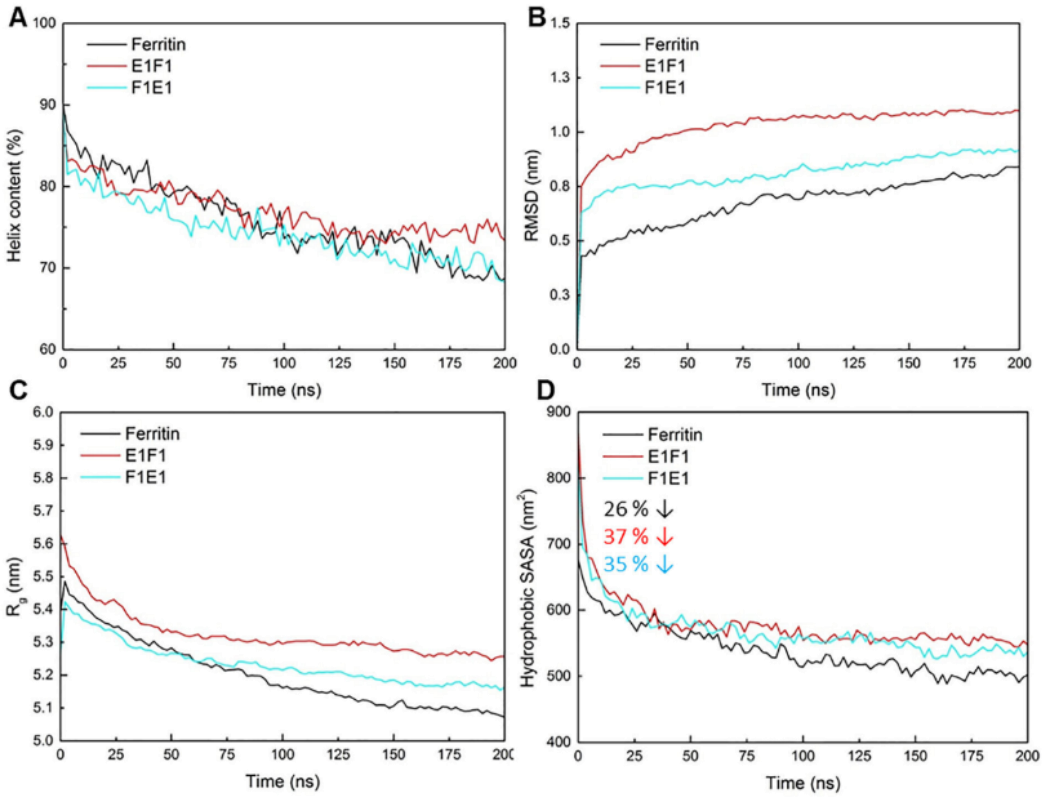
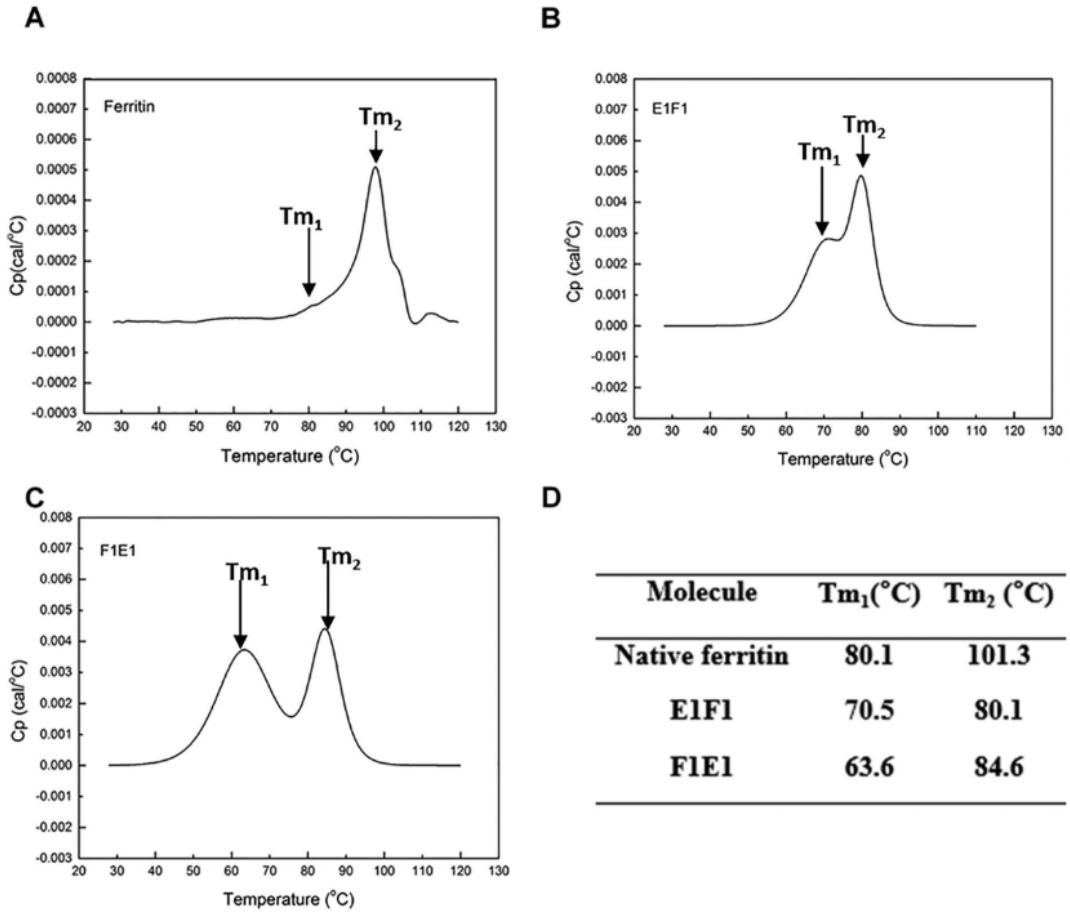


Figure 7. Simulation data for assembled structures of ferritin, E1F1, and F1E1 using GROMACS at 150 °C for 200 ns. (A) α-Helix content. (B) RMSD. (C) R_g. (D) Hydrophobic SASA; ferritin dropped 26%, E1F1 dropped 37%, and F1E1 dropped 35%, compared to the initial structure.

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842



Molecule	Tm_1(°C)	Tm_2 (°C)
Native ferritin	80.1	101.3
E1F1	70.5	80.1
F1E1	63.6	84.6

Figure 8. DSC results for ferritin, E1F1, and F1E1 (protein concentration: 1 mg/mL; sample buffer: 20 mM PB; pH 7.0). (A) Native ferritin. (B) E1F1. (C) F1E1. (D) Tm_1 and Tm_2 values for native ferritin, E1F1, and F1E1 from DSC.

100 ns, there were measurable thermal effects on secondary structures of E1F1 and F1E1. However, regarding the structural stability in terms of RMSD, the comparisons between E1F1 and F1E1 were not obvious.

To support a better understanding of the thermal effect on protein assembly structures and more observable thermal stability comparisons, 24 subunits of ferritin, E1F1, and F1E1 were simulated at a relatively high temperature of 150 °C, well above the reported denaturation temperature of ferritin at 102 °C.[55] Due to computation capacity limitations, an extreme condition was proposed as 150 °C for 200 ns.[21,56] The helical content losses for 24 subunits of ferritin, E1F1, and F1E1 were 21, 16, and 21%, respectively (Figure 7A). The simulation results also indicated that the three assembly proteins were still integrated after 200 ns. Compared with trimer structures, thermal stabilities for all three assembly proteins were improved as the secondary bonds in all three proteins survived the thermal effect to preserve protein tertiary structures. The assembly of protein subunits is driven by secondary bonds, such as noncovalent interactions, including hydrogen bonding, electrostatic association, hydrophobic effects, and van der Waals forces.[57,58] It was also observed that thermal stabilities for each protein were different at 150 °C regarding their backbone RMSD. Figure 7B reveals that ferritin was the most thermally stable over 200 ns with the lowest RMSD, and F1E1 was found to be more thermally stable than E1F1, indicated by a lower RMSD obtained. RMSD related to the ferritin core part without consideration of linkers and epitopes was also calculated to determine how fusion proteins E1F1 and F1E1 were perturbed compared to wild-type ferritin (Figure S3, Supporting Information). It showed that the RMSDs for core parts of E1F1 and F1E1 were similar but higher than that of wild-type ferritin. It demonstrates the less stable core parts of E1F1 and F1E1 than wild-type ferritin. R_g results (Figure 7C) also indicated that F1E1 had a more compact structure than E1F1 at 150 °C. Computed hydrophobic SASA for each protein decreased against time (Figure 7D). This indicates that since simulation started, less hydrophobic surface area was exposed to water. The hydrophobic SASA for E1F1 was found to be lost the most as 37%. To conclude, MD simulation-predicted native ferritin presents the best thermal stability, while C-terminus insertion to ferritin showed better thermal stability against high temperatures up to 150 °C than N-terminus insertion.

To validate this simulation prediction, thermal stability was characterized by experiments using DSC. As shown in Figure 8A, the first peak (Tm_1) suggested where the protein structure began to partially change but will integrate the structure remaining. After the second peak (Tm_2), protein structures were completely destroyed. The Tm_1 range for three proteins

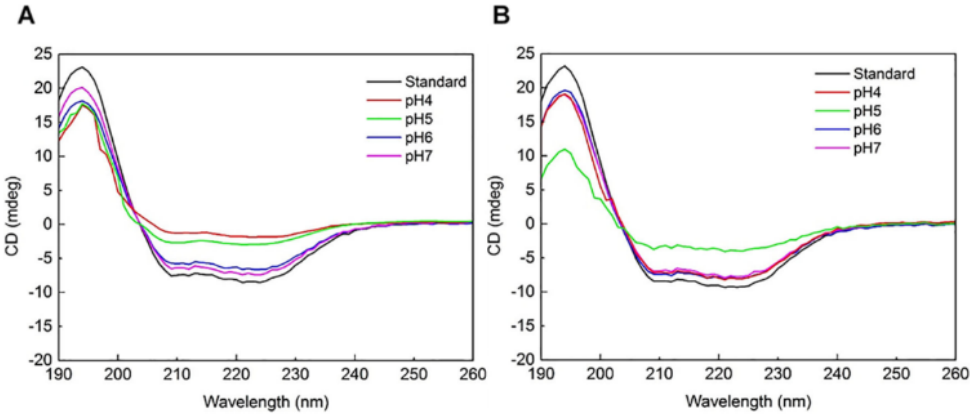


Figure 9. CD spectra of E1F1 (A) and F1E1 (B) (sample concentration: 1 mg/mL; sample buffer: 20 mM PB, pH 7.0). The spectra range shows that the secondary structures of E1F1 are completely changed at pH 5. F1E1 is stable over the tested pH, except at pH 5. The standard curve represents the signal measurement for tested proteins without pH adjustments.

was around 63.6−80.1 °C (Figure 8A−C). Native ferritin had robust thermal stability, and the structure was completely denatured above 100 °C (Figure 8A). In comparison with native ferritin, adding epitopes decreased the maximum tolerated temperature. For both E1F1 and F1E1, the proteins were completely disassembled below 90 °C (Figure 8B,C). The denaturation temperature for C-terminus insertion was slightly higher than that for N-terminus insertion (Tm_2 values 85 and 80 °C, respectively). The experimental results are well aligned with the MS simulation result; C-terminus insertion showed better thermal stability than N-terminus insertion.

For detailed investigation regarding protein secondary structure changes, four temperatures (25, 70, 80, and 90 °C) were selected based on DSC results. CD showed that the helical content of proteins was connected to the secondary structural changes of ferritin, E1F1, and F1E1 at defined temperatures (Figure S4, Supporting Information). Both E1F1 and F1E1 showed less tolerant thermal stabilities than native ferritin. The overall decrease of the secondary structure of helix content for ferritin is 15% at 90 °C, while E1F1 and F1E1 lost 25 and 29% helix content, respectively (Table S4, Supporting Information). N-terminus insertion displayed preserved helicity until 80 °C. After 80 °C, there was a dramatic helical content reduction. In contrast, C-terminus insertion appeared to have weaker helicity above 70 °C. All these results from CD are consistent with the results from DSC and indicate that the engineered ferritin is slightly less thermally tolerant than native ferritin for temperatures up to 70 °C.

Overall, both molecular simulation and experimental results clearly confirmed a change of the insertion site on an engineered ferritin and impacted significantly on the thermal stability of the protein. Based on the simulation results, the E1F1 monomer showed similar thermal properties to the native ferritin monomer in terms of helical content and the RMSD value up to 90 °C. In comparison, the F1E1 monomer showed less thermal stability and major helix content loss. However, the thermal stability for the trimer structure of F1E1 was enhanced. At temperatures in the range of 25−90 °C, the trimer structures of the native ferritin, E1F1, and F1E1 all presented similar trends regarding the helix content change. It was found that the assembly structure of the protein could significantly increase resistance to high temperature. Engineered ferritins and native ferritin could still retain their structures at 150 °C for 200 ns in simulation. Considering only terminus insertions, E1F1 showed slightly stronger helicity than F1E1 at 150 °C for 200 ns, which was consistent with CD results. On the other hand, F1E1 showed lower fluctuated regions and better structure stability than E1F1. This could explain that the tertiary structure of F1E1 presented better thermal stability in DSC results. By comparing the hydrophobic SASA for these two proteins, E1F1 had a larger hydrophobic SASA than F1E1, which implies that more hydrophobic side chains are exposed to water and an unfolding process is favored, thus resulting in weaker thermal stability. Hence, C-terminus insertion was more thermally stable than N-terminus insertion when considering the assembled structure.

Effect of pH Change on Engineered Ferritins. To investigate the effect of pH on engineered ferritins, equal amounts of purified E1F1 and F1E1 were adjusted to pH 4, 5, 6, and 7, and all samples were inculcated in CD for 5 min. By observing the band at wavelengths of 210−230 nm, F1E1 showed an obvious shift only at pH 5, whereas E1F1 has two obvious curve shifts at pH 4 and pH 5 (Figure 9A,B). The theoretical pI for E1F1 and F1E1 is approximately 5.13; therefore, at pH 5, both E1F1 and F1E1 secondary structures shifted because of the protein isoelectric point (pI) effect. At pH equal to pI, purified proteins have zero net charge and the attraction force becomes dominant, resulting in proteins that are more likely to aggregate and precipitate.[59] Nevertheless, F1E1 was more tolerant to a lower pH than E1F1 as the secondary structure of E1F1 was completely changed from pH 5, while F1E1 was able to maintain the secondary structure at pH 4. For N-terminus insertion, epitopes are located on the surface of the ferritin particle. The aggregation of epitopes is possible for E1F1 at pH 4. C-terminus insertion buries the epitopes inside the ferritin cage and thus protects them from outside low-pH conditions. Consequently, C-terminus insertion is expected to be more tolerant to a lower pH than N-terminus insertion.

Effect of Chemical Denaturants on Engineered Ferritins. To test the stability of E1F1 and F1E1 against chemical denaturation, CD spectra at a wavelength of 222 nm were recorded with varying guanidinium chloride (GdnHCl) concentrations. The samples of all three proteins were equilibrated with GdnHCl solution for 20 min. Native ferritin

https://doi.org/10.1021/acs.jpcb.1c00276
J. Phys. Chem. B 2021, 125, 3830−3842

was used as the reference. Native ferritin was highly chemically stable. Ferritin lost 50% of its native structure at 5 M GdnHCl (Figure 10).

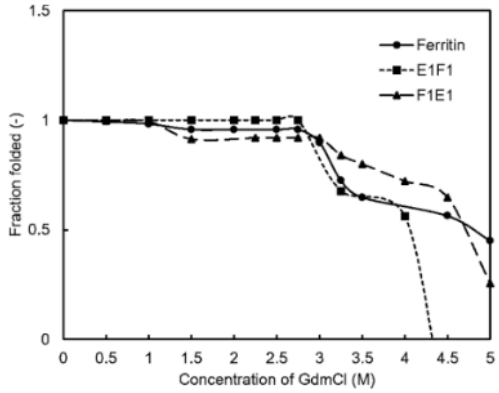


Figure 10. CD spectra of ferritin, E1F1, and F1E1 with varied guanidinium chloride concentrations from 0 to 5 M at 20 °C for 20 min.

Compared with ferritin, both E1F1 and F1E1 were less tolerant to chemical denaturants. Transition of E1F1 occurred from 3 M GdnHCl. Fifty percent of the native structures in E1F1 and F1E1 was lost at 4.25 and 4.60 M GdnHCl, respectively. It could be clearly observed that F1E1 showed better chemical stability than E1F1. E1F1 was completely unfolded at 4.5 M GdnHCl, whereas for F1E1 and ferritin, folded structures still existed at pH 4.5. There is still limited knowledge about the mechanism of how the chemical denaturant GdnHCl interacts with ferritin, F1E1, and E1F1 by binding to the specific sites or by altering the interactions between protein and water molecules. One possible explanation for the different behaviors toward chemical denaturants by E1F1 and F1E1 is the hydrophobic effect.[60] The smaller hydrophobic patch in F1E1 (shown in Table 1) and also the simulation result of hydrophobic SASA encourage the protein to interact with water molecules, which moves toward a low entropy folded state. In contrast, for protein E1F1, as the EBNA1 epitope and soft linkers are mainly composed of hydrophobic amino acids (Pro, Val, Gly, Ala, and Phe), they were completely exposed outside the ferritin cage, which contributed a larger amount of hydrophobic side chains to the surrounding solvent than ferritin and E1F1. The contribution of denaturant co-solvents to this hydrophobic effect was even stronger. It has been found from both MD simulation and experiments by England et al.[61] that guanidinium in particular interacts with the hydrophobic regions of a protein through its flat and nonpolar surface to reduce the unfavorableness of its exposure to the solvent. Hence, larger hydrophobic regions drive a stronger interaction between guanidinium and proteins. In the end, C-terminus insertion is more stable than N-terminus insertion against the chemical denaturant GdnHCl.

CONCLUSIONS

The insertion site of epitopes in ferritin as an engineered nanovaccine plays an important role in the modified protein's stability and performance. In this study, two insertion sites in ferritin were considered. E1F1 was produced by inserting the epitope and soft linkers at the N-terminus of ferritin where epitopes were on the surface of nanoparticles, while F1E1 inserted epitopes at the C-terminus located inside the ferritin cage. Via a combined study of simulation and experiments, we demonstrated that N-terminus insertion (E1F1) and C-terminus insertion (F1E1) have different protein properties. Both simulation and experimental data suggested that engineered ferritin with N-terminus insertion was more hydrophobic and less stable than that of C-terminus insertion. The less hydrophobic SASA provided F1E1 less interactions between nonpolar residues with water molecules. Therefore, a more tolerant pH and chemical denaturant stability was obtained. This work improves our understanding on ferritin nanoparticles as a vaccine carrier. It also provides a method by combining molecular simulation and experiments to predict engineered protein stability with epitope insertion, which potentially shortens the development process for nanoparticle protein-based vaccine design.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c00276.

Stability comparisons for the monomer of E1F1 with different lengths of linkers, simulation results from GROMACS for one subunit of engineered ferritins (E1F1 and F1E1), CD spectra of engineered ferritins, cubic box dimensions for constructed protein structures, protein sequences, CD, and simulation results of helical content of engineered ferritins (PDF)

AUTHOR INFORMATION

Corresponding Author

Jingxiu Bi − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0002-9152-0786; Email: jingxiu.bi@adelaide.edu.au

Authors

Yiran Qu − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0002-3536-925X

Lijie Wang − *Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;* orcid.org/0000-0003-0743-7988

Shuang Yin − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Bingyang Zhang − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Yan Jiao − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0003-1329-4290

Yan Sun − *Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;* orcid.org/0000-0001-5256-9571

Anton Middelberg − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c00276

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide. The authors acknowledge Dr. Fabien Voisin from the HPC support team for providing useful input to quickly obtain the desired results for analysis and processing of big data. The DS team provided the software support. The authors also thank Adelaide Microscopy Centre for assisting to access TEM. Prof. Songping Zhang and Prof. Yongdong Liu from the Institute of Process Engineering, Chinese Academy of Sciences, provided support in molecular characterization by CD and DSC. Y.Q. thanks Lukas Gerstweiler as a general reviewer to provide editing feedback.

■ ABBREVIATIONS

AS, ammonium sulfate; BSA, bovine serum albumin; CD, circular dichroism; CTL, cytotoxic T-lymphocytes; DSC, differential scanning calorimetry; EBV, Epstein−Barr virus; EBNA1, Epstein−Barr nuclear antigen 1; E1F1, ferritin with insertion of EBNA1 at the N-terminus; EV71, Enterovirus 71; F1E1, ferritin with insertion of EBNA1 at the C-terminus; GdnHCl, guanidinium chloride; MALS, multiple-angle light scattering; MD, molecular dynamics; pI, isoelectric point; R_g, radius of gyration; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SASA, solvent accessible surface area; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SEC, size exclusion chromatography; HA, hemagglutinin; HIC, hydrophobic interaction chromatography; TEM, transmission electron microscopy

■ REFERENCES

(1) Angelov, B.; Angelova, A.; Filippov, S. K.; Narayanan, T.; Drechsler, M.; Štěpánek, P.; Couvreur, P.; Lesieur, S. DNA/Fusogenic Lipid Nanocarrier Assembly: Millisecond Structural Dynamics. *J. Phys. Chem. Lett.* **2013**, *4*, 1959−1964.

(2) Doll, T. A. P. F.; Raman, S.; Dey, R.; Burkhard, P. Nanoscale assemblies and their biomedical applications. *J. R. Soc., Interface* **2013**, *10*, 20120740.

(3) Feger, G.; Angelov, B.; Angelova, A. Prediction of Amphiphilic Cell-Penetrating Peptide Building Blocks from Protein-Derived Amino Acid Sequences for Engineering of Drug Delivery Nanoassemblies. *J. Phys. Chem. B* **2020**, *124*, 4069−4078.

(4) Joshi, H.; Lewis, K.; Singharoy, A.; Ortoleva, P. J. Epitope engineering and molecular metrics of immunogenicity: A computational approach to VLP-based vaccine design. *Vaccine* **2013**, *31*, 4841−4847.

(5) López-Sagaseta, J.; Malito, E.; Rappuoli, R.; Bottomley, M. J. Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 58−68.

(6) Lua, L. H. L.; Connors, N. K.; Sainsbury, F.; Chuan, Y. P.; Wibowo, N.; Middelberg, A. P. J. Bioengineering virus-like particles as vaccines. *Biotechnol. Bioeng.* **2014**, *111*, 425−440.

(7) Han, J.-A.; Kang, Y. J.; Shin, C.; Ra, J.-S.; Shin, H.-H.; Hong, S. Y.; Do, Y.; Kang, S. Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine* **2014**, *10*, 561−569.

(8) He, D.; Marles-Wright, J. Ferritin family proteins and their use in bionanotechnology. *New Biotechnol.* **2015**, *32*, 651−657.

(9) Harrison, P. M. The structure and function of ferritin. *Biochem. Educ.* **1986**, *14*, 154−162.

(10) Wang, Z.; Xu, L.; Yu, H.; Lv, P.; Lei, Z.; Zeng, Y.; Liu, G.; Cheng, T. Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomater. Sci.* **2019**, *7*, 1794−1800.

(11) Li, C. Q.; Soistman, E.; Carter, D. C. Ferritin nanoparticle technology... A new platform for antigen presentation and vaccine development. *Ind. Biotechnol.* **2006**, *2*, 143−147.

(12) Kanekiyo, M.; Wei, C.-J.; Yassine, H. M.; McTamney, P. M.; Boyington, J. C.; Whittle, J. R. R.; Rao, S. S.; Kong, W.-P.; Wang, L.; Nabel, G. J. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **2013**, *499*, 102.

(13) Sliepen, K.; Ozorowski, G.; Burger, J. A.; van Montfort, T.; Stunnenberg, M.; LaBranche, C.; Montefiori, D. C.; Moore, J. P.; Ward, A. B.; Sanders, R. W. Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **2015**, *12*, 82.

(14) Deshpande, S.; Masurkar, N. D.; Girish, V. M.; Desai, M.; Chakraborty, G.; Chan, J. M.; Drum, C. L. Thermostable exoshells fold and stabilize recombinant proteins. *Nat. Commun.* **2017**, *8*, 1442.

(15) Zhen, Z.; Tang, W.; Todd, T.; Xie, J. Ferritins as nanoplatforms for imaging and drug delivery. *Expert Opin. Drug Delivery* **2014**, *11*, 1913−1922.

(16) Luzzago, A.; Cesareni, G. Isolation of point mutations that affect the folding of the H chain of human ferritin in E. coli. *EMBO J.* **1989**, *8*, 569−576.

(17) Malleshappa Gowder, S.; Chatterjee, J.; Chaudhuri, T.; Paul, K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *Sci. World J.* **2014**, *2014*, 971258.

(18) Strub, C.; Alies, C.; Lougarre, A.; Ladurantie, C.; Czaplicki, J.; Fournier, D. Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochem.* **2004**, *5*, 9.

(19) Argos, P.; Rossmann, M. G.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. Thermal stability and protein structure. *Biochemistry* **1979**, *18*, 5698−5703.

(20) Novák, P.; Havlíček, V. 4—Protein Extraction and Precipitation. In *Proteomic Profiling and Analytical Chemistry*, 2nd ed.; Ciborowski, P., Silberring, J., Eds.; Elsevier: Boston, 2016; pp 51−62.

(21) Schaller, A.; Connors, N. K.; Dwyer, M. D.; Oelmeier, S. A.; Hubbuch, J.; Middelberg, A. P. J. Computational study of elements of stability of a four-helix bundle protein biosurfactant. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 47−58.

(22) Camilloni, C.; Guerini Rocco, A.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys. J.* **2008**, *94*, 4654−4661.

(23) Alizadeh-Rahrovi, J.; Shayesteh, A.; Ebrahim-Habibi, A. Structural stability of myoglobin and glycomyoglobin: a comparative molecular dynamics simulation study. *J. Biol. Phys.* **2015**, *41*, 349−366.

(24) Du, J.; Dong, J.; Du, S.; Zhang, K.; Yu, J.; Hu, S.; Yin, H. Understanding Thermostability Factors of Barley Limit Dextrinase by Molecular Dynamics Simulations. *Front. Mol. Biosci.* **2020**, *7*, 51.

(25) Gu, J.; Tong, H.; Sun, L.; Lin, Z. Molecular dynamics perspective on the thermal stability of mandelate racemase. *J. Biomol. Struct. Dyn.* **2019**, *37*, 383−393.

(26) Jiang, X.; Chen, G.; Wang, L. Structural and dynamic evolution of the amphipathic N-terminus diversifies enzyme thermostability in the glycoside hydrolase family 12. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21340−21350.

(27) Bhadra, P.; Siu, S. W. I. Refined Empirical Force Field to Model Protein−Self-Assembled Monolayer Interactions Based on AMBER14 and GAFF. *Langmuir* **2019**, *35*, 9622−9633.

(28) Tazikeh-Lemeski, E. Binding Free Energy and the structural changes determination in hGH protein with different concentrations of copper ions (A molecular dynamics simulation study). *J. Theor. Comput. Chem.* **2016**, *15*, 1650045.
(29) Zhang, D.; Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Sci. Rep.* **2017**, *7*, 44651.
(30) Kumar, C. V.; Swetha, R. G.; Anbarasu, A.; Ramaiah, S. Computational Analysis Reveals the Association of Threonine 118 Methionine Mutation in PMP22 Resulting in CMT-1A. *Adv. Bioinf.* **2014**, *2014*, 502618.
(31) Gill, M.; McCully, M. E. Molecular dynamics simulations suggest stabilizing mutations in a de novo designed α/β protein. *Protein Eng., Des. Sel.* **2019**, *32*, 317−329.
(32) Chan, M.; Mukund, S.; Kletzin, A.; Adams, M.; Rees, D. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **1995**, *267*, 1463−1469.
(33) Lee, Y.; Lee, J.-j.; Kim, S.; Lee, S.-C.; Han, J.; Heu, W.; Park, K.; Kim, H. J.; Cheong, H.-K.; Kim, D.; Kim, H.-S.; Lee, K. W. Dissecting the Critical Factors for Thermodynamic Stability of Modular Proteins Using Molecular Modeling Approach. *PLoS One* **2014**, *9*, No. e98243.
(34) Kamaraj, B.; Purohit, R. In Silico Screening and Molecular Dynamics Simulation of Disease-Associated nsSNP in TYRP1 Gene and Its Structural Consequences in OCA3. *BioMed Res. Int.* **2013**, *2013*, 697051.
(35) Stănciuc, N.; Banu, I.; Turturică, M.; Aprodu, I. pH and heat induced structural changes of chicken ovalbumin in relation with antigenic properties. *Int. J. Biol. Macromol.* **2016**, *93*, 572−581.
(36) Zheng, F.; Vermaas, J. V.; Zheng, J.; Wang, Y.; Tu, T.; Wang, X.; Xie, X.; Yao, B.; Beckham, G. T.; Luo, H. Activity and Thermostability of GH5 Endoglucanase Chimeras from Mesophilic and Thermophilic Parents. *Appl. Environ. Microbiol.* **2019**, *85*, No. e02079.
(37) Jiang, Z.; You, L.; Dou, W.; Sun, T.; Xu, P. Effects of an Electric Field on the Conformational Transition of the Protein: A Molecular Dynamics Simulation Study. *Polymers* **2019**, *11*, 282.
(38) Maiorov, V. N.; Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J. Mol. Biol.* **1994**, *235*, 625−634.
(39) Sviben, D.; Bertoša, B.; Hloušek-Kasun, A.; Forcic, D.; Halassy, B.; Brgles, M. Investigation of the thermal shift assay and its power to predict protein and virus stabilizing conditions. *J. Pharm. Biomed. Anal.* **2018**, *161*, 73−82.
(40) Chen, Z.; Fu, Y.; Xu, W.; Li, M. Molecular Dynamics Simulation of Barnase: Contribution of Noncovalent Intramolecular Interaction to Thermostability. *Math. Probl. Eng.* **2013**, *2013*, 504183.
(41) Manjunath, K.; Sekar, K. Molecular Dynamics Perspective on the Protein Thermal Stability: A Case Study Using SAICAR Synthetase. *J. Chem. Inf. Model.* **2013**, *53*, 2448−2461.
(42) Pikkemaat, M. G.; Linssen, A. B. M.; Berendsen, H. J. C.; Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng., Des. Sel.* **2002**, *15*, 185−192.
(43) Lobanov, M. Y.; Bogatyreva, N. S.; Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **2008**, *42*, 623−628.
(44) Chakravarty, S.; Varadarajan, R. Elucidation of Factors Responsible for Enhanced Thermal Stability of Proteins: A Structural Genomics Based Study. *Biochemistry* **2002**, *41*, 8152−8161.
(45) Ding, Y.; Cai, Y.; Han, Y.; Zhao, B.; Zhu, L. Application of principal component analysis to determine the key structural features contributing to iron superoxide dismutase thermostability. *Biopolymers* **2012**, *97*, 864−872.
(46) Farrell, P. J. Epstein−Barr Virus and Cancer. *Annu. Rev. Pathol.: Mech. Dis.* **2019**, *14*, 29−53.
(47) Duellman, S. J.; Thompson, K. L.; Coon, J. J.; Burgess, R. R. Phosphorylation sites of Epstein-Barr virus EBNA1 regulate its function. *J. Gen. Virol.* **2009**, *90*, 2251−2259.
(48) He, L.; Zhu, J. Computational tools for epitope vaccine design and evaluation. *Curr. Opin. Virol.* **2015**, *11*, 103−112.
(49) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(50) Berhanu, W. M.; Hansmann, U. H. E. Side-chain hydrophobicity and the stability of $A\beta_{16-22}$ aggregates. *Protein Sci.* **2012**, *21*, 1837−1848.
(51) Erdur, H.; Scholz, V.; Streitz, M.; Hammer, M.; Meisel, C.; Schönemann, C.; Wandinger, K.-P.; Rosche, B. EBNA1 antigen-specific CD8+ T cells in cerebrospinal fluid of patients with multiple sclerosis. *J. Neuroimmunol.* **2016**, *294*, 14−17.
(52) Burrows, S. R.; Rossjohn, J.; McCluskey, J. Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol.* **2006**, *27*, 11−16.
(53) Miles, J. J.; Borg, N. A.; Brennan, R. M.; Tynan, F. E.; Kjer-Nielsen, L.; Silins, S. L.; Bell, M. J.; Burrows, J. M.; McCluskey, J.; Rossjohn, J.; Burrows, S. R. TCRα Genes Direct MHC Restriction in the Potent Human T Cell Response to a Class I-Bound Viral Epitope. *J. Immunol.* **2006**, *177*, 6804−6814.
(54) Sharma, R.; Sagurthi, S. R.; Narahari Sastry, G. Elucidating the preference of dimeric over monomeric form for thermal stability of Thermus thermophilus isopropylmalate dehydrogenase: A molecular dynamics perspective. *J. Mol. Graphics Modell.* **2020**, *96*, 107530.
(55) McNally, J. R.; Mehlenbacher, M. R.; Luscieti, S.; Smith, G. L.; Reutovich, A. A.; Maura, P.; Arosio, P.; Bou-Abdallah, F. Mutant L-chain ferritins that cause neuroferritinopathy alter ferritin functionality and iron permeability. *Metallomics* **2019**, *11*, 1635−1647.
(56) Day, R.; Bennion, B. J.; Ham, S.; Daggett, V. Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding. *J. Mol. Biol.* **2002**, *322*, 189−203.
(57) Chen, H.; Zhang, S.; Xu, C.; Zhao, G. Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions. *Chem. Commun.* **2016**, *52*, 7402−7405.
(58) Stoffelen, C.; Huskens, J. Soft Supramolecular Nanoparticles by Noncovalent and Host−Guest Interactions. *Small* **2016**, *12*, 96−119.
(59) Zellner, M.; Winkler, W.; Hayden, H.; Diestinger, M.; Eliasen, M.; Gesslbauer, B.; Miller, I.; Chang, M.; Kungl, A.; Roth, E.; Oehler, R. Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. *Electrophoresis* **2005**, *26*, 2481−2489.
(60) Pace, C. N. [14]Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in Enzymology*; Academic Press: 1986; Vol. *131*, pp 266−280.
(61) England, J. L.; Pande, V. S.; Haran, G. Chemical Denaturants Inhibit the Onset of Dewetting. *J. Am. Chem. Soc.* **2008**, *130*, 11854−11855.

CHAPTER 4 IMPACT OF FLEXIBLE LINKER LENGTH ON PROTEIN STABILITY OF ENGINEERED FERRITIN AS A VACCINE CARRIER

Statement of Authorship

Title of Paper	Impact of flexible linker length on protein stability of engineered ferritin as a vaccine carrier
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Qu, Y.; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Impact of flexible linker length on protein stability of engineered ferritin as a vaccine carrier. *International Journal of Biological Macromolecules*

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Designed experiment, performed experiments; analysed data; writing manuscript		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth Davey		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	11/10/2021

Name of Co-Author	Yan Sun
Contribution to the Paper	Proof read the manuscript

Signature		Date	20/09/2021

Name of Co-Author	Anton Middelberg		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	13/10/21

Name of Co-Author	Jingxiu Bi		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/2021

Impact of flexible linker length on protein stability of engineered ferritin as a vaccine carrier

Yiran Qu, Kenneth Davey, Yan Sun, Anton Middelberg, Jingxiu Bi*

Y. Qu, A/Prof. Dr. K. Davey A/Prof. Dr. J. Bi

School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide
South Australia 5005 (Australia)

Prof. Dr. A. Middelberg

Division of Research and Innovation, The University of Adelaide, SA 5005 (Australia)

Prof. Dr. Y. Sun
Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072 (China)

Submitted to International Journal of Biological Macromolecules

Abstract

Epitope-based chimeric (EBC) vaccines have received research attention because of demonstrated safety and efficacy. The linker in EBC vaccine, can significantly affect protein and hence vaccine stability. Despite this empirical understanding, there are limited systematic reports as to how protein stability is affected by flexible linker length. Here we report for the first time, the impact of flexible linker length on stability of engineered ferritins in design for a stable EBC vaccine. Four (4) engineered ferritins were synthesized by inserting Epstein-Barr nuclear antigen 1 (EBNA1 or 'E') and flexible linkers (3 or 15 residues) at the N- or C-terminus of human ferritin heavy chain (HFn or 'F'), namely, $E_1L_{15}F_1$, $E_1L_3F_1$, $F_1L_{15}E_1$ and $F_1L_3E_1$. It was found that: 1) long linkers boosted protein stability against thermal and chemical denaturants compared with short linkers, especially for linkers inserted at N-terminus, and; 2) for N-terminus insertion, $E_1L_{15}F_1$ long linker group was significantly less hydrophobic than $E_1L_3F_1$ short linker group. Therefore, design of long flexible linkers for stable protein may be important to ensure epitope-based chimeric vaccine is stable and thus safe and efficacious. Findings will inform future studies as to the importance of considering linker length in the design of EBC vaccines.

KEYWORDS: ferritin, stability, flexible linker

4.1 Introduction

With the present COVID-19 pandemic, there is increasing interest in epitope-based chimeric vaccine development. This is because of advantages including: 1) a broad selection of particle sizes (5 to 100 nm) [1], 2) cost-effective production systems [2], and; 3) highly repetitive structures to induce immune responses [2-4]. Epitope-based chimeric vaccine is a protein-based vaccine that consists of three (3) major parts, namely, 1) nanoparticle protein cage, 2) epitope, and 3) linker.

Protein stability is important because: 1) structural stability of proteins is critical to presentation of antigens on major histocompatibility complex (MHC) to induce potent immune response [5,6], and; 2) stable proteins can be readily expressed and purified [7,8]. A stable protein therefore ensures that epitope-based chimeric vaccine is safe and efficacious.

The linker is a key component in design of epitope-based chimeric vaccine. Its purpose is to provide significant separation between nanoparticle protein cage and epitope to maintain individual structure and hence function. A lack of appropriate linker therefore leads to problems, for example, protein misfolding [9,10], protein aggregation and low yield in protein production [9,11], and/or impaired bioactivity [9,12,13]. The selection of linker directly affects protein stability [9,15]. Therefore, the linker is important in design for stable epitope-based chimeric vaccine.

Linkers are classified into 3 broad types, 1) flexible, 2) rigid, and; 3) cleavable, shown in Figure 4.1. Flexible linkers are designed for protein cages that require a particular degree of movement, or interaction [9]. Compared with flexible linkers, rigid linkers, such as polyproline motifs [16] and α-helical linker [17], separate protein cage and epitopes more efficiently in rigid structures [9]. Rigid linkers have potential to interfere with protein folding [18]. Cleavable linkers are used for releasing free functional domains *in vivo* [9].

Flexible linkers are commonly applied therefore in epitope-based chimeric vaccine, because these: 1) can be readily designed with negligible changes on conformational stability of the protein nanocage, and; 2) provide significant distance between epitope and protein nanocage to function independently [9,19,20].

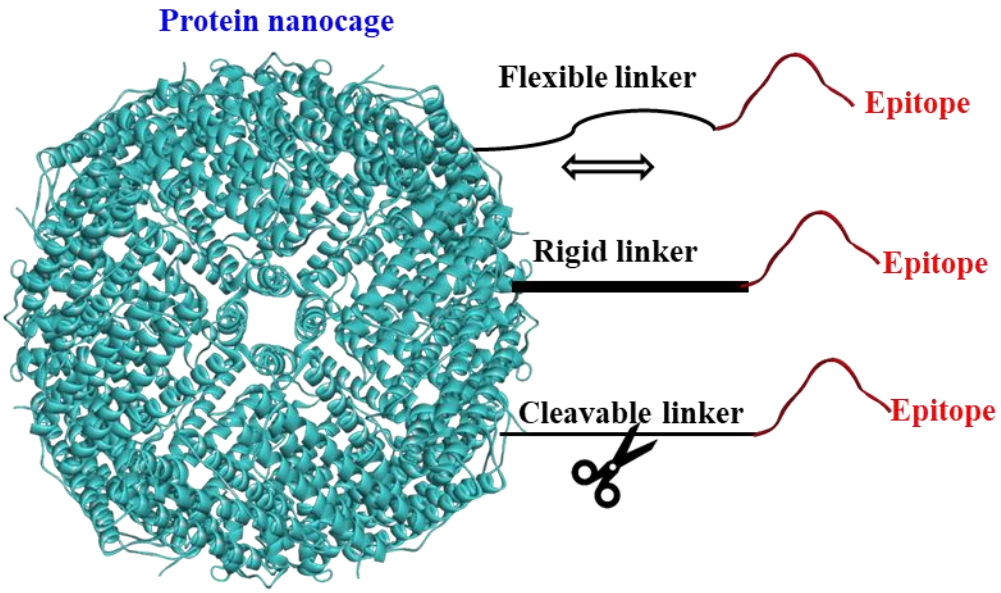


Figure 4.1 Schematic of 3 linker types by using software Discovery Studio 2021 [21]. Human ferritin heavy chain is used as example for the protein nanocage and is marked in blue-colour. Epitope is marked red. Linkers are marked black. 1) *Flexible linker* provides flexible distance between epitope and protein nanocage; 2) *Rigid linker* gives fixed distance between epitope and protein cage, and; 3) *Cleavable linker* releases epitope from protein nanocage.

Amino acid composition and length of flexible linker are two key design parameters. Flexible linkers are synthesized by small non-charged residues [22]. Gly and Ser residues, known as 'GS linker', are widely applied in fusion proteins because of good solubility and flexibility [9]. The property of a GS linker is altered by adjusting glycine content. Gly-rich linkers are highly significantly flexible [23]. Sabourin *et al*. designed a flexible linker $(Gly)_8$ between the open reading frame of a yeast gene and nine (9) Myc epitopes [24]. This type of linker increases accessibility of an epitope to antibodies and/or to improving protein folding [9], [25]. Rosmalen *et al*. reported that Forster resonance energy transfer (FRET) efficiency was overall lower for linkers with less glycine [15]. This finding explained that stiffness of polypeptide linkers

increases with decreasing glycine content. The minimum and maximum linker length, based on recent reports, is between 2 and 31 amino acids. Linker length is commonly selected from 5 to 11 residues [26]. This length needs to be optimal to maintain protein folding and stability. Changes in linker length and composition make a significant difference to protein folding kinetics [27].

There is however a present lack of understanding on the impact of flexible linker length on protein stability. Robinson *et al*. pointed out that a linker with 19 residues synthesized the most stable protein [27]. The addition or deletion of amino acids decreased protein stability. Nagi *et al*. demonstrated that that there is an inverse correlation between linker length and protein stability [28]. Rop protein was used as a model system, and a series of flexible linkers with 1 to 10 Gly residues was applied. They reported that all variants retained functionality. Stability against thermal and chemical denaturation was reduced with increasing flexible linker length. However, experimental results reported by Chen *et al*. were contradictory. This is because it was claimed that a longer flexible linker improves activity and stability of displayed esterase over shorter flexible linker [29]. Lau *et al*. performed molecular dynamics simulation to show that longer flanking linkers provide flexibility for the presentation of rotavirus VP8* antigen module on VP1, resulting in good expression of the modular proteins [30]. The mechanism of impact of linker length on protein folding and stability is still not fully understood. These studies however have underscored that the length of flexible linker is a significant design parameter for the expression of stable folded fusion protein.

In this study, the impact of flexible linker length on protein stability of epitope-based chimeric ferritin-based vaccine was investigated using an experimental approach. Short (GGS) and long (GGSGG GGSGG GGSGG) flexible linkers were applied between the epitope Epstein-Barr

nuclear antigen 1 (EBNA1) and nanoparticle protein cage human ferritin heavy chain (HFn). Two (2) insertion sites were used on human ferritin heavy chain, 1) N-terminus, and; 2) C-terminus, to give four (4) engineered ferritins plus a control, 1) $E_1L_3F_1$ (N-terminus insertion, 3 residues linker), 2) $F_1L_3E_1$ (C-terminus insertion, 3 residues linker), 3) $E_1L_{15}F_1$ (N-terminus insertion, 15 residues linker), 4) $F_1L_{15}E_1$ (C-terminus insertion, 15 residues linker), and; 5) control wild type HFn (without linker and epitope). Detailed information of gene sequences is provided in Table S4.1, Supporting Information. Protein hydrophobicity and protein stability were determined against thermal and chemical denaturants to quantify the impact of flexible linker length.

Findings highlighted that long flexible linkers inserted at N-terminus of HFn significantly reduced protein surface hydrophobicity and boosted protein stability against thermal and chemical denaturants, compared with short flexible linkers. The differences between long and short flexible linker inserted at C-terminus of HFn however were negligible. Therefore, flexible linker design is important to produce stable proteins that ensure safe and efficacious epitope-based chimeric vaccine.

Findings are expected to be immediate interest and benefit to a range of researchers in the design of flexible linkers to produce stable and efficacious vaccines.

4.2 Materials and methods

4.2.1 Expression and purification of engineered ferritin

PET 30a vector was used to synthesize plasmids harboring genes of 4 types of Epstein-Barr virus nuclear antigen-1 (EBNA1) fused HFn, namely, $E_1L_3F_1$, $F_1L_3E_1$, $E_1L_{15}F_1$ and $F_1L_{15}E_1$ and control HFn.

Target genes were inserted between *Nde I* and *BamH I* restriction sites [31]. The plasmids were transformed into *E. coli* BL21 (DE3) [31]. The expression of the vector ferritin and four (4) engineered ferritins were the same, following a similar experimental procedure as previous published work [32].

Cells were collected by centrifugation at 13,751 xg. Cell pellets were suspended in lysis buffer (20 mM phosphate buffer, 2 mM EDTA, pH 7.0) and disrupted by sonication on ice with control temperature below 37 °C (Scientx Sonicator at 360 W: 4 s on and 6 s off for 10 min) [31]. Bacterial lysate was centrifuged at 13,751 xg for 30 min at 4 °C to remove cell debris. Supernatant contained target proteins.

Protein purification for these five proteins was similar [31]. Two steps including heating precipitation and hydrophobic interaction chromatography (HIC) were applied (Table S4.2, Supporting Information). Protein supernatant collected after centrifugation was diluted with Milli-Q water to a final total protein concentration of 5 mg mL^{-1}. NaCl and NaAc-HAc were added to precipitate host cell proteins. pH of HFn, $F_1L_3E_1$ and $F_1L_{15}E_1$ was adjusted with NaAc-HAc (1M) to 4.5, whilst for $E_1L_3F_1$ and $E_1L_{15}F_1$, pH was adjusted to 5.5 and 5.0, respectively. Heat (60 °C, 10 min) was applied to help precipitate impure proteins. After that, centrifugation (19,802 xg at 4 °C for 10 min) was used to collect supernatant. Resultant supernatant was ready

to perform hydrophobic interaction chromatography (HIC).

A HiTrap Butyl FF column (GE Healthcare, USA) was used for HIC. The equilibration buffer of HFn, $F_1L_3E_1$ and $F_1L_{15}E_1$, was 100 mM phosphate buffer, 1.2 M ammonium sulfate (AS), pH 6.5. For $E_1L_3F_1$ and $E_1L_{15}F_1$ the salt concentration of AS in the equilibrated buffer was reduced to, respectively, 0.8 M and 1.0 M. The eluted buffer was the same for all proteins as 20 mM phosphate buffer, pH 6.5. The chromatography procedure using AKTA Pure (GE Healthcare, USA) was exactly same as previous published work [31]. Protein purity level was determined by 12 % (w/v) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

4.2.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS separation and stacking gel layers were prepared as 12 % (w/v) and 5 % (w/v), respectively. Samples preparation was same as previously [31]. Separation was conducted at 200 V for 40 min using PowerPac Universal Power Supply (Bio Rad, USA). The gel was stained with 0.25 % (w/v) Coomassie R-250 for 60 min. The background was washed using a buffer of 10 % (v/v) ethanol (AR) (ChemSupply, AU) and 10 % (v/v) acetic acid (ChemSupply, AU) for 60 min.

4.2.3 Transmission electron microscopy (TEM)

Samples preparation was same as previously [31]. Grids were analysed with a Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

4.2.4 Size exclusion chromatography coupled with multiple-angle laser light scattering (SEC-MALS)

Sample preparation was same as reported previously [31]. Superose 6 10/300 GL (GE Healthcare, USA) was connected to a high-performance liquid chromatography (HPLC) system (Shimazdu, Japan) coupled with Wyatt Optilab refractive index (RI) and Wyatt DAWN MALLS detector (USA). Hydrodynamic radius was determined by the software ASTRA (Wyatt Technology, USA).

4.2.5 Hydrophobicity

Hydrophobicity was determined by hydrophobic interaction chromatography amongst HFn, $E_1L_3F_1$, $F_1L_3E_1$, $E_1L_{15}F_1$ and $F_1L_{15}E_1$ following a similar procedure [31].

Sample volume was adjusted to 2 mL with protein concentration of 1 mg mL^{-1}. Each protein sample was mixed with 8 mL of equilibration buffer of 0.8 M ammonium sulfate buffer, 100 mM phosphate buffer containing Na_2HPO_4 and NaH_2PO_4, pH 6.5. The elution buffer was 20 mM phosphate buffer, pH 6.5. A HiTrap Butyl FF column (GE Healthcare, USA) was connected to AKTA Pure (GE Healthcare, USA) and was equilibrated by the equilibration buffer before loading protein samples. The column was eluted with the elution buffer by linear gradient elution of 0 to 100 % elution buffer in ten (10) column volumes. Retention time for each protein sample was recorded using software Unicorn 7 (GE Healthcare, USA).

4.2.6 Thermal stability by differential scanning calorimetry (DSC)

Nano DSC (TA Instruments, USA) was performed to analyze protein thermal stability. Samples were prepared at protein concentration 5 mg mL^{-1} in 200 mM phosphate buffer (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. Sample volume was 0.3 mL. DSC was equilibrated by the setting of 30

min with scan rate of 90 °C h^{-1} at a temperature ramp from 30 to 110 °C. Data were analysed using software *Nano Analyze* (TA Instruments, USA). Raw data from the experiment were subtracted from the reference buffer scan.

The baseline was fitted and subtracted from the thermogram using a sigmoidal baseline function. Differences in heat capacity of the folded and unfolded states of protein were determined [33]. The midpoint of transition temperatures (T_m) and calorimetric enthalpy (ΔH) were determined at end of analysis. The unit of ΔH is kcal mol^{-1}, where mol is defined as molar quantity for proteins based on equation (4.1).

Molar quantity (mol) = Protein concentration (g mL^{-1}) × Sample volume (mL) / Molar weight (g mol^{-1}) (4.1)

4.2.7 Stability against thermal- and chemical denaturants by fluorescence spectrometer

Fluorescence spectrometry (RF-5301PC, Shimadzu, Japan) was used to determine tertiary protein change against thermal- and chemical denaturants.

To determine thermal stability samples were prepared at protein concentration of 0.1 mg mL^{-1} in 200 mM phosphate buffer (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. An excitation wavelength of 280 nm was used, and emission was recorded from 300 to 600 nm with a scanning rate of 1200 nm min^{-1}, using 1.0 cm path length cuvette. Each protein sample was analyzed at room temperature (RT), 30, 40, 50, 60, 70, 75, 80, 85, 90, 95 and 100 °C for 10 min, and then scanned three (3) times. Data were analyzed using *Origin 2021* [34]. Maximum fluorescence intensity change (MFIC) was computed by equation (4.2):

$$MFIC = (I_2 - I_1)/I_1 \times 100\ \% \quad (4.2)$$

where I_1 = protein maximum fluorescence intensity without treatment (a.u), and I_2 = protein maximum fluorescence intensity following treatment (a.u).

To determine impact of chemical denaturants (GdnHCl and urea) on protein stability, HFn, $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$, and $F_1L_{15}E_1$ were adjusted to the concentration of 0.3 mg mL^{-1} in 20 mM phosphate buffer at pH 7.0. Each 1 mL of protein was treated with varying concentration of GdnHCl, or urea, to give a final concentration of 0.1 mg mL^{-1}; final GdnHCl concentration of 0, 1, 2, 3, 4, 5, 6 M, and; final urea concentration of 0 1, 2, 3, 4, 5, 6, 7 and 8 M, respectively, shown in Table S4.3, Supporting Information. The treated protein sample volume was 3 mL. Treatment time intervals were 0, 1, 2, 4, 8, 12 and 24 h. At 0 h, samples were equilibrated with chemical denaturant (GdnHCl or urea) for 5 min prior to recording spectra. The baseline was recorded by using same concentration of buffer as protein sample and was subtracted from fluorescence values for treated protein samples.

4.2.8 Thermal stability by UV-vis spectrometer

UV-VIS spectrophotometry (UV-2600, Shimadzu, Japan) was used to determine protein quaternary structure change. Wavelength range was set from 185 nm to 300 nm. The slit-width was 1.0 and light source wavelength was 323 nm. Samples were prepared at protein concentration of 0.1 mg mL^{-1} in 200 mM phosphate buffer (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. Background buffer (200 mM phosphate buffer) was scanned as baseline. Absorbance was auto-zero. Each protein sample was treated at room temperature (RT: 28 °C), 70, 75, 80, 85, 90, 95 and 100 °C for 10 min, and then scanned triplicates. Data were analyzed using Origin 2021 (OriginLab Corporation, Northampton, MA, USA).

4.3 Results and discussion

4.3.1 Characterization of engineered ferritin

Characterization of engineered ferritin was used to compare protein structural differences. Human ferritin heavy chain (HFn) was used as a control.

The molecular weight (MW) exhibited in SDS-PAGE for monomer of $E_1L_{15}F_1$ and $F_1L_{15}E_1$ were similar (~ 23 kDa), and were slightly larger than $E_1L_3F_1$ and $F_1L_3E_1$, shown in Figure 4.2A. These findings are consistent with theoretical monomer MW prediction of $E_1L_3F_1$ and $F_1L_3E_1$ at 22.6 kDa and $E_1L_{15}F_1$ and $F_1L_{15}E_1$ at 23.3 kDa. MW of all engineered ferritins were larger than wild-type HFn having MW of 21 kDa. MW differences were also demonstrated by size-exclusion chromatography. The lower the value by elution volume, the greater MW of assembled proteins. As is illustrated in Figure 4.2B (and Table 4.1), assembled $E_1L_{15}F_1$ was eluted with the shortest elution volume suggesting largest size (elution volume: 12.35 mL; hydrodynamic radius: 7.26 nm).

The particle size of engineered ferritin with C-terminus insertion ($F_1L_3E_1$: 6.62 nm; $F_1L_{15}E_1$: 6.65 nm) were consistently less than that for N-terminus insertion ($E_1L_3F_1$: 6.86 nm; $E_1L_{15}F_1$: 7.26 nm). This is likely because the linkers and epitopes are located inside of the ferritin cage when insertion is at the C-terminus. By comparison with a 3-residues linker it was established that MW and particle size with 15 residues linkers were meaningfully increased. Importantly overall, TEM results confirmed that the engineered ferritins were assembled as nanoparticles, Figure 4.2C.

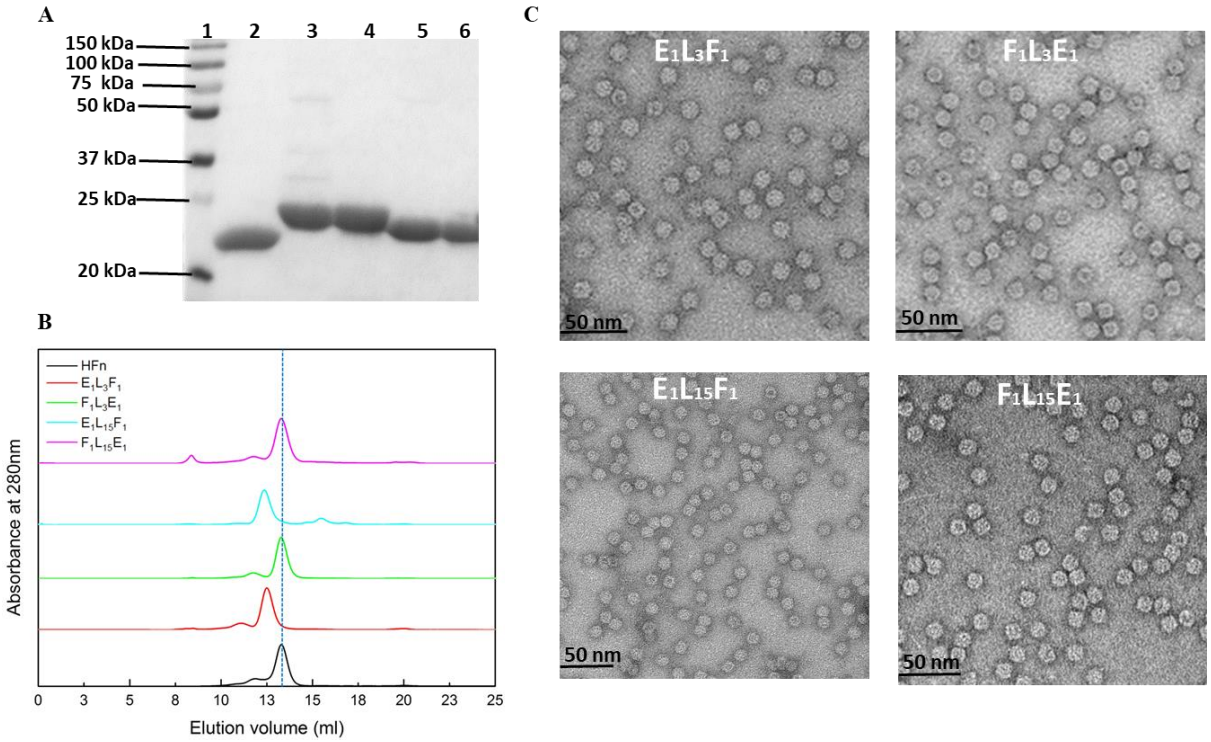


Figure 4.2 Characterization of engineered ferritin. A) SDS-PAGE analysis post-purification: lane 1, marker; lane 2, HFn; lane 3, $E_1L_{15}F_1$; lane 4, $F_1L_{15}E_1$; lane 5, $E_1L_3F_1$; lane 6, $F_1L_3E_1$. B) Size exclusion chromatography Multi Angle Light Scattering (SEC-MALS). C) TEM images.

Table 4.1 Summary comparison of particle size and molecular weight (MW) for engineered ferritins.

Protein	Assembled MW (kDa)	Elution volume of SEC (mL)	Hydrodynamic radius by SEC-MALS (nm)
HFn	504.0	13.30	6.58 (± 0.76 %)
$E_1L_3F_1$	542.4	12.49	6.86 (± 0.86 %)
$F_1L_3E_1$	542.4	13.23	6.62 (± 0.78 %)
$E_1L_{15}F_1$	559.2	12.35	7.26 (± 0.68 %)
$F_1L_{15}E_1$	559.2	13.26	6.65 (± 0.58 %)

4.3.2 Hydrophobicity of engineered ferritin via hydrophobic interaction chromatography (HIC)

To determine the impact of flexible linker length on protein surface hydrophobicity, the five proteins, HFn, $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$, were loaded to HIC and protein surface hydrophobicity differences were compared.

Proteins were eluted by reducing salt concentration. The shorter the retention time before appearance of the elution peak the weaker the protein binding to the column. Therefore, the less hydrophobic the protein.

Figure 4.3 demonstrates that $E_1L_3F_1$ was the most hydrophobic protein. This is evidenced by the longest retention time of 37.3 min, when compared with others; HFn = 31.1, $E_1L_{15}F_1$ = 35.7, $F_1L_3E_1$ = 30.9 and $F_1L_{15}E_1$ = 31.1 min. The intensity signal at 280 nm for $E_1L_3F_1$ was the strongest. This finding underscores that $E_1L_3F_1$ binds the most strongly with HIC, and that there was only minor protein lost compared with others in flow-through.

A comparison of $E_1L_3F_1$ (37.3 min) with $E_1L_{15}F_1$ (35.7 min) found that, for N-terminus insertion, protein hydrophobicity decreased with increasing flexible linker length. This is highly likely a consequence of a steric shielding caused by flexible linkers. The linkers investigated here are flexible polypeptide linkers, which are composed mainly of glycine. Simulations reported by Rozycki *et al.* highlighted that for glycine rich linkers, the longer the linker, the lower is its elastic stiffness [35]. Therefore, a 15-residues linker is more flexible than a 3-residue linker.

As is shown in Figure 4.3B, 3-residues linkers group presents one possible conformation where epitopes and linkers are located on the surface of nanoparticle, together with minor sheltering of its original hydrophobicity.

However, when linker length increased the flexibility of linkers was also significantly increased. Long flexible linkers perform a similar function as polyethylene glycol (PEG). A distinctive hydrated layer may form to hide the surface of the nanoprotein cage, which results in steric shielding [36,37]. Reduced hydrophobicity was found for long linker compared with short linker, Figure 4.3B. However, for C-terminus insertion, both of short and long linkers were located inside the ferritin cage. This results in a minor change on protein surface hydrophobicity. Findings presented as Figure 4.3A confirm that HFn, $F_1L_3E_1$ (30.9 min) and $F_1L_{15}E_1$ (31.1 min) have similar surface hydrophobicity.

For large protein molecules, surface hydrophobicity significantly impacts protein folding and structural stability [38]. Proteins with a larger hydrophobic surface are more likely to aggregate [39], and result in decreasing protein solubility [40]. Strub *et al*. reported that decreasing protein solvent accessible surface area (SASA) increased protein thermal stability [41]. Our research has

shown [31] that protein with stronger hydrophobicity are more likely to be less thermally and chemically stable than others. It is concluded therefore that $E_1L_3F_1$ is the least stable.

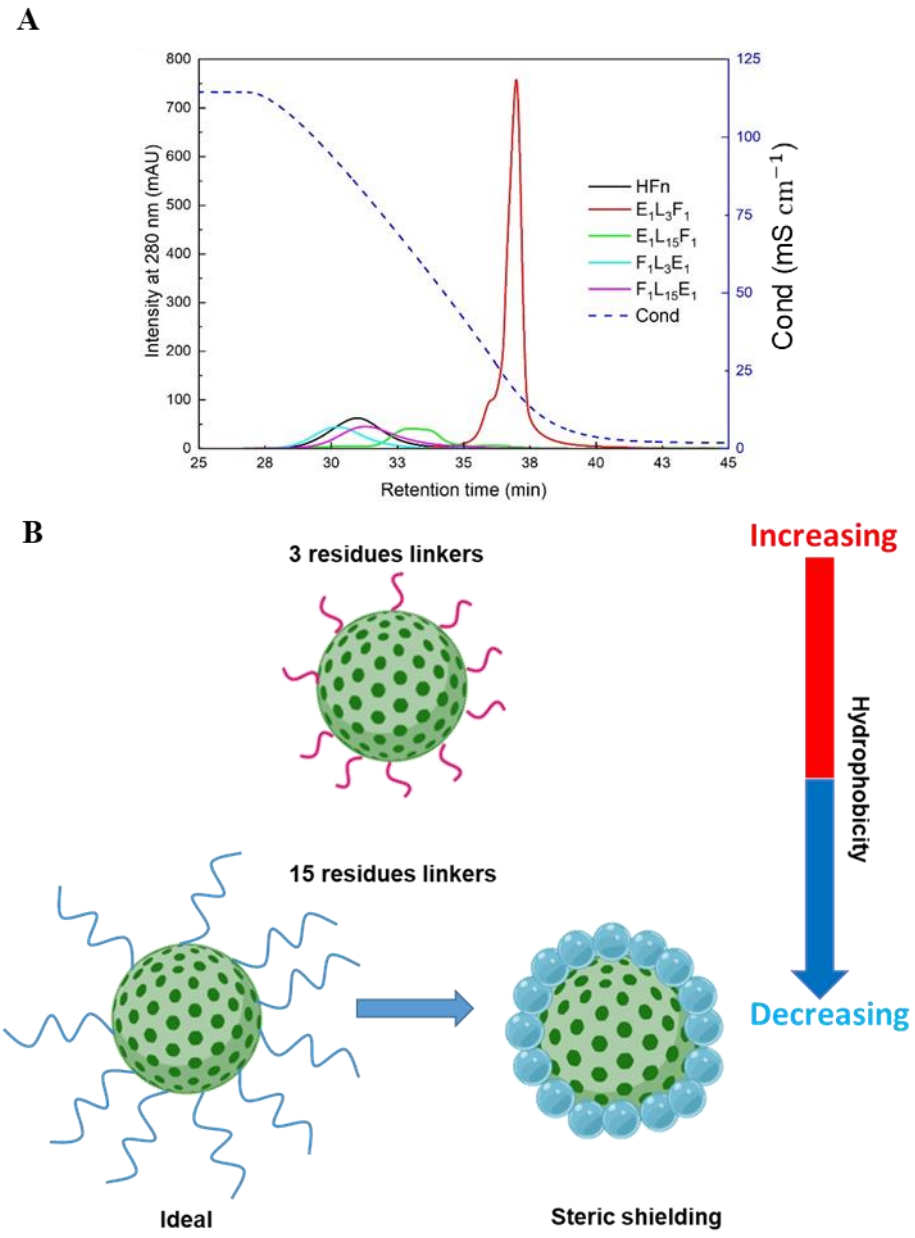


Figure 4.3 Hydrophobicity of engineered ferritin with short, or long flexible linker. A) Hydrophobic interaction chromatography of engineered ferritin: HFn = 31.10, $E_1L_3F_1$ = 37.27, $E_1L_{15}F_1$ = 35.69, $F_1L_3E_1$ = 30.8, and $F_1L_{15}E_1$ = 31.09, min. B) Schematic of short (3 residues: red-colour) and long (15 residues: blue) flexible linkers of proteins.

4.3.3 Thermal stability of engineered ferritin

Determination of thermal stability by differential scanning calorimetry (DSC)

To fully understand the impact of flexible linker length on protein thermal stability, differential scanning calorimetry (DSC) was used to characterise the four engineered ferritins ($E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$) and the HFn control.

DSC characterizes protein thermal stability, conformational change and domain folding integrity via determination of heat capacity as a function of temperature [41,42]. The two main parameters obtained by DSC are the midpoint temperature of the transition (T_m), and calorimetric enthalpy (ΔH). T_m values for proteins are used to predict conformational stability [43]. A decrease in T_m underscores a destabilizing impact on protein structure. T_m can be assumed as the temperature at which a protein denatures [44]. Proteins with greater T_m have stronger thermal stability. ΔH represents the energy consumed to unfold proteins [45]. This value is always positive for proteins because unfolding is endothermic. T_m and ΔH are used to indicate whether a protein undergoes two-state transition or forms intermediates during thermal denaturation [33]. A protein with multiple domains may experience more than one unfolding event, giving multiple T_m values [46,47].

It is assumed that three possible stages are involved for engineered ferritin during DSC thermal denaturation, initial, intermediate and final stage (Figure 4.4A). At initial state, the protein remains 'native-folded'. Following thermal denaturation there are however different intermediates. Figure 4.4A illustrates schematically three (3) different intermediates, 1) minor change; 2) medium change, and; 3) major change. Minor change refers to when the protein begins to be loosen or open up, while remaining essentially as an assembled, folded structure. There is a minor enthalpy change observed representing minor protein conformation change,

normally referring to T_{m1}. Intermediates with medium change begin losing quaternary structure and tertiary structures and disassembling. A more significant enthalpy change is observed than for the minor change (T_{m1}). The conformation of protein subunits becomes disordered. Intermediates with major change almost completely lose all tertiary and secondary structures. This is shown in greater enthalpy change than for the medium change (T_{m2} or T_{m3}). Ultimately, proteins lose all secondary and tertiary structure and unfold. This is seen in the last peak (T_{m2} or T_{m3}) in the DSC thermogram. For a direct comparison of protein thermal stability, T_m^* was focused, which was defined as the protein disassembled temperature when protein starts disassembling by losing quaternary and tertiary structures (medium change).

Reducing linker length destabilized proteins as evidenced by comparing individual T_m^* values. Wild-type HFn was used as a control because it has the most thermally stable structure with T_m^* = 99.1°C and ΔH_2 = 462 kcal mol^{-1} (Figure 4.4B, Table 4.2). This measured enthalpy change is significantly greater than the published value (T_m = 90.9 °C and ΔH = 402.8 kJ mol^{-1} = 96.3 kcal mol^{-1}, for HFn with protein concentration of 1 mg mL^{-1}) [48]. It was found that data associated with the first peak in HFn DSC thermogram (T_{m1} = 81.1 °C and ΔH_1 = 121 kcal mol^{-1}) was similar to the published value. In this thermogram, initial disassembly of the particle at T_{m1} is observed to be followed by protein denaturation at T_{m2}. For the long linker group, $E_1L_{15}F_1$ and $F_1L_{15}E_1$, two transition temperatures T_{m1} and T_{m2} were observed, whilst for the short linker group, $E_1L_3F_1$ and $F_1L_3E_1$, three transition temperatures T_{m1}, T_{m2} and T_{m3} were seen (Figures 4.4C, 4.4D, 4.4E and 4.4F).

For N-terminus insertion, T_m^* for $E_1L_{15}F_1$ was 98.6 °C, compared with $E_1L_3F_1$ (87.9 °C) (Table 4.2). $E_1L_{15}F_1$ (N-terminus insertion) in the intermediate stage exhibited a medium change (Figure 4.4D). Compared with published data for HFn (ΔH = 96.3 kcal mol^{-1}), the enthalpy

change for the first peak was significantly high and was similar to the second peak with ΔH_1 = 322 and ΔH_2 = 367 kcal mol^{-1} (Table 4.2). With a T_{m1} = 91.6 °C, $E_1L_{15}F_1$ was denatured via losing quaternary and tertiary structure. In comparison with $E_1L_{15}F_1$, $E_1L_3F_1$ became looser at 75.5 °C (T_{m1}), with minor enthalpy change ΔH_1 = 96.2 kcal mol^{-1}. It was denatured at 87.9 °C (T_{m2}), and unfolded at 96.8 °C (T_{m3}). It is concluded that $E_1L_{15}F_1$ is therefore more thermally stable than $E_1L_3F_1$, evidenced by a greater protein denaturing temperature.

The C-terminus insertion, $F_1L_{15}E_1$ (Figure 4.4F) exhibited the most similar thermogram to HFn (Figure 4.4B). T_m^* for $F_1L_{15}E_1$ was 93.9 °C, compared with $F_1L_3E_1$ (90.1 °C) (Table 4.2). For $F_1L_{15}E_1$, the enthalpy change of the first peak was minor compared with that of the second peak, ΔH_1 = 80.8 and ΔH_2 = 451 kcal mol^{-1}. T_{m2} for $F_1L_{15}E_1$ was therefore considered as the denaturing temperature T_m^*. $F_1L_3E_1$ initiated conformational change against thermal denaturants at 77.5 °C (T_{m1}) (Figure 4.4E and Table 4.2). The enthalpy change at the first peak was greater than for $F_1L_{15}E_1$ (long linker group) at 189, compared with 80.8 kcal mol^{-1}. At 90.1 °C (T_{m2}) the intermediate for $F_1L_3E_1$ underwent highly significant change and was almost denatured as was demonstrated by a significant enthalpy change of ΔH_2 = 372 kcal mol^{-1}. The denatured protein was unfolded at 96.2 °C (T_{m3}). $F_1L_{15}E_1$ therefore has greater thermal stability than $F_1L_3E_1$, however the denaturing temperature difference is less than that for the N-terminus insertion.

It is concluded that the protein with strongest hydrophobicity, $E_1L_3F_1$, is least thermally stable. Additionally, long flexible linkers stabilize proteins against thermal denaturants. However, because these results from DSC do not permit conclusive findings of exact quaternary and tertiary structural change in proteins, an additional protein thermal stability investigation was undertaken.

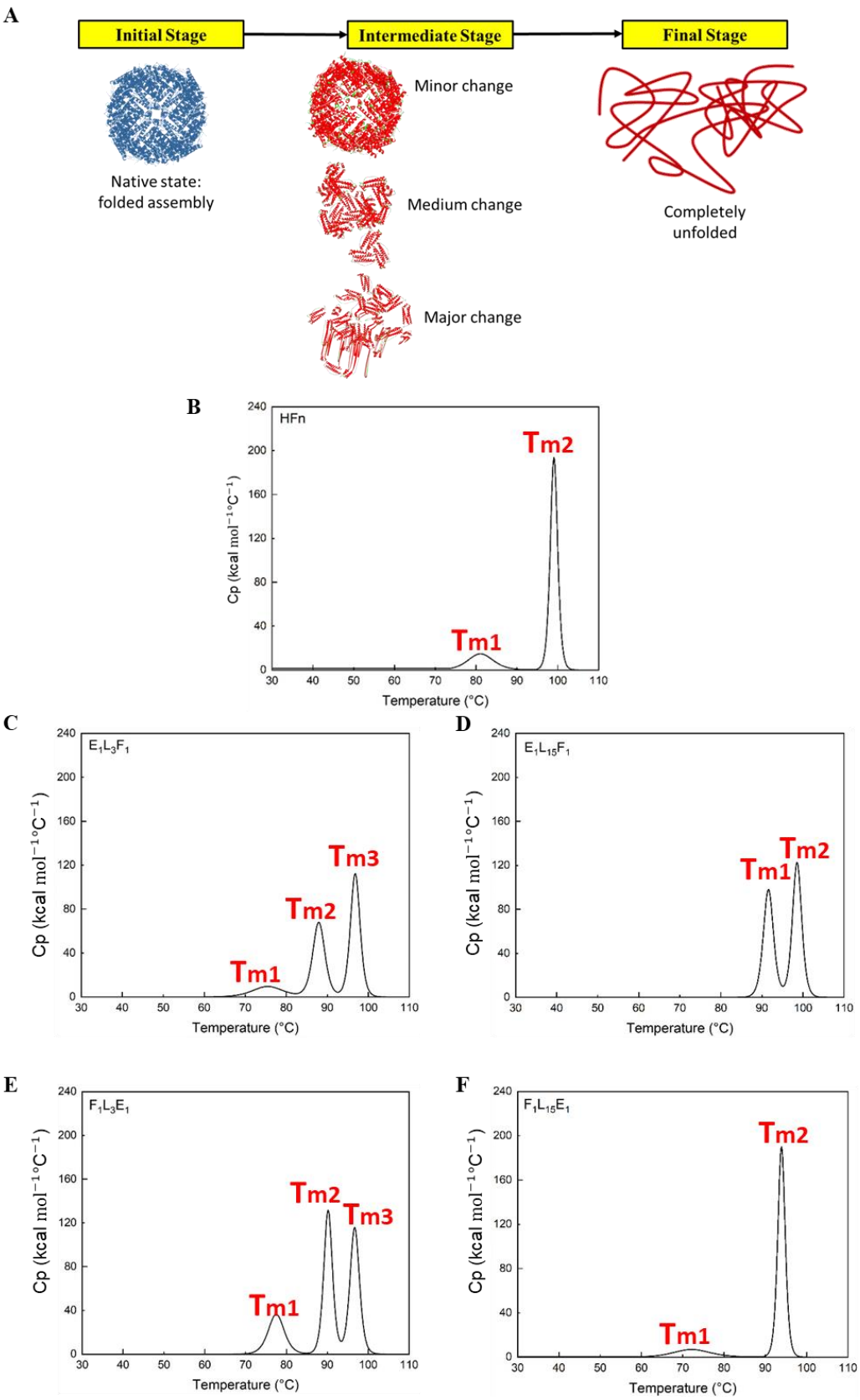


Figure 4.4 DSC (differential scanning calorimetry) protein thermal denaturation. A) Showing multiple-domain proteins have multiple-stages during DSC. Engineered ferritin (HFn) was used illustratively. Diagrams made with *Discovery Studio 2021* [21] and *Biorender*©. DSC thermogram for: B) HFn. C) $E_1L_3F_1$. D) $E_1L_{15}F_1$. E) $F_1L_3E_1$. F) $F_1L_{15}E_1$.

Table 4.2 Disassembled temperature (Tm) and enthalpy change (ΔH) for engineered ferritin generated by differential scanning calorimetry (DSC).

Protein		**Disassembled temperature T_m (°C)**				**Enthalpy change ΔH (kcal mol^{-1})**		
		T_{m1}	**T_{m2}**	**T_{m3}**	**T_m***	**ΔH_1**	**ΔH_2**	**ΔH_3**
Wild type	HFn	81.1	99.1	-	99.1	121	462	-
N-	$E_1L_3F_1$	75.5	87.9	96.8	87.9	96.2	265	350
terminus	$E_1L_{15}F_1$	91.6	98.6	-	98.6	322	367	-
C-	$F_1L_3E_1$	77.5	90.1	96.2	90.1	189	372	355
terminus	$F_1L_{15}E_1$	72.2	93.9	-	93.9	80.8	451	-

Thermal stability of engineered ferritin by fluorescence spectrometer

Fluorescence spectrometry and UV-VIS spectrometry permit a significantly detailed determination of the impact of linker length on protein tertiary and quaternary structure changes when applying thermal denaturation.

Fluorescence spectrometry is used widely to study the impact of denaturants on protein folding, and to understand protein stability [42,49-51]. The MFIC, Equation (4.2), was used to quantify the impact of thermal stress on protein tertiary structure [42, 50]. The greater the value, the greater the impact of heating on protein structure, and the less thermally stable is the protein. As is shown in Figure 4.5A, HFn, the control, exhibited the least fluorescence intensity change. This underscores therefore the most thermally stable protein structure. With insertion of linkers and epitopes, thermal stability of engineered ferritin decreased. All had significant tertiary structural change from a temperature of 75 °C. At a temperature of 80 °C, $E_1L_{15}F_1$ (N-terminus insertion) exhibited boosted thermal stability compared with $E_1L_3F_1$. At lower temperatures than 95 °C, $F_1L_{15}E_1$ exhibited boosted thermal stability. At 95 °C, $F_1L_{15}E_1$ and $F_1L_3E_1$ exhibited similar tertiary structural change.

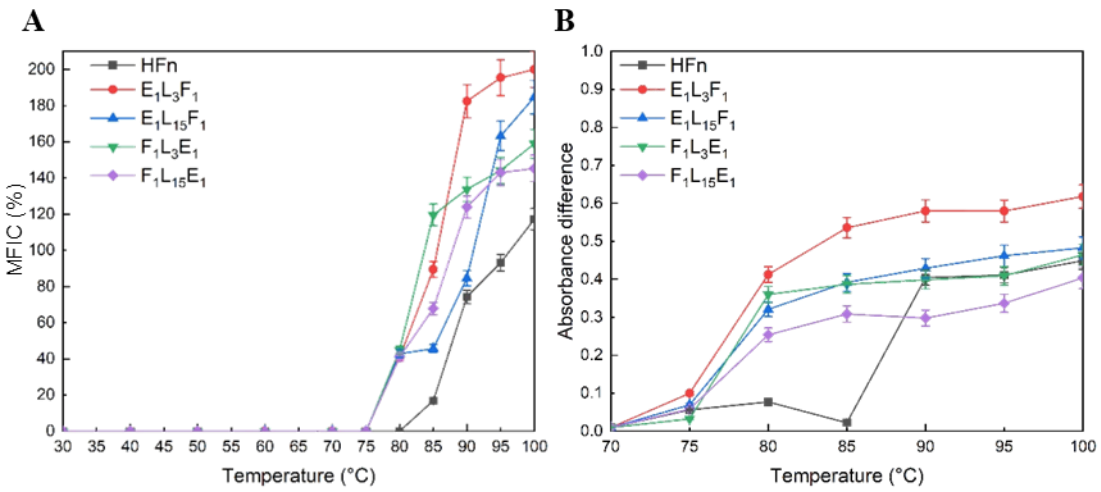


Figure 4.5 Impact of temperature on tertiary and quaternary structure of engineered ferritin. A) Maximum fluorescence intensity change (MFIC) for engineered ferritin against temperature denaturant. B) Absorbance difference at 207 nm between absorbance following heating and absorbance before heating.

Thermal stability of engineered ferritin by UV-Vis spectrometer

Although UV-VIS cannot aid a direct, detailed explanation of protein structural change, the change in absorbance at wavelength of 200 to 230 nm caused by the peptide bond of protein in solution, evidences protein quaternary structural change [52], [53]. As is seen in Figure 4.5B, HFn (control) was the most thermally stable. There was no significant quaternary structural change observed in HFn up to a temperature of 85 °C, whereas $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$ underwent quaternary change at about 80 °C. At about 90 °C, $F_1L_{15}E_1$ exhibited the least quaternary structural change, whilst $E_1L_3F_1$ exhibited the most. This finding underscores assumptions of the intermediate stage in DSC. At temperatures approaching 90 °C, $F_1L_{15}E_1$ maintained most of its quaternary structure, together with minor change in enthalpy. However, $E_1L_3F_1$, $E_1L_{15}F_1$, and $F_1L_3E_1$ all showed more significant structural change together with greater enthalpy change against temperature denaturant.

Overall, engineered ferritin with long flexible linkers has less absorbance change than that with short linkers. This finding confirms the advantage of long flexible linker in maintaining thermal

stability of proteins.

Investigation of thermal stability of engineered ferritin by fluorescence spectrometry and UV-VIS spectrometry was consistent with the DSC results. Wild-type HFn was the most thermally stable protein of all five. Long flexible linkers (15 residues) retain greater distance between epitope and HFn nanocage, compared with short linkers (3 residues) that reduce the interaction between exogenous peptides (epitopes and linkers) and wild-type protein nanocage. This is likely the reason that long linker group exhibits better thermal stability than the short linker does. Additionally, linkers inserted at the N-terminus alter protein surface hydrophobicity more significantly than at C-terminus. The most hydrophobic protein ($E_1L_3F_1$) was the least thermally stable.

It is concluded therefore that engineered ferritin with long flexible linkers has enhanced protein thermal stability, especially for N-terminus insertion.

4.3.4 Chemical denaturants stability using fluorescence spectrometer

Guanidine hydrochloride (GdnHCl) and urea are two (2) common chemical denaturants used to unfold proteins to investigate conformational stability [54].

To determine the impact of GdnHCl on denaturation of the nanoparticle vaccines, the four engineered ferritin, $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$ were treated with varying concentrations of GdnHCl of 0, 1, 2, 3, 4, 5 and 6 M for time intervals of 0, 1, 2, 4, 8, 12 and 24 h. HFn was the control. A greater MFIC was observed for proteins with weaker stability against chemical denaturants. It was found (Figure 4.6) that GdnHCl with concentration < 2M resulted in negligible tertiary structural change with all five proteins. All ferritin were observed

to begin losing tertiary structure from 2 M GdnHCl. In the initial period of the experiment (0 h) all ferritin had < 100 % change in tertiary structure at 4 M GdnHCl compared with no GdnHCl-treatment, HFn = 18, $E_1L_3F_1$ = 62, $E_1L_{15}F_1$ = 41, $F_1L_3E_1$ = 61 and $F_1L_{15}E_1$ = 21 %, Figure 4.6A. All proteins lost tertiary structure at 6 M GdnHCl, as evidenced by intensity change > 100 % (Figure 4.6A).

The long linker group ($E_1L_{15}F_1$ = 202.6, $F_1L_{15}E_1$ = 178 %) had less tertiary change than the short linker group ($E_1L_3F_1$ =227, $F_1L_3E_1$ = 226 %). At the end of the experiment (24 h), 4 proteins lost natural tertiary structure at 3 M GdnHCl, namely, $E_1L_3F_1$ = 311, $E_1L_{15}F_1$ = 293, $F_1L_3E_1$ = 156 and $F_1L_{15}E_1$ = 101 %) (Figure 4.6F), whereas HFn MFIC was 73 %.

Overall, for all time intervals, the control HFn exhibited the most stable tertiary structure against GdnHCl denaturant. This is evidenced by the lowest MFIC value. From 2 h, the trend-lines for long linker group ($E_1L_{15}F_1$ and $F_1L_{15}E_1$) remained below short linker group ($E_1L_3F_1$ and $F_1L_3E_1$), Figure 4.6. This finding confirms that long linker at both N-terminus and C-terminus insertions, conferred greater protein stability against GdnHCl than did a short linker.

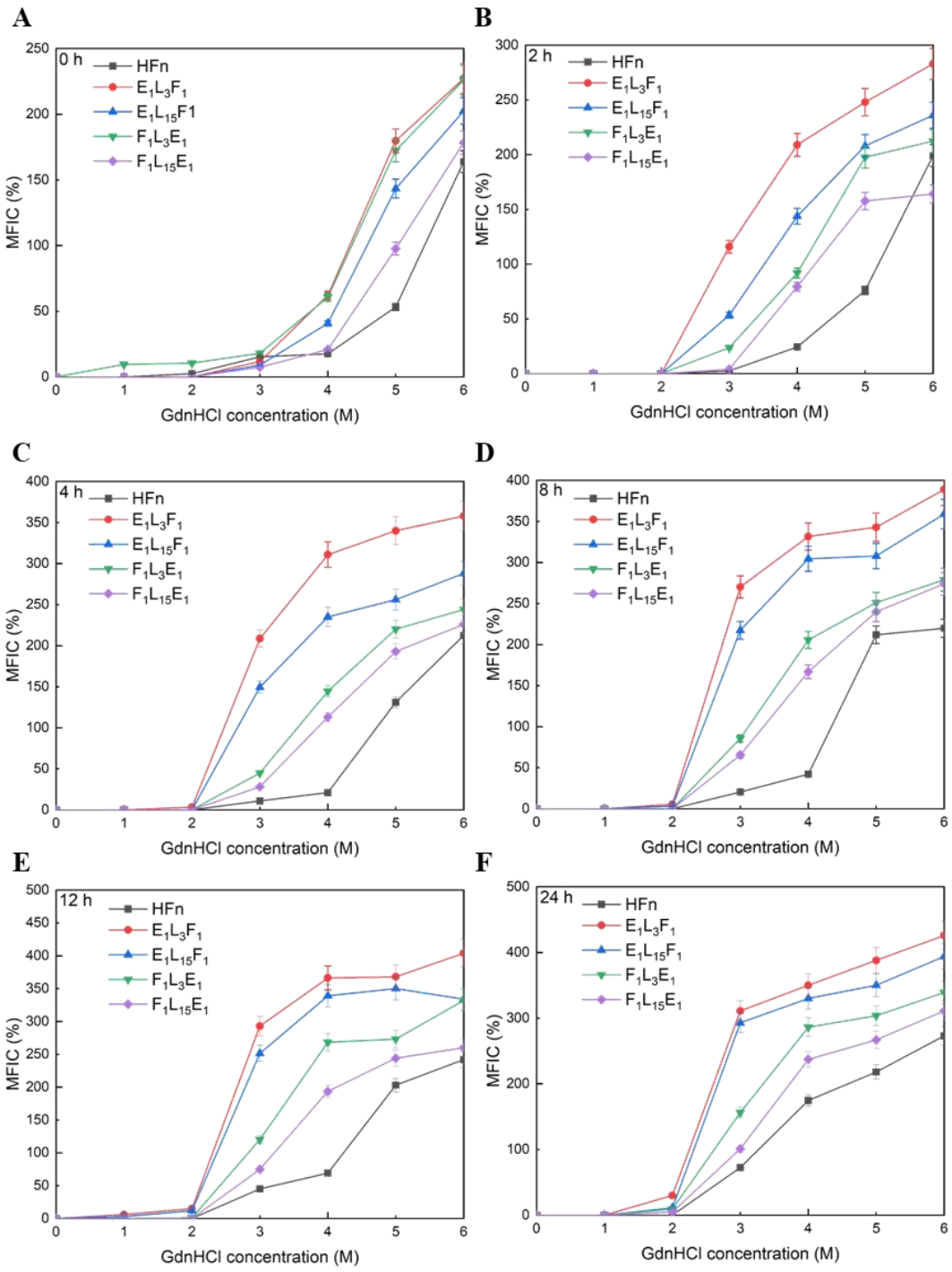


Figure 4.6 Maximum fluorescence intensity change (MFIC) caused by GdnHCl chemical denaturant on five (5) proteins, HFn, $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$. Time interval treated with GdnHCl: A) 0 h. B) 2 h. C) 4 h. D) 8 h. E) 12 h. F) 24 h.

GdnHCl and urea denaturation result in a different estimates of protein stability [55,56]. The mechanism for urea unfolding of proteins is different to that for GdnHCl. To test how linker length impacts protein stability against urea denaturants, experiments were conducted by treating proteins HFn, $E_1L_3F_1$, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$ with varying concentration of urea of 0, 1, 2, 3, 4, 5, 6, 7 and 8 M for time intervals of 0, 1, 2, 4, 8, 12 and 24 h. GdnHCl at the same concentration as urea denatured proteins more, as indicated by a higher MFIC. All five proteins showed < 50 % change in natural tertiary structure up to 4 M urea, Figure 4.7. Except for HFn, up to 4 h the trend-lines of MFIC for the other four engineered ferritins were highly similar at concentrations of urea (Figure 4.7A, 4.7B and 4.7C). For example, at 0 h the intensity change for both the long linker group ($E_1L_{15}F_1$ = 61 %, $F_1L_{15}E_1$ = 61 %) and the short linker group ($E_1L_3F_1$ = 62 %, $F_1L_3E_1$ = 64 %) were similar at 8 M urea (Figure 4.7A). Up to 8 h with urea treatment, all five proteins had tertiary structural change of < 100 % at 8 M (Figure 4.7A, 4.7B, 4.7C and 4.7D). $E_1L_3F_1$ lost tertiary structure at 8 M urea following treatment with urea for > 8 h. The fluorescence intensity change was > 100 %, whereas others, HFn, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$ retained partial, natural tertiary structures (Figure 4.7E and 4.7F). At the end of experiments (24 h), $E_1L_3F_1$ had 158 % of natural tertiary structure change with 8 M urea, whilst the change for HFn, $E_1L_{15}F_1$, $F_1L_3E_1$ and $F_1L_{15}E_1$ were, respectively, 81 %, 67 %, 97 % and 99 %, (Figure 4.7E).

In comparing long linker with short linker groups, it is seen that for N-terminus insertion ($E_1L_3F_1$ and $E_1L_{15}F_1$), a long linker is advantageous in boosting protein stability against urea denaturant. For C-terminus insertion ($F_1L_3E_1$ and $F_1L_{15}E_1$) there was significant difference between long and short linkers on protein stability against urea denaturant.

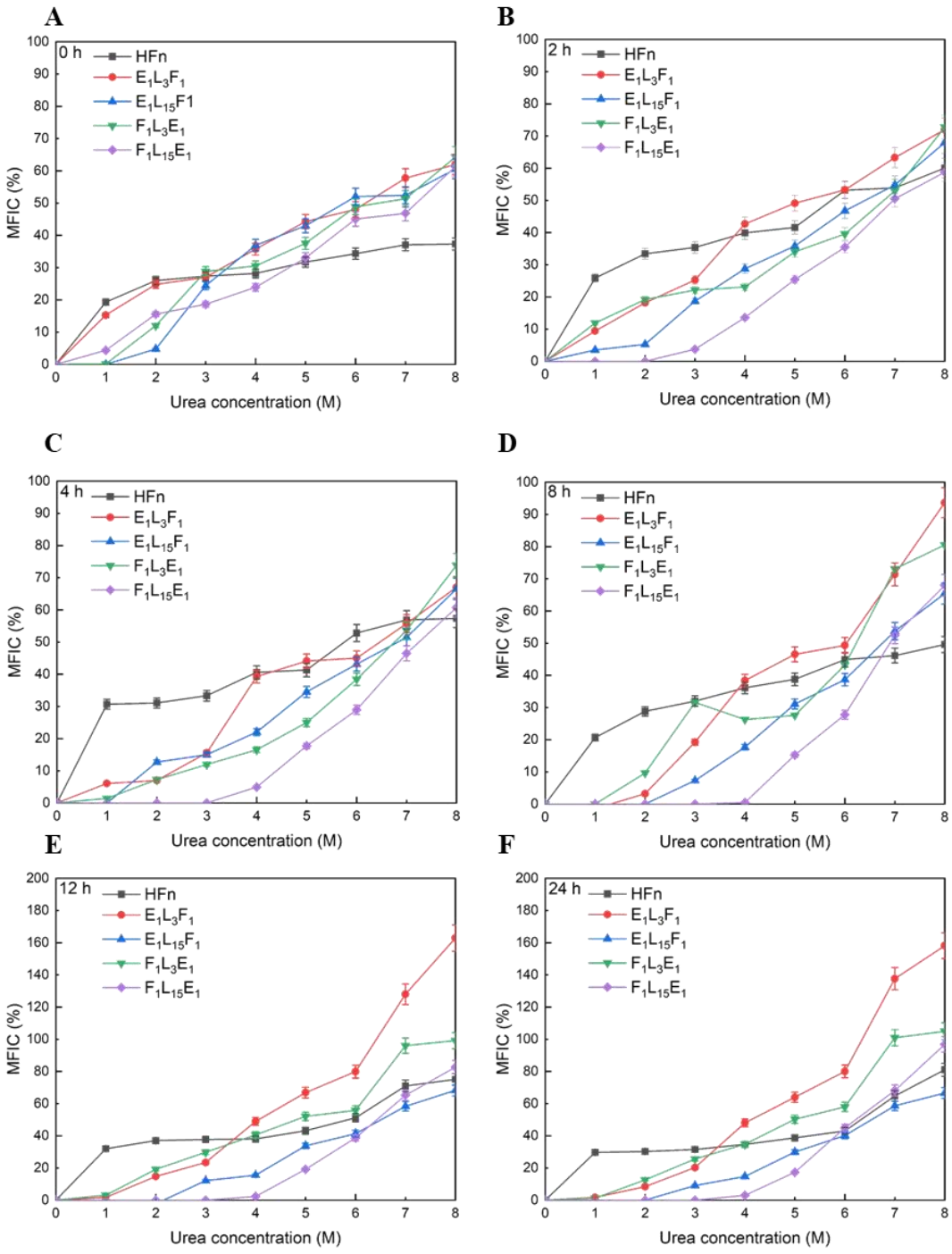


Figure 4.7 Maximum fluorescence intensity change (MFIC) caused by urea chemical denaturant on five (5) proteins, HFn, E1L3F1, E1L15F1, F1L3E1 and F1L15E1. Time interval treated with urea: A) 0 h. B) 2 h. C) 4 h. D) 8 h. E) 12 h. F) 24 h.

There is therefore a difference in impact on protein stability with the two chemical denaturants, GdnHCl and urea, between long and short flexible linkers. The long linker exhibited a boost in improving protein stability against GdnHCl for both N-terminus and C-terminus insertion of epitopes. With urea unfolding of proteins, a long linker inserted at N-terminus gave the protein nanocage greater stability against chemical denaturation. This finding however is not clear for C-terminus insertion.

This difference is explained by the mechanism of chemical denaturation of proteins. GdnHCl is more conclusive than urea in denaturing proteins [57]. The mechanism of chemical denaturants in unfolding protein is by facilitating solvation of hydrophobic regions of proteins [58]. In denaturing proteins urea after follows native → denatured state transition in a single step, whilst for GdnHCl intermediates are observed at low concentration of denaturant [54]. This is caused by the fact that guanidine cation (Gdn^+) at low concentration of GdnHCl (0 to 0.3 M) stabilizes proteins via binding to negative sites of proteins [54,59]. At greater concentration of GdnHCl, it acts as chemical denaturant. Rashid *et al*. reported urea has a chaotropic impact that disrupts the hydrogen bonding network between water molecules, and thereby decreases protein stability [54]. In contrast, GdnHCl has both chaotropic and ionic impact [54,56,60]. This fact explains that GdnHCl unfolds proteins through breaking both hydrophobic and electrostatic interactions between protein and solution.

It is concluded therefore that, for N-terminus insertion, long linker ($E_1L_{15}F_1$) and short linker ($E_1L_3F_1$) groups have a different hydrophobicity. $E_1L_3F_1$ is more hydrophobic and interacts more strongly with GdnHCl/urea. However, for C-terminus insertion, $F_1L_3E_1$ and $F_1L_{15}E_1$ interact with GdnHCl/urea similarly, because of similar protein surface hydrophobicity. Therefore, the stability of $F_1L_3E_1$ and $F_1L_{15}E_1$ against urea denaturant is similar. Dissimilarity

of stability against GdnHCl depends on ionic solute electrostatic interaction. Compared with short flexible linker (3 residues), long flexible linker (15 residues) maintains epitopes at a sufficient distance from the C-terminus of HFn.

The interaction between negative-charged amino acids (Asp and Glu) of EBNA1 epitope and Gdn^+ cations, affects meaningfully less on the HFn nanocage. $F_1L_{15}E_1$ is therefore less impacted by GdnHCl compared with $F_1L_3E_1$ (Figure 4.6).

4.4 Conclusions

A comprehensive experimental comparison of engineered ferritin with both long and short flexible linkers has shown that:

1. For N-terminus insertion, a long linker group, $E_1L_{15}F_1$, leads to a vaccine nanoparticle that is significantly less hydrophobic than when a short linker group, $E_1L_3F_1$, is used. This difference is likely because of steric shielding from long flexible linkers.
2. For C-terminus insertion, a long linker group has similar hydrophobicity to that for a short linker, as the linker and epitope are shielded within the nanoparticle cage.
3. $E_1L_{15}F_1$ exhibited greater thermal stability, together with GdnHCl-denaturant and urea-denaturant stability, than $E_1L_3F_1$.
4. $F_1L_{15}E_1$ exhibited greater thermal stability, together with GdnHCl-denaturant stability, than $F_1L_3E_1$.
5. Stability against urea denaturation for both $F_1L_{15}E_1$ and $F_1L_3E_1$ was similar.

It is concluded that flexible linker length molecular design can be used to significantly boost protein stability, and therefore assist the purification and production of safe and efficacious epitope-based chimeric vaccine. Flexible linker design therefore obviates production of unstable and unefficacious vaccines.

Findings will be of immediate interest to researchers in developing safe, stable and efficacious epitope-based chimeric vaccine.

4.5 References

1 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine: Nanotechnology, Biology and Medicine* **2**, 137-149, doi:https://doi.org/10.1016/j.nano.2006.07.009 (2006).

2 Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances* **41**, 107547, doi:https://doi.org/10.1016/j.biotechadv.2020.107547 (2020).

3 Bachmann, M. F. & Jennings, G. T. Vaccine delivery: A matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **10**, 787-796, doi:10.1038/nri2868 (2010).

4 Mohsen, M. O. *et al.* Major findings and recent advances in virus-like particle (VLP)-based vaccines. *Seminars in Immunology* **34**, 123-132, doi:10.1016/j.smim.2017.08.014 (2017).

5 Scheiblhofer, S. *et al.* Influence of protein fold stability on immunogenicity and its implications for vaccine design. *Expert Review of Vaccines* **16**, 479-489, doi:10.1080/14760584.2017.1306441 (2017).

6 McHugh, C. A. *et al.* Improved stability of a protein vaccine through elimination of a partially unfolded state. *Protein science : a publication of the Protein Society* **13**, 2736-2743, doi:10.1110/ps.04897904 (2004).

7 Structural Genomics, C. *et al.* Protein production and purification. *Nature Methods* **5**, 135-146, doi:10.1038/nmeth.f.202 (2008).

8 Morales-Camacho, J. I. *et al.* Expression, purification and thermal stability evaluation of an engineered amaranth protein expressed in *Escherichia coli*. *Electronic Journal of Biotechnology* **22**, 44-51, doi:https://doi.org/10.1016/j.ejbt.2016.04.001 (2016).

9 Chen, X. *et al.*Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

10 Zhao, H. L. *et al.* Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Expression and Purification* **61**, 73-77, doi:10.1016/j.pep.2008.04.013 (2008).

11 Amet, N. *et al.* Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharmaceutical research* **26**, 523-528, doi:10.1007/s11095-008-9767-0 (2009).

12 Bai, Y. *et al.* Recombinant granulocyte colony-stimulating factor-transferrin fusion

protein as an oral myelopoietic agent. *Proceeding of the National Academy of Sciences of the United States of America* **102**, 7292-7296, doi:10.1073/pnas.0500062102 (2005).

13 Bai, Y. & Shen, W. C. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharmaceutical Research* **23**, 2116-2121, doi:10.1007/s11095-006-9059-5 (2006).

14 Wei, Y. *et al.* Evaluation of lumazine synthase from Bacillus anthracis as a presentation platform for polyvalent antigen display. *Protein Science* **26**, 2059-2072, doi:10.1002/pro.3243 (2017).

15 van Rosmalen, M. *et al.* Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. *Biochemistry* **56**, 6565-6574, doi:10.1021/acs.biochem.7b00902 (2017).

16 Schuler, B. *et al.* Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proceeding of the National Academy of Sciences of the United States of America* **102**, 2754-2759, doi:10.1073/pnas.0408164102 (2005).

17 Arai, R. *et al.* Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering* **14**, 529-532, doi:10.1093/protein/14.8.529 (2001).

18 George, R. A. & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Engineering, Design and Selection* **15**, 871-879, doi:10.1093/protein/15.11.871 (2002).

19 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102, doi:10.1038/nature12202 https://www.nature.com/articles/nature12202#supplementary-information (2013).

20 Sliepen, K. *et al.* Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **12**, 82, doi:10.1186/s12977-015-0210-4 (2015).

21 Discovery Studio v. 2021 (Dassault Systèmes, San Diego, 2021).

22 Klein, J. S. *et al.* Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering Design and Selection* **27**, 325-330, doi:10.1093/protein/gzu043 (2014).

23 Reddy Chichili, V. P. *et al.* Linkers in the structural biology of protein-protein interactions. *Protein science : a publication of the Protein Society* **22**, 153-167, doi:10.1002/pro.2206 (2013).

24 Sabourin, M. *et al.* A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. *Yeast (Chichester, England)* **24**, 39-45,

doi:10.1002/yea.1431 (2007).

25 Sabourin, M. *et al.* A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. *Yeast* **24**, 39-45, doi:10.1002/yea.1431 (2007).

26 Reddy Chichili, V. P. *et al.* Linkers in the structural biology of protein–protein interactions. *Protein Science* **22**, 153-167, doi:https://doi.org/10.1002/pro.2206 (2013).

27 Robinson, C. R. & Sauer, R. T. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proceedings of the National Academy of Sciences* **95**, 5929, doi:10.1073/pnas.95.11.5929 (1998).

28 Nagi, A. D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Folding and Design* **2**, 67-75, doi:10.1016/s1359-0278(97)00007-2 (1997).

29 Chen, H. *et al.* Effect of linker length and flexibility on the clostridium thermocellum esterase displayed on bacillus subtilis spores. *Applied Biochemistry and Biotechnology* **182**, 168-180, doi:10.1007/s12010-016-2318-y (2017).

30 Lua, L. H. L. *et al.* Synthetic biology design to display an 18kDa rotavirus large antigen on a modular virus-like particle. *Vaccine* **33**, 5937-5944, doi:https://doi.org/10.1016/j.vaccine.2015.09.017 (2015).

31 Qu, Y. *et al.* Stability of engineered ferritin nanovaccines investigated by combined molecular simulation and experiments. *The Journal of Physical Chemistry B* **125**, 3830-3842, doi:10.1021/acs.jpcb.1c00276 (2021).

32 Qu, Y. *et al.* Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* **39**, 4830-4841, doi:https://doi.org/10.1016/j.vaccine.2021.07.021 (2021).

33 Durowoju, I. B. *et al.* Differential scanning calorimetry - a method for assessing the thermal stability and conformation of protein antigen. *Journal of Visualized Experiments*, 55262, doi:10.3791/55262 (2017).

34 OriginLab (Northampton, MA, USA, 2021).

35 Różycki, B. *et al.* The length but not the sequence of peptide linker modules exerts the primary influence on the conformations of protein domains in cellulosome multi-enzyme complexes. *Physical Chemistry Chemical Physics* **19**, 21414-21425, doi:10.1039/C7CP04114D (2017).

36 Mu, Q. *et al.* Molecular insight into the steric shielding effect of PEG on the conjugated staphylokinase: Biochemical characterization and molecular dynamics simulation. *PLOS ONE* **8**, e68559, doi:10.1371/journal.pone.0068559 (2013).

37 Labouta, H. I. *et al.* Surface-grafted polyethylene glycol conformation impacts the transport of PEG-functionalized liposomes through a tumour extracellular matrix model. *RSC Advances* **8**, 7697-7708, doi:10.1039/C7RA13438J (2018).

38 Malleshappa Gowder, S. *et al.* Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *The Scientific World Journal* **2014**, 7, doi:10.1155/2014/971258 (2014).

39 Lan, H. *et al.* The role of surface properties on protein aggregation behavior in aqueous solution of different pH values. *AAPS PharmSciTech* **21**, 122, doi:10.1208/s12249-020-01663-7 (2020).

40 Wagner, J. R. *et al*. Relation between solubility and surface hydrophobicity as an indicator of modifications during preparation processes of commercial and laboratory-prepared soy protein isolates. *Journal of Agricultural and Food Chemistry* **48**, 3159-3165, doi:10.1021/jf990823b (2000).

41 Strub, C. *et al.* Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry* **5**, 9, doi:10.1186/1471-2091-5-9 (2004).

42 Nemtseva, E. V. *et al.* Experimental approach to study the effect of mutations on the protein folding pathway. *PloS one* **14**, e0210361-e0210361, doi:10.1371/journal.pone.0210361 (2019).

43 Lee, B.-R. *et al.* Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. *Scientific reports* **6**, 35182-35182, doi:10.1038/srep35182 (2016).

44 Cao, X. M. *et al.* Effects of protein and phosphate buffer concentrations on thermal denaturation of lysozyme analyzed by isoconversional method. *Bioengineered* **7**, 235-240, doi:10.1080/21655979.2016.1197629 (2016).

45 Johnson, C. M. Differential scanning calorimetry as a tool for protein folding and stability. *Archives of Biochemistry and Biophysics* **531**, 100-109, doi:https://doi.org/10.1016/j.abb.2012.09.008 (2013).

46 Mazurenko, S. *et al.* Exploration of protein unfolding by modelling calorimetry data from reheating. *Scientific Reports* **7**, 16321, doi:10.1038/s41598-017-16360-y (2017).

47 Rastegari, A. A. *et al.* Thermal stability of pepsin: A predictive thermodynamic model of a multi-domain protein. *Biochemistry and Biophysics Reports* **9**, 295-301, doi:https://doi.org/10.1016/j.bbrep.2017.01.005 (2017).

48 Yu, J. *et al.* Thermostable iron oxide nanoparticle synthesis within recombinant ferritins from the hyperthermophile Pyrococcus yayanosii CH1. *RSC Advances* **9**, 39381-39393,

doi:10.1039/C9RA07397C (2019).

49 Monsellier, E. & Bedouelle, H. Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. *Protein Engineering, Design and Selection* **18**, 445-456, doi:10.1093/protein/gzi046 (2005).

50 Anand, S. *et al.* Differential thermal stability, conformational stability and unfolding behavior of Eis proteins from Mycobacterium smegmatis and Mycobacterium tuberculosis. *PLOS ONE* **14**, e0213933, doi:10.1371/journal.pone.0213933 (2019).

51 Pignataro, M. F. *et al.* Evaluation of peptide/protein self-assembly and aggregation by spectroscopic methods. *Molecules* **25**, doi:10.3390/molecules25204854 (2020).

52 Antosiewicz, J. M. & Shugar, D. UV–Vis spectroscopy of tyrosine side-groups in studies of protein structure. Part 2: selected applications. *Biophysical Reviews* **8**, 163-177, doi:10.1007/s12551-016-0197-7 (2016).

53 Saraiva, M. A. Interpretation of α-synuclein UV absorption spectra in the peptide bond and the aromatic regions. *Journal of Photochemistry and Photobiology B: Biology* **212**, 112022, doi:https://doi.org/10.1016/j.jphotobiol.2020.112022 (2020).

54 Rashid, F. *et al.* Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of human placental cystatin (HPC). *The Protein Journal* **24**, 283-292, doi:10.1007/s10930-005-6749-5 (2005).

55 Povarova, O. I. *et al.* Differences in the pathways of proteins unfolding induced by urea and guanidine hydrochloride: Molten globule state and aggregates. *PLOS ONE* **5**, e15035, doi:10.1371/journal.pone.0015035 (2010).

56 Monera, O. D. *et al.* Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein science : a publication of the Protein Society* **3**, 1984-1991, doi:10.1002/pro.5560031110 (1994).

57 Lim, W. K., *et al.* Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proceedings of the National Academy of Sciences* **106**, 2595, doi:10.1073/pnas.0812588106 (2009).

58 England, J. L. & Haran, G. Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. *Annual Review of Physical Chemistry* **62**, 257-277, doi:10.1146/annurev-physchem-032210-103531 (2011).

59 Mayr, L. M. & Schmid, F. X. Stabilization of a protein by guanidinium chloride. *Biochemistry* **32**, 7994-7998, doi:10.1021/bi00082a021 (1993).

60 Das, D. K. *et al.* Effect of an ionic liquid on the unfolding of human serum albumin: a

fluorescence correlation spectroscopy study. *Chemphyschem* **13**, 1949-1955, doi:10.1002/cphc.201100421 (2012).

4.6 Supporting information

4.6.1 Supporting figures

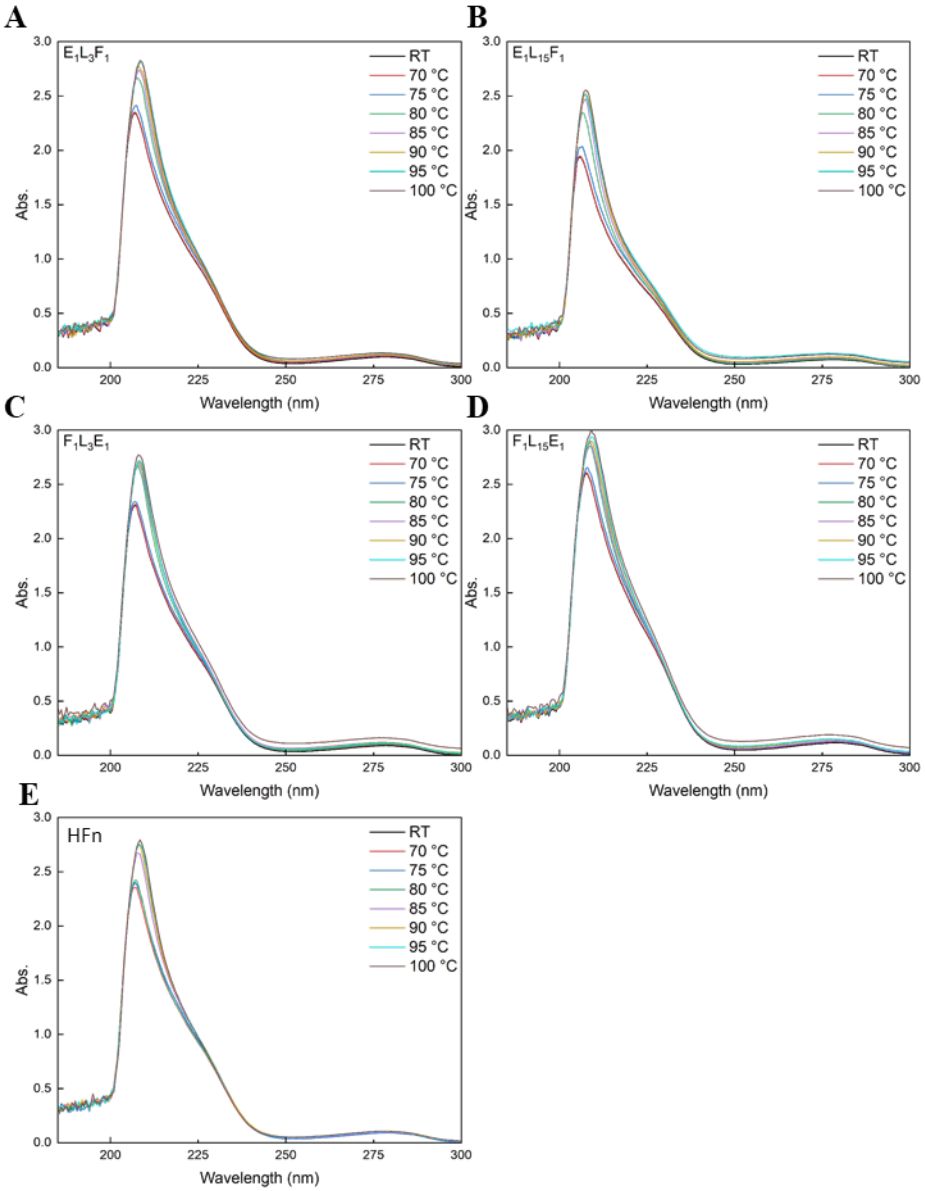


Figure S4.1 Quaternary structural change in engineered ferritin against thermal denaturant by UV-Vis. A) $E_1L_3F_1$. B) $E_1L_{15}F_1$. C) $F_1L_3E_1$. D) $F_1L_{15}E_1$. E) HFn.

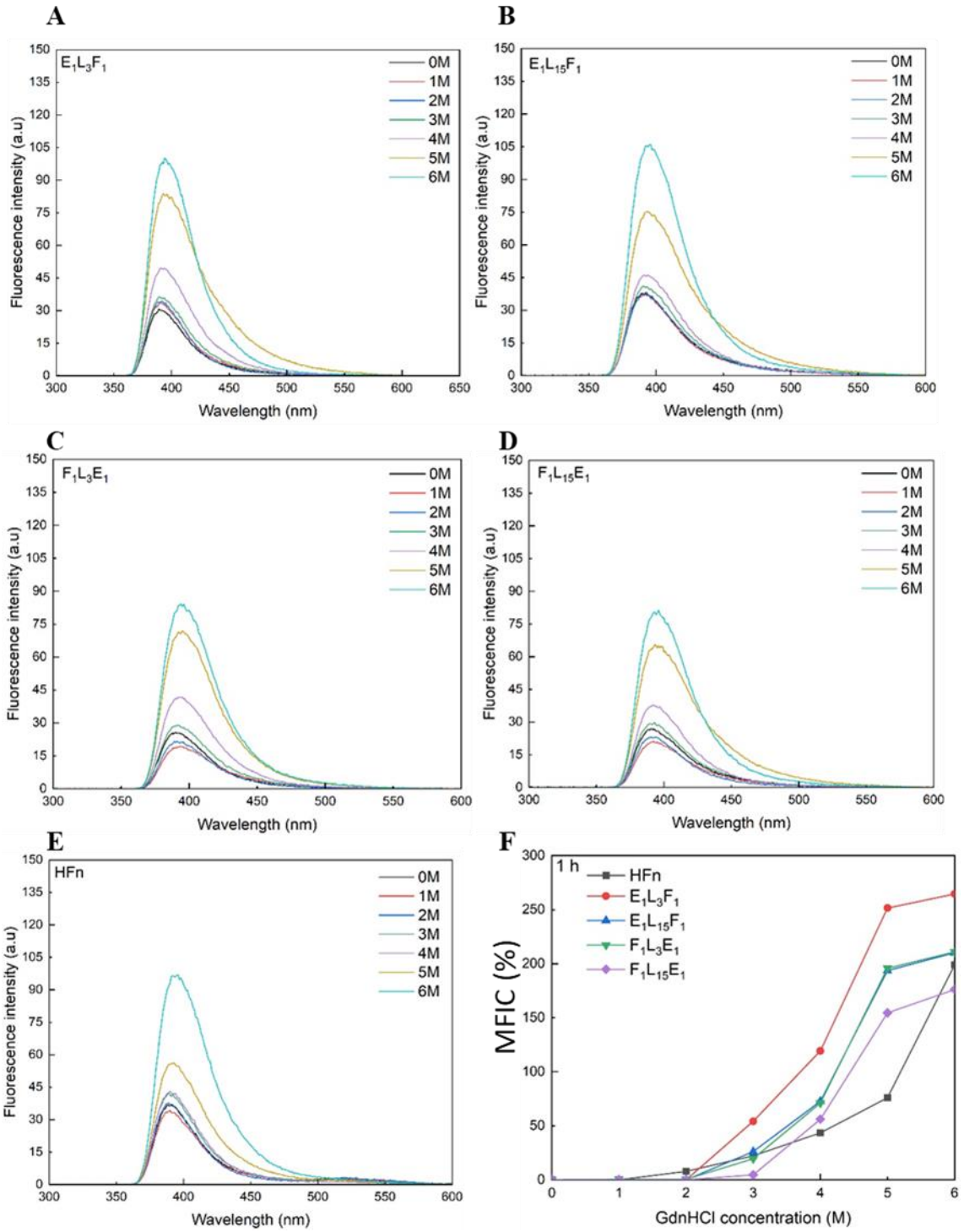


Figure S4.2 Tertiary structural change in engineered ferritin against GdnHCl denaturant. A) E1L3F1. B) E1L15F1. C) F1L3E1. D) F1L15E1. E) HFn. F) Maximum fluorescence intensity change (MFIC) at 1 h.

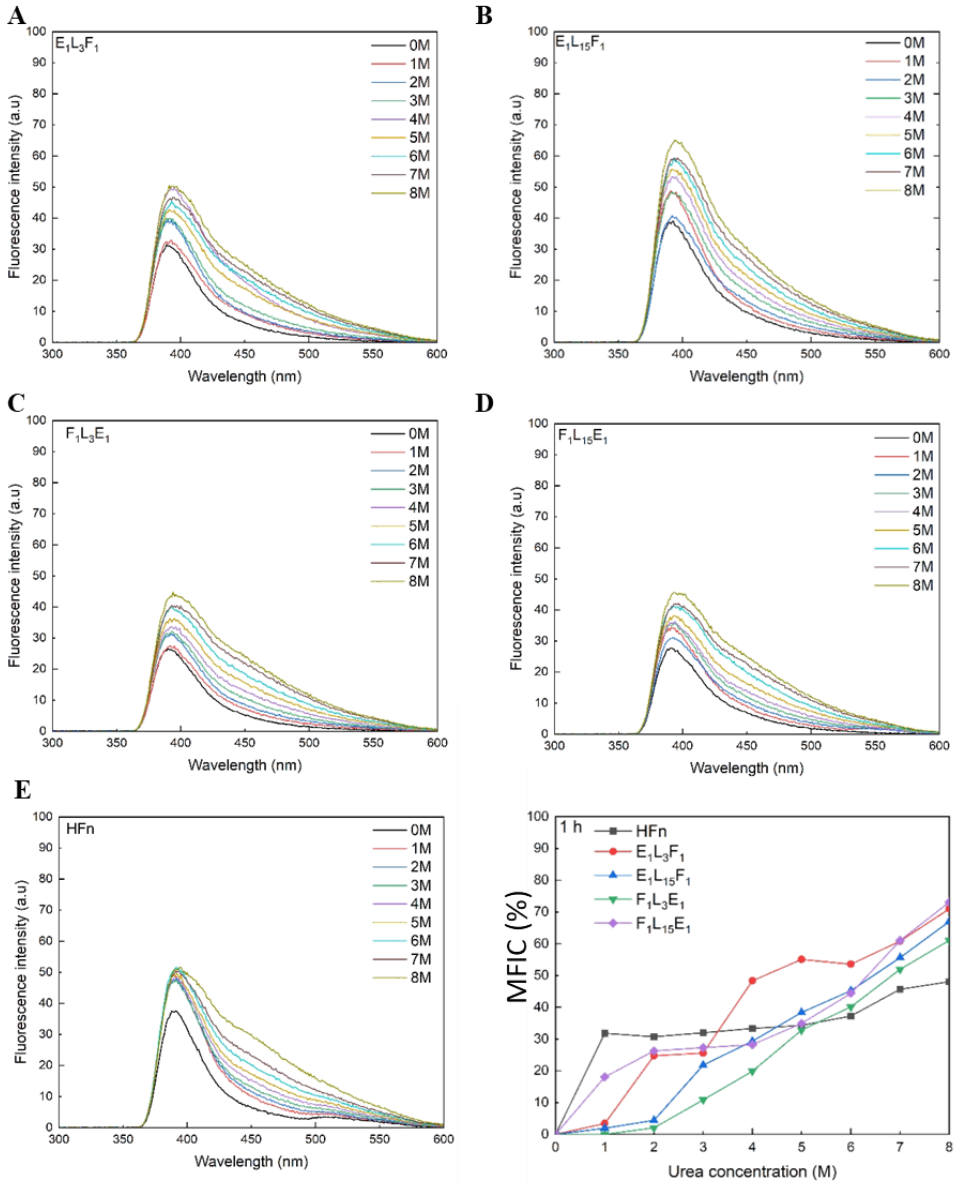


Figure S4.3 Tertiary structural change in engineered ferritin against urea denaturant. A) E1L3F1. B) E1L15F1. C) F1L3E1. D) F1L15E1. E) HFn. F) Maximum fluorescence intensity change (MFIC) at 1 h.

4.6.2 Supporting tables

Table S4.1 Gene sequence.

Name	Flexible linker group	Sequence
$E_1L_3F_1$		HPVGEADYFEYGGSTTASTSQVRQNYHQDSEAAINRQINLELYASYVYLS MSYYFDRDDVALKNFAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIK KPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETH YLNEQVKAIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES
$F_1L_3E_1$	Short	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKN FAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIKKPDCDDWESGLNAM ECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVKAIKELGDH VTNLRKMGAPESGLAEYLFDKHTLGDSDNESGGSHPVGEADYFEY
$E_1L_{15}F_1$	Long	HPVGEADYFEYGGSGGGGSGGGGSGGTTASTSQVRQNYHQDSEAAINRQI NLELYASYVYLSMSYYFDRDDVALKNFAKYFLHSHEEREHAEKLMKLQNQ RGGRIFLQDIKKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKN DPHLCDFIETHYLNEQVAIKELGDHVTNLRKMGAPESGLAEYLFDKHTLG DSDNES
$F_1L_{15}E_1$		TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKN FAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIKKPDCDDWESGLNAM ECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVKAIKELGDH VTNLRKMGAPESGLAEYLFDKHTLGDSDNESGGSGGGGSGGGGSGGHPVG EADYFEY
HFn	Control	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKN FAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIKKPDCDDWESGLNAM ECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVKAIKELGDH VTNLRKMGAPESGLAEYLFDKHTLGDSDNES

Table S4.2 Purification methods.

Protein	Step 1: Heating precipitation	Step 2: HIC
HFn	5M NaCl (1:5) 1M NaAc-HAc (pH4.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF
$E_1L_3F_1$	5M NaCl (1:5) 1M NaAc-HAc (pH5.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 0.8M AS, 100 mM PB, pH6.5 Load samples to Butyl FF
$E_1L_{15}F_1$	5M NaCl (1:5) 1M NaAc-HAc (pH5.0) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.0M AS, 100 mM PB, pH6.5 Load samples to Butyl FF
$F_1L_3E_1$	5M NaCl (1:5) 1M NaAc-HAc (pH4.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF
$F_1L_{15}E_1$	5M NaCl (1:5) 1M NaAc-HAc (pH4.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF

Table S4.3 Chemical denaturants testing conditions.

Time intervals (h)	GdnHCl concentration (M)	Urea concentration (M)
0	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
1	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
2	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
4	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
8	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
12	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8
24	0	0
	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
		7
		8

4.6.3 Other supporting information not included in online publication

Figure S4.4 shows that $E_1L_3F_1$ (lane 4) and $F_1L_3E_1$ (lane 8) both achieved soluble expression by using expression condition 37 °C, 200 rpm, 4 h. There were minor inclusion bodies expression.

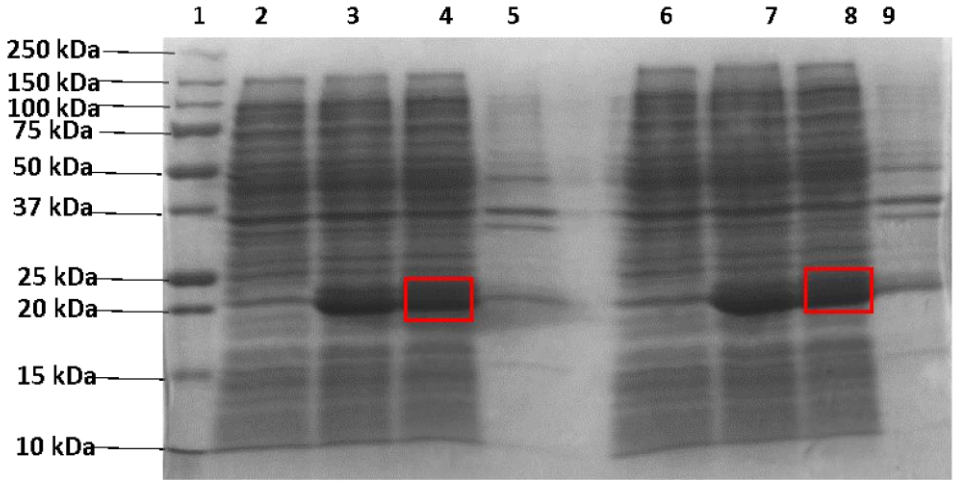


Figure S4.4 SDS PAGE image of engineered ferritins ($E_1L_3F_1$ and $F_1L_3E_1$) expression. Lane 1: Protein marker; lane 2: $E_1L_3F_1$ before adding IPTG; lane 3: $E_1L_3F_1$ after adding IPTG; lane 4: $E_1L_3F_1$ soluble expression; lane 5: $E_1L_3F_1$ inclusion bodies expression; lane 6: $F_1L_3E_1$ before adding IPTG; lane 7: $F_1L_3E_1$ after adding IPTG; lane 8: $F_1L_3E_1$ soluble expression; lane 9: $F_1L_3E_1$ inclusion bodies expression. Red box indicated target proteins E1F1/F1E1.

Figure S4.5 indicated the optimal condition for $E_1L_3F_1$ in the first step of purification was pH 5.5, 60 °C, 10 min, and total protein concentration 4 mg mL^{-1}. At pH 5.0, all $E_1L_3F_1$ proteins were precipitated. Figure S4.6 indicated the optimal condition for $F_1L_3E_1$ in the first step of purification was pH 4.5, 60 °C, 10 min, and total protein concentration 4 mg mL^{-1}.

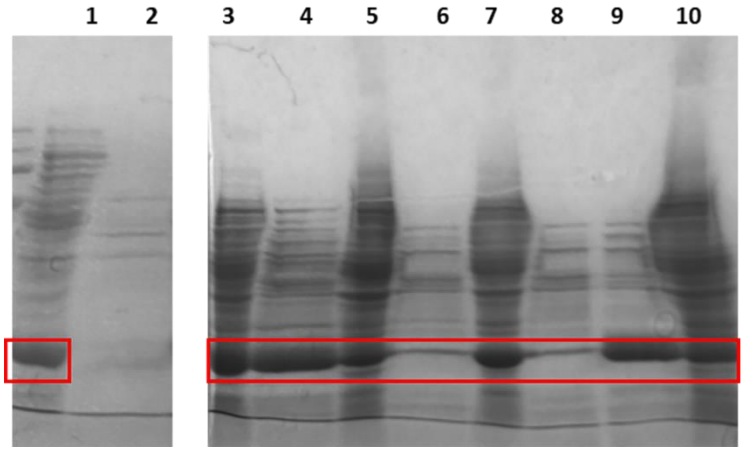


Figure S4.5 SDS PAGE image of $E_1L_3F_1$ acid and heat precipitation at various conditions. Lane 1: $E_1L_3F_1$ before purification (Gel 1); lane 2: supernatant, pH 5.0, 60 °C, 10 min; lane 3: $E_1L_3F_1$ before purification (Gel 2); lane 4: supernatant, pH 5.5, 50 °C, 10 min; lane 5: precipitation, pH 5.5, 50 °C, 10 min; lane 6: supernatant, pH 5.0, 50 °C, 5 min; lane 7: precipitation, pH 5.0, 50 °C, 5 min; lane 8: supernatant, pH 5.0, 50 °C, 10 min; lane 9: supernatant, pH 5.5, 60 °C, 10 min; lane 10: precipitation, pH 5.5, 60 °C, 10 min. Red box indicated target protein $E_1L_3F_1$.

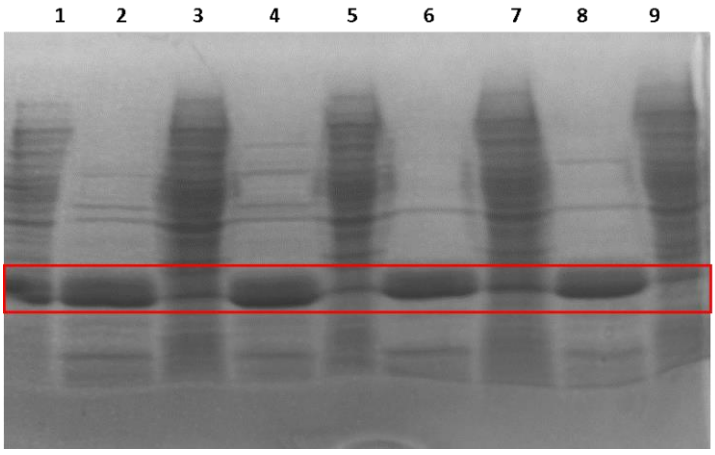


Figure S4.6 SDS PAGE image of $F_1L_3E_1$ acid and heat precipitation at various conditions. Lane 1: $F_1L_3E_1$ before purification; lane 2: supernatant, pH 4.5, 50 °C, 10 min; lane 3: precipitation, pH 4.5, 50 °C, 10 min; lane 4: supernatant, pH 5.0, 50 °C, 10 min; lane 5: precipitation, pH 5.0, 50 °C, 10 min; lane 6: supernatant, pH 4.5, 60 °C, 10 min; lane 7: precipitation, pH 4.5, 60 °C, 10 min; lane 8: supernatant, pH 5.0, 60 °C, 10 min; lane 9: precipitation, pH 5.0, 60 °C, 10 min. Red box indicated target protein $F_1L_3E_1$.

CHAPTER 5 ENGINEERED DESIGN OF HELIX E STRUCTURE ON FERRITIN NANOPARTICLES

Statement of Authorship

Title of Paper	Engineered design of helix E structure on ferritin nanoparticles
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Qu, Y.; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Engineered design of helix E structure on ferritin nanoparticles. *ACS Applied Bio Materials.*

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Designed experiment, performed experiments; analysed data; writing manuscript		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	11/10/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth Davey		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	11/10/2021

Name of Co-Author	Yan Sun		
Contribution to the Paper	Experiment design, Proof read the manuscript		
Signature		Date	11/10/2021

Name of Co-Author	Anton Middelberg		
Contribution to the Paper	Experiment design, ~~proof read~~ edited the manuscript		
Signature		Date	13/10/21

Name of Co-Author	Jingxiu Bi		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/2021

Engineered design of helix E structure on ferritin nanoparticles

Yiran Qu, Kenneth Davey, Yan Sun, Anton Middelberg, Jingxiu Bi*

Y. Qu, A/Prof. Dr. K. Davey A/Prof. Dr. J. Bi

School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide
South Australia 5005 (Australia)

Prof. Dr. A. Middelberg

Division of Research and Innovation, The University of Adelaide, SA 5005 (Australia)

Prof. Dr. Y. Sun
Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072 (China)

Submitted to ACS Applied Bio Materials

Abstract

Insertion of an immunogenic epitope at the C-terminus of ferritin has shown potential to produce stable and efficacious vaccine. There is however limited understanding of how C-terminus insertion affects ferritin protein stability. Helix E at the C-terminus attracts interest because there are contradictory reports as to whether it has a role in protein stabilization. Here we report, for the first time, combining of molecular dynamics simulation (MDS) with experiment to engineered design helix E at the C-terminus of human ferritin heavy chain (F_1) engineered inserted with Epstein-Barr nuclear antigen 1 (EBNA1, E_1) and flexible linker (L_3) residues (to give $F_1L_3E_1$). Hot spots on Helix E of the C-terminus were predicted by MDS at aa167 (Glu) and aa171 (Asp). Five (5) variants of $F_1L_3E_1$ were constructed by considering hot spots and alteration of electrostatic or hydrophobic interfaces, namely: 1) C1, hot spots substituted with non-charged residue Gln; 2) C2, hot spots substituted with positive-charged residue Arg; 3) C3, hydrophobic residues substituted with the most hydrophobic residues Val and Ile; 4) C4, hydrophobic residues substituted with the most hydrophilic residues Gln and Asn; and 5) C5, heptad repeat structure in Helix E disrupted by substituting 'a' and 'd' heptad residues with non-charged polar residue Gln. It was found that Helix E is essential to maintain integrated protein stability and that changing the hydrophobic interface (C3 and C4) had more significant effects on protein folding and stability than changing the electrostatic interface (C1 and C2). It was confirmed by both MDS and experiment that variants C1, C2 and C5 were able to fold to form stable conformational structures with similar protein surface hydrophobicity to that for $F_1L_3E_1$. However, they are less thermally stable than $F_1L_3E_1$. Significant changes in hydrophobicity drove significant protein aggregation for variants C3 and C4. It is concluded that molecular design of the C-terminus in engineered ferritin, especially Helix E, is important to ensure epitope-based chimeric vaccine can be safe (aggregate free) and efficacious.

KEYWORDS: Human ferritin heavy chain, molecular dynamics simulation, hot spot

5.1 Introduction

There is a great potential to develop ferritin-based vaccine because of numerous benefits: convenient display of antigens, high thermal stability, good biodegradability and safety. Our research has shown that engineered ferritin carrying epitopes at the C-terminus demonstrates greater thermal-, pH- and chemical denaturants stability than engineered ferritin carrying epitopes at the N-terminus [1]. Immunological studies have also shown that engineered ferritin with C-terminus insertion of epitopes induces stronger cell-mediated immune response than for the equivalent N-terminus insertion [2]. These studies suggest further understanding of the C-terminus design space will aid the design of stable vaccine candidates.

Research has demonstrated that substitutions of amino acid residues can affect both protein stability and protein folding pathway [3-8]. Current studies regarding engineered ferritin focus on modifications at the C-terminus of ferritin, because the C-terminal region contributes significantly to protein stability and assembly [9]. The N-terminal region, BC loop and threefold axis contribute minimally to ferritin folding and assembly, while the DE loop and C-terminal region play more important roles. Ingrassia *et al.* found that the last 6 non-helical residues (177-182: DSDNES) (Figure 5.1) have no obvious effects on ferritin stability, and the extension of this area slightly reduced the solubility and capacity to assemble ferritin cages [9]. This research particularly addressed the effect of E Helix on helical structure of the molecule along with protein folding. Similar results were found by Fan *et al.*[10]. They indicated that the BC Helix in bacterioferritin from *Escherichia coli* (BFR, a protein cage similar to ferritin) was less important than the E Helix for self-assembly [10]. Removal of the E Helix resulted in a destabilized protein [10 11]. On the other hand, Luzzago *et al.* claimed that E helices were not

important for human ferritin H-chain assembly [12]. The contradictory results from these studies highlights the importance of further understanding the role of E Helix on ferritin stability.

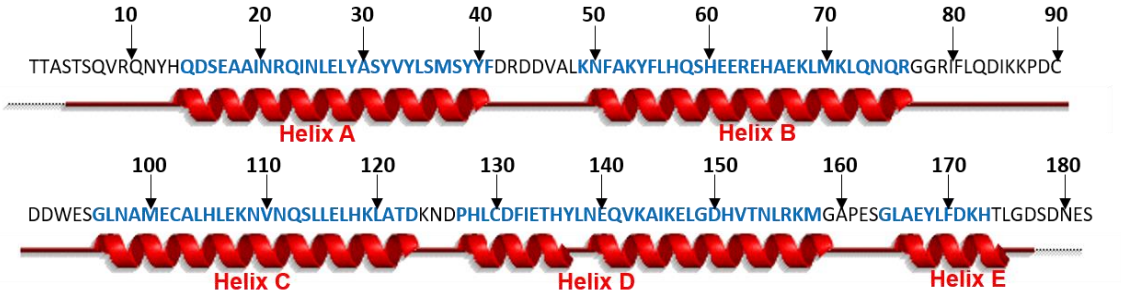


Figure 5.1 Schematic representation of ferritin human heavy chain secondary structure by EMBL-EBI. Residues with alpha helices are blue-bold shown in sequence.

Alpha helix bundles often comprise a heptad repeat having a characteristic pattern of hydrophilic and hydrophobic amino acids every seven residues [13,14]. The residues within the heptad are designated ‘a b c d e f g’ (Figure 5.2A). Positions ‘a’ and ‘d’ are generally hydrophobic. Here, we treated the ‘165’ and ‘168’ positions of ferritin as ‘a’ and ‘d’ (Figure 5.2B). Five (5) variants were designed aiming to investigate the effect of electrostatic interface and hydrophobic interface on protein stability. The information about these 5 variants are outlined in Table 5.1. C1 and C2 are variants that alter the electrostatic interface by replacing negatively charged residues Glu (E) and Asp (D) with uncharged residue Gln (Q), and positively charged residue Arg (R), respectively. C3, C4 and C5 were designed to alter the hydrophobic interface. C3 was designed to replace all non-charged residues with very hydrophobic residues Val (V) and Ile (I), while C4 was formed to replace with the most hydrophilic residues Q and Asn (N). C5 was designed to break the heptad repeat by changing the ‘a’ and ‘d’ positions with non-charged hydrophilic reside Q.

To our knowledge, no research has been reported that investigated the effects of altering the

electrostatic or hydrophobic interfaces on the stability of engineered ferritins. The aim of this paper is to identify the key residues at the C-terminus of engineered ferritin affecting protein stability, and further understand how altering electrostatic or hydrophobic interfaces affects C-terminus folding and assembling. In this paper, molecular dynamics simulation (MDS) was firstly applied to predict hot spots in the C-terminus of engineered ferritin carrying Epstein-Barr nuclear antigen 1 (EBNA1) with 3 residues of flexible linker ($F_1L_3E_1$). Five (5) variants were further constructed by altering the electrostatic or hydrophobic interfaces of $F_1L_3E_1$ according to the identified hot spots. Hydrophobic and thermal stability study for assembled variants were further predicted by MDS following the validation by experiments. It was found that altering the electrostatic or hydrophobic interfaces on the E Helix decreased protein stability. C3 (L165V-A166I-Y168I-L169V-F170I) and C4 (L165Q-A166N-Y168Q-L169N-F170Q), which were variants changing native hydrophobicity of the C-terminus, were found to be the most unstable regarding protein folding and conformational assembly. Variants C1 (E167Q-D171Q), C2 (E167R-D171R) and C5 (C5-L165Q-Y168Q) exhibited similar conformational stability and protein surface hydrophobicity compared with $F_1L_3E_1$, however, significantly less thermal stability was observed.

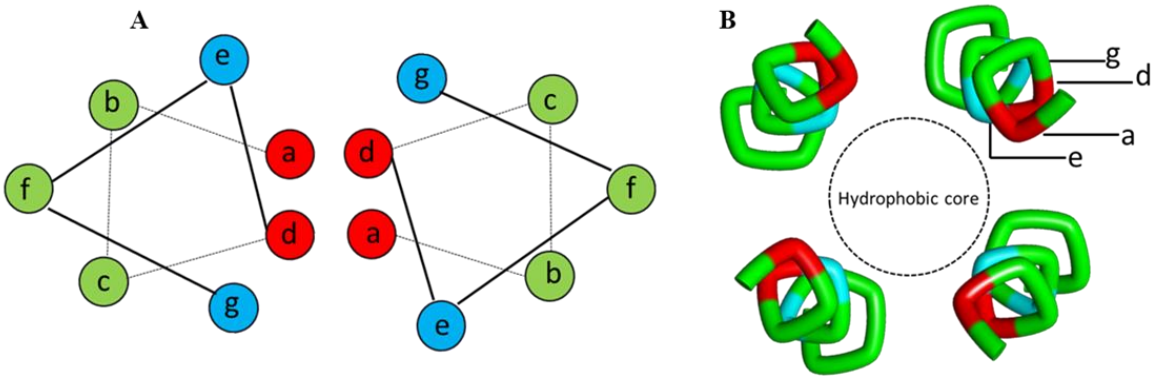


Figure 5.2 Schematic of heptad repeat. A) Mimic diagram of heptad repeat showing position 'a b c d e f g ' in a hydrophobic core. Position 'a' and 'd' are hydrophobic amino acids (red). Position 'e' and 'g' are charged amino acids (blue). B) Hydrophobic core taken from human ferritin heavy chain showing positions 'a, 'd' 'e' and 'g'. Hydrophobic residues are marked in red; Charged residues are marked in blue; Other residues are marked in green.

Table 5.1 Variant information and sequence. Red denotes residue changes.

Variant	Sequence from position "164" to "173"
$F_1L_3E_1$ (Control)	GLAEYLFDKH
C1 (No charge)	GL AQYLF QKH
C2 (Same charge)	GL ARYLF RKH
C3 (Most hydrophobic)	GV IEIVI DKH
C4 (Most hydrophilic)	GQ NEQNQ DKH
C5 (Break L Y interface)	GQ AEQLF DKH

5.2 Materials and methods

5.2.1 Protein data bank (PDB) files build-up

The PDB files of the Human heavy chain ferritin (PDB ID: 2FHA) and the Epstein-Barr virus nuclear antigen-1 (EBNA1) (PDB ID: 2FZ3) were generated from the RCSB Protein Databank. Discovery Studio (DS) 2021 was used to construct recombinant protein structures. The model epitope was truncated from EBNA1 comprising aa 407 to aa 417. By inserting EBNA1 epitope at the C-terminus of each ferritin monomer with three residues linker (GGS), PDB file named $F_1L_3E_1$ Assembly was prepared (Table S5.1, Supporting Information). By applying 'Mutate' function in DS 2021, PDB files for five (5) assembled variants (C1, C2, C3, C4 and C5) were built.

To study the interaction between each C-terminus, two C-termini from the hydrophobic core of assembled $F_1L_3E_1$ (aa 164 to aa 173) were cut (Figure 5.3). This procedure was repeated for five variants.

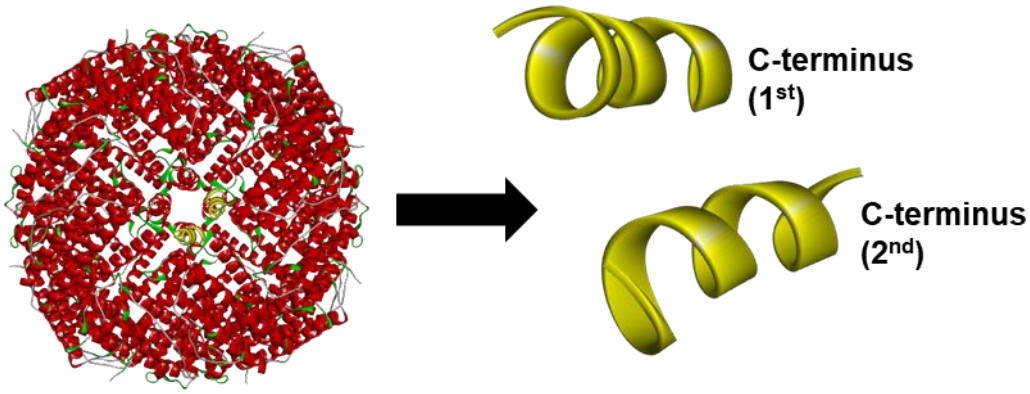


Figure 5.3 Mimic diagram to illustrate PDB files build-up for study of C-terminus to C-terminus interaction by DS 2021.

5.2.2 Molecular dynamics simulation (MDS)

GROMACS 2018.3 was used to perform MDS of above built-up protein structures in a water environment [15]. The Amber-99sb force field was used to generate topology files [16,17]. Each protein structure was solvated in a cubic box using the SPC/E water model. The settings were similar to previous published work [1]. The dimensions of the cubic box for each structure are summarized in Table S5.2. The minimum distance between the protein and the wall of the unit cell was set to be 1.0 nm. The protein/ion/water system was firstly energy minimized and then equilibrated in a 100-picosecond MDS with positional restraints on the heavy atoms of the protein. The equilibration was performed at a temperature of 300 K and a pressure of 1 bar by coupling to an external heat source and an isotropic pressure bath. The position restraints were then released and the MDS for assembled structures was produced in 200 nanoseconds and the snapshots were stored per 100 picoseconds for the MD trajectory analysis. For C-terminus to C-terminus interaction study, a 600 ns production MDS was performed.

For hydrophobicity study by simulation, hydrophobic solvent accessible surface area (SASA) was determined by using the *gmx sasa* tool from GROMACS based on the completed MDS results for $F_1L_3E_1$, C1, C2, C3, C4 and C5 in water at 27 °C (300 K).

Thermal stability studies for assembled structures of $F_1L_3E_1$, C1, C2, and C5. MDS were produced in 200 nanoseconds at 423 K. The snapshots were stored per 100 picoseconds for the MD trajectory analysis. Root mean square deviation (RMSD), radius of gyration (R_g) and root mean square fluctuation (RMSF) were determined using *gmx rmsd* tool, *gmx gyrate* tool and *gmx rmsf* tool from GROMACS. Helix content was calculated using the *gmx dssp* tool from GROMACS.

5.2.3 Molecular mechanics/Poisson-Boltzmann solvent accessible surface area (MM-PBSA)

The g_mmpbsa software [18] was executed to calculate the binding free energy and residue decomposition of C-terminus using the MM-PBSA method [19-21]. The MM-PBSA method was applied as follows

$$\Delta G_{bind} = \Delta G_{polar} + \Delta G_{nonpolar} \quad (5.1)$$

$$\Delta G_{polar} = \Delta G_{elec} + \Delta G_{PB} \quad (5.2)$$

$$\Delta G_{nonpolar} = \Delta G_{vdW} + \Delta G_{np} \quad (5.3)$$

where ΔG_{bind} is the binding energy; ΔG_{polar} is the polar energy; $\Delta G_{nonpolar}$ is the nonpolar energy; ΔG_{elec} is the electrostatic energy; ΔG_{PB} is the electrostatic solvation energy; ΔG_{vdW} is the van der Waals energy ΔG_{np} is the nonpolar solvation energy.

The free-energy contribution of each residue can be classified as polar and nonpolar interactions. Polar energy is the contribution from electrostatic interactions, while nonpolar energy is the contribution from hydrophobic interactions.

5.2.3 Protein expression

The pET 30a vector was used to construct plasmids harbouring genes of HFn-EBNA1 (F1L3E1) (control) and its variants (C1, C2, C3, C4 and C5) inserted between the *Nde I* and *BamH I* restriction sites. The protein sequence is shown in Table S5.3 (Supporting Information). Plasmids were transformed into *E. coli* BL21 (DE3) (Invitrogen, USA). Epitope EBNA1 (HPVGEADYFEY) was inserted at the C-terminus of HFn with three residues of linkers (GGS). Single colonies engineered to express each protein ($F_1L_3E_1$, C1, C2, C3, C4 and C5) were selected from an agar plate and incubated in 50 mL LB-ampicillin (100 μg mL^{-1}) medium. The culture was grown for 16 h at 37 °C in a shaking incubator at 180 rpm. In every 500 mL LB-

ampicillin (100 µg mL^{-1}) medium, 10 mL of cultured medium from 50 mL was added. The culture was grown at 37 °C in a shaking incubator at 200 rpm until OD_{600} reached 0.8, and then protein expression was induced using 1 µM IPTG (Sigma Aldrich, USA) at 37 °C for 4 h at 200 rpm in a shaking incubator. Cells were collected and resuspended after IPTG induction for four (4) h. After sonication on ice (Scientx sonicator at 360 W: 4 sec on and 6 sec off for 10 min) with lysis buffer (20 mM PB, 2 mM EDTA, pH 7.0), bacterial lysate was centrifuged at 13,751 xg for 30 min to remove cell debris. Reducing SDS-PAGE (Section 5.2.5) was used to confirm protein expression.

5.2.4 Protein purification

The procedure to purify proteins ($F_1L_3E_1$, C1, C2, C3, C4 and C5) included two (2) steps ($F_1L_3E_1$, C1 and C5) or four (4) steps (C2, C3 and C4). Details are provided in Table S5.4, Supporting Information. The first step was pre-purification of proteins by acid, salt and heat precipitation. Supernatant of cell lysate containing soluble expressed protein was diluted with Milli-Q water to adjust total protein concentration to 4 mg mL^{-1}. The pH of diluted supernatant was adjusted with NaCl (5M) and NaAc-HAc (1M) as per Table S5.4. The mixture was then heated to 60 °C and maintained for 10 min to precipitate protein impurities. Precipitated protein impurities were removed by centrifugation (19,802 xg at 4°C for 30 min).

The second step, except for C4, aimed to remove nucleic acid. Resultant supernatant was mixed with an ammonium sulfate (AS) solution (3M), and a phosphate buffer (1M) to final AS concentration 1.2 M at pH 6.5 (Table S5.2). Hydrophobic interaction chromatography (HIC) was applied to remove nucleic acid. A HiTrap Butyl FF column (GE Healthcare Life Sciences, USA) was used in binding mode. AKTA Pure (GE Healthcare Life Sciences, USA) was adopted

in all liquid chromatography. The equilibration buffer was 100 mM phosphate buffer with 1.2 M AS, pH 6.5, and the elution buffer was 20 mM phosphate buffer, pH 6.5. After discarding the flow through fraction, target protein was directly eluted by 100 % elution buffer in 10 column volumes (CVs).

For other variants (C2, C3 and C4), the additional two steps involved desalting and ion exchange chromatography (IEC) and were required to improve protein purity. Before applying IEC as Step 3, protein samples were adjusted to the same pH and salt concentration as the equilibration buffer (20 mM PB, pH7.5). A HiPrep Desalting column with Sephadex G-25 resin (GE Healthcare Life Sciences, USA) was used by flow- through purification. For step 4, Sartobind Q membrane absorbers (3 mL) (Sartorius Stedim Biotech, Germany) were applied for the last step of purification of C2, C3 and C4 by a bind-elute process. Proteins were eluted with 100 % elution buffer (20 mM PB, 2M NaCl, pH7.5) in 16 CVs using an AKTA Pure System (GE Healthcare Life Sciences, USA). All peaks were collected to test purity using SDS-PAGE.

5.2.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein expression and purification was performed with 12 % (w/v) SDS separation gel with a 5 % (w/v) SDS stacking layer. Sample preparation and procedures to stain and wash gels were similar to previous published work [1].

5.2.6 Transmission electron microscopy (TEM)

TEM was conducted to characterize the conformations of purified $F_1L_3E_1$, C1, C2, C3, C4 and C5. Samples preparation was similar to previous published work [1]. Grids were analysed with a

Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

5.2.7 Size exclusion chromatography multi-angle light scattering (SEC-MALS)

SEC-MALS was performed to determine the hydrodynamic radius of purified proteins ($F_1L_3E_1$, C1, C2, C3, C4 and C5). Superose 6 Increase 10/300 GL (GE Healthcare, USA) was connected to a high performance liquid chromatography (HPLC) system (Shimazdu, Japan) coupled with a multi-angle laser light scattering (MALS) Wyatt DAWN® HELEOS II and Optilab T-rEx (Wyatt Technology, Santa Barbara, CA USA). The equilibration buffer/mobile phase was 20 mM Phosphate buffer, pH 7.0. Protein concentration was 10 mg mL^{-1}. Loaded volume was 50 µl. Retention time and absorbance at 280 nm were recorded. MALS data were processed with ASTRA software (v. 6.1).

5.2.8 Dynamic light scattering (DLS)

A DLS Zetasizer Nano ZS90 (Malvern, UK) was used to characterization protein sizes. Purified protein samples were adjusted to a protein concentration of 0.3 mg mL^{-1}. The instrument was equilibrated at 25 ºC and every sample was measured three (3) times.

5.2.9 Hydrophobicity

Hydrophobicity was determined by hydrophobic interaction chromatography amongst $F_1L_3E_1$, C1, C2, C3, C4 and C5. Sample volume was adjusted to 2 mL with protein concentration 1 mg mL^{-1}. Each protein sample was mixed with 8 mL of equilibrated buffer of 1.0 M ammonium sulfate buffer, 100 mM phosphate buffer containing Na_2HPO_4 and NaH_2PO_4, pH 6.5. A HiTrap Butyl FF column (GE Healthcare, USA) was used. The elution buffer was 20 mM phosphate

buffer, pH 6.5. The column was equilibrated and protein samples loaded to a HiTrap Butyl FF column. The column was eluted with a buffer by linear gradient elution of 0 to 100 % elution buffer in ten (10) CVs. Retention time for each protein sample was recorded using AKTA Pure (GE Healthcare, USA). After each run, the column was regenerated using 1.0 M sodium hydroxide.

5.2.10 Thermal stability by differential scanning calorimetry (DSC)

Nano DSC (TA Instruments, USA) was used to determine the thermal stability of proteins. Samples were prepared at protein concentration 5 mg mL^{-1} in 200 mM phosphate buffer (Na_2HPO_4 and NaH_2PO_4) at pH 7.0. For DSC measurement equilibration was set at 30 min with scan rate of 90 °C h^{-1} for a temperature ramp from 30 to 110 °C. Data were analysed using software *Nano Analyze* (TA Instruments, USA). Raw data from the experiment were subtracted from the reference buffer scan. The baseline was fitted and subtracted from the thermogram using a sigmoidal baseline function. Differences in heat capacity of the folded and unfolded states of protein were determined [22]. The midpoint of transition temperatures (T_m) and calorimetric enthalpy (ΔH) were determined at end of analysis.

5.3 Results and discussion

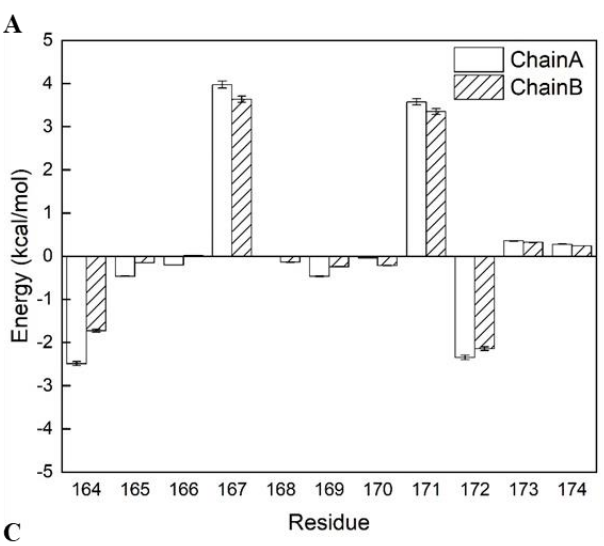
5.3.1 Hot spots determination by MDS

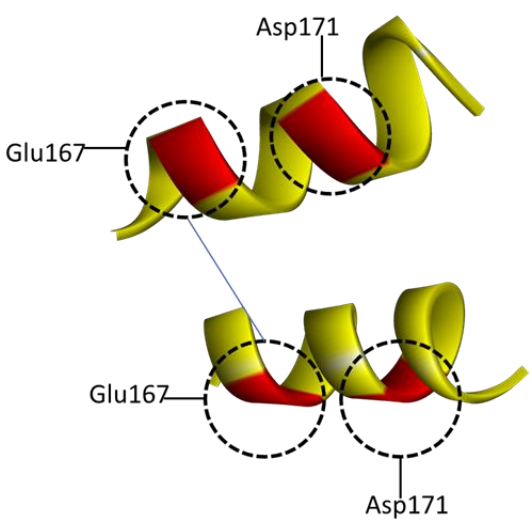
To determine the key residues that contribute significantly to protein stability, MDS coupled with MM-PBSA were used. The hot spots are defined as the residues making significant contributions to the binding free energy with a criterion of ± 2.5 kcal mol^{-1} [21] [23]. Negative binding energies indicate the equilibrium is in favour of the bound complex, however, positive binding energies indicate the equilibrium is tilted towards the unbound complex [24]. The hot spots in C-terminus of $F_1L_3E_1$ were identified as Glu 167 and Asp 171 with free binding energy of 3.97 kcal mol^{-1} and 3.58 kcal mol^{-1}, respectively (Figure 5.4). Positive values also indicate that these two residues favour the formation of clusters of monomers against complex formation.

To further validate the importance of these two residues, amino acids (aa) 167 and 171, C-terminus variants C1 and C2 were constructed by replacing Glu and Asp with Gln (C1) and Arg (C2), respectively. The other three variants (C3, C4 and C5) were also constructed without replacing aa 167 and aa 171, but rather by altering the hydrophobic interface. Figure 5.5 shows the hot spots in each variants, highlighted in red. Table 5.2 summarises the energy distribution by the main residues in C-terminus of variants (C1, C2, C3, C4 and C5). It was found that in C1, aa 167 and aa 171 were not hot spots anymore compared with $F_1L_3E_1$, however, aa 172 became a new hot spot, with free binding energy of 3.40 kcal mol^{-1}. In C2, aa 167 and aa 171 were still hot spots with increasing free binding energy of 9.18 and 8.17 kcal mol^{-1}, respectively. There was an additional hot spot found at the position of aa 172. For C3, C4 and C5, because residues at aa 164 and aa 171 were not changed, hot spots remained the same. Interestingly, for C3 and C4, one (1) additional hot spot was found at aa 172 with free binding energies of -5.99 and -2.77 kcal mol^{-1}, respectively. The negative binding energies suggested that residues favour

the bound complex. For C5, there was no additional hot spot. The newly replaced residues presented positive free binding energy, however, they were significantly close to zero, which is unlikely to affect protein folding.

Findings demonstrate that aa 167 and aa 171 are two hot spots that potentially affect the stability of the C-terminus. Except for C1, other variants all showed significantly greater binding free energy than the control $F_1L_3E_1$. Besides aa 167 and aa 171, aa 172 was found to be another hot spot after mutating native hot-spot residues. C5 is the variant without replacing residues at previous hot spots, therefore, no additional hot spot was found. It can be concluded here that all five variants contained hot spots, which significantly affect native C-terminus stability. However, all of these assumptions require validation using experimental data.





C

	Gly164	Leu165	Ala166	**Glu167**	Tyr168	Leu169	Phe170	**Asp171**	Lys172	His173	Thr174
Chain A	-2.49	-0.46	-0.20	**3.97**	0.00	-0.47	-0.04	**3.58**	-2.35	0.36	0.27
Chain B	-1.72	-0.15	0.02	**3.64**	-0.14	-0.24	-0.22	**3.35**	-2.14	0.32	0.24

Figure 5.4 Hot spot predictions in C-terminus of $F_1L_3E_1$ by MDS. A) Energy distribution of residues at aa 164 to aa 174 in C-terminus of $F_1L_3E_1$. B) Mimic diagram of hot spots in C-terminus of $F_1L_3E_1$ by Discovery Studio. Hot spots are indicated in red. C) Energy distribution table. Hot spots are highlighted in bold.

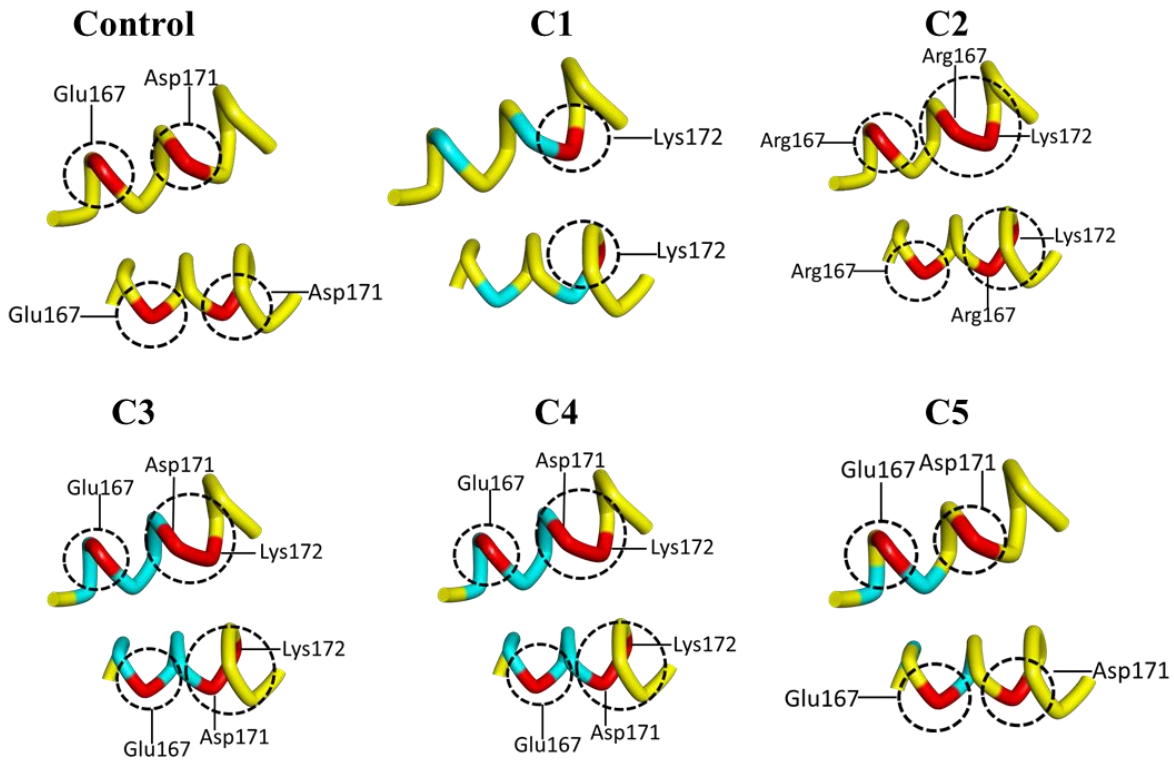


Figure 5.5 The identified hot spots predicted in C-terminus of 5 variants by MDS. Hot spots are indicated in red. Mutated residues are indicated in blue.

Table 5.2 Energy distribution by main residues in C-terminus of variants. Bold indicates the hot spots; Red indicates replaced residues.

Protein	**Binding free energy contribution of each residue in C-terminus variants (kcal mol^{-1})**										
$F_1L_3E_1$	Gly164	Leu165	Ala166	**Glu167**	Tyr168	Leu169	Phe170	**Asp171**	Lys172	His173	Thr174
	-2.49	-0.46	-0.20	**3.97**	0.00	-0.47	-0.04	**3.58**	-2.35	0.36	0.27
C1	Gly164	Leu165	Ala166	Gln167	Tyr168	Leu169	Phe170	Gln171	**Lys172**	His173	Thr174
	-1.94	-1.37	0.05	0.21	-0.97	-0.30	0.14	0.51	**3.40**	0.55	0.32
C2	Gly164	Leu165	Ala166	**Arg167**	Tyr168	Leu169	Phe170	**Arg171**	**Lys172**	His173	Thr174
	-1.77	0.01	-0.03	**9.18**	-0.41	-0.55	-0.27	**8.71**	**8.78**	-0.16	0.08
C3	Gly164	Val165	Ile166	**Glu167**	Ile168	Val169	Ile170	**Asp171**	**Lys172**	His173	Thr174
	-1.67	-1.30	-1.68	**4.62**	-0.20	-1.04	-0.46	**4.82**	**-5.99**	-1.20	-0.37
C4	Gly164	Gln165	Asn166	**Glu167**	Gln168	Asn169	Gln170	**Asp171**	**Lys172**	His173	Thr174
	-2.40	0.06	0.10	**4.47**	0.18	0.29	-0.00	**4.34**	**-2.77**	0.07	0.12
C5	Gly164	Gln165	Ala166	**Glu167**	Gln168	Leu169	Phe170	**Asp171**	Lys172	His173	Thr174
	-2.36	0.49	0.05	**3.58**	0.27	-0.39	-0.13	**3.56**	-2.44	0.26	0.19

5.3.2 Molecular characterization by experiments

Characterization of engineered ferritin and variants was used to examine protein folding and compare protein structural differences. $F_1L_3E_1$ was used as a control.

Theoretical monomer MW predictions based on amino acid sequence for these six (6) proteins were highly similar ($F_1L_3E_1$: 22.6, C1: 22.6, C2: 22.7, C3: 22.5, C4: 22.60, and C5: 22.60 kDa). Therefore, the MW exhibited in SDS-PAGE for monomer of $F_1L_3E_1$ and variants (C1, C2, C3, C4 and C5) were similar (~ 23 kDa), Figure 5.6A. The purity levels $F_1L_3E_1$ (lane 2), C1, C2 and C5 (lane 3, 4 and 7, respectively) were similar (> 90 %). Purities for C3 and C4 (lane 5 and 6) were less than 70 %. Assembled protein MW differences were also demonstrated by size-exclusion chromatogram (SEC). The lower the value by elution volume, the greater MW of assembled proteins. As is illustrated in Figure 5.6B (and Table 5.3), the elution volume for C1, C2 and C5 were similar to the control (~ 12 mL). C3 displayed two obvious peaks on SEC (Peak 1: 11 mL; Peak 2: 13 mL). According to theoretical MW, peak 2 is assumed to be assembled single C3 protein particle, and peak 1 referred to the aggregated C3 particles. The theoretical assembled MW for C4 was similar to the control, however, a lower MW was indicated by SEC.

To observe particle size of variants, hydrodynamic radius by MALS and particle diameter by DLS were determined, shown in Table 5.3. The hydrodynamic radius for C1 (6.96 nm), C2 (7.33 nm) and C5 (7.72 nm) were found to be greater than the control ($F_1L_3E_1$: 6.62 nm). For C3, peak 2 from SEC was assumed to represent an assembled C3 protein particle, which exhibited hydrodynamic radius as 6.12 nm with greater error compared with others. The hydrodynamic radius for C4 was unmeasurable because the reading from MALS was wildly inaccurate (R_h = 1.39 nm ± 19.30 %) (Figure S5.1, Supporting information). DLS was applied

as supporting information to compare particle size of these variants. Compared with the readings from MALS, the deviation obtained from DLS was greater. However, the trend between these two different measurements was similar. For C4, the particle diameter was found to be 8.72 nm, which was less than the expected (ferritin: 12 nm) [25].

TEM images confirmed that whether protein particles were well assembled or not. Compared with the control, C1, C2, C5 exhibited clear nanoparticles with similar particle size (Figure 5.6C). For C3, nanoparticles with different particle sizes were observed (~ 30-100 nm) (Figure 5.6C, Figure S5.2, Supporting Information). The sizes of these particles were greater than the expected. These were more likely aggregates. There was no assembled C4 found using TEM (Figure 5.6C).

Protein characterization is an efficient indicator to determine if protein folds properly and protein conformationally stable. The particle size and assembled protein MW can indicate if protein folds properly. The proportion of protein aggregation suggests protein conformational stability [26]. The aggregation of protein is mainly driven by hydrophobic effects. The reduction in free surface energy by the removal of hydrophobic residues from contact with the solvent drives protein aggregation [27]. It was found that alternation of native electrostatic [28] or hydrophobic interface [29] affected differently on protein folding and protein conformational stability. Zhou *et al.* claimed that charged residues differ from nonpolar residues (e.g., Leu and Ile) [28]. The hydrophobic interactions of nonpolar residues (either hydrophobic or hydrophilic) are the driving force for protein folding and conformational stability [28 30 31].

Our findings demonstrated that alteration of hydrophobic interfaces affected native protein folding and conformational stability more seriously than alteration of the electrostatic interface

at the C-terminus of engineered ferritin. Figure S5.4 (Supporting Information) has shown all five variants were expressed with soluble proteins in *E. coli*. The purification protocol for C1 and C5 was similar to our previous study [1]. However, because of lower tolerance to thermal and pH denaturants for variants C2, C3 and C4, an additional step using anion membrane adsorbers was applied. Protein purification using membrane adsorbers offers advantages of fast, gentle and effective separation of target proteins from impurities. The successful purification of C2 indicated this additional step was non-invasive to protein stability. C1 and C2 are the variants that change native electrostatic interface of protein. C1 (replacing with non-charged residues) and C2 (replacing with positive charged residues) both folded properly and were stable such that they could be purified in two steps to achieve purity around 90% and correct particle size was also detected. C5 is the variant breaking the native hydrophobic interface at position 'a' and 'd' in a heptad repeat helix. Protein conformational stability of C5 variant was influenced negligibly, suggested by negligible proportion of protein aggregation and correct particle size. The purification of C3 and C4 failed to achieve a satisfactory purity level possibly because of changes to the protein itself (unfolded and/or unstable). Variant C3 replaced residues with the most hydrophobic residues Val and Ile, which may eventually result in greater protein hydrophobicity change, therefore, greater proportion of protein aggregation in C3 is observed. Variant C4 replaced residues with the most hydrophilic residues Gln (Q) and Asn (N). The flip of native hydrophobic interface affects significantly on protein proper folding. It seems that C4 was not assembled/folded correctly, therefore, small particles with smaller size and low MW were observed.

It can be concluded based on variants characterization that, variants C1, C2 and C5 achieved more than 90% purity level with reasonable assembled molecular weight and particle size, which provided more reliable protein-stability-study data compared with C3 and C4. Variants

C3 and C4 obtained unsatisfied purity level (< 70 %), which may be caused by less protein stability and therefore greater difficulty in protein purification. Variant C3 is highly likely to aggregate, indicated by obvious two peaks in SEC and broad deviation of particle size in TEM. Variants C4 exhibited less particle size and molecular weight, which is most likely caused by incorrect protein folding or unstable proteins.

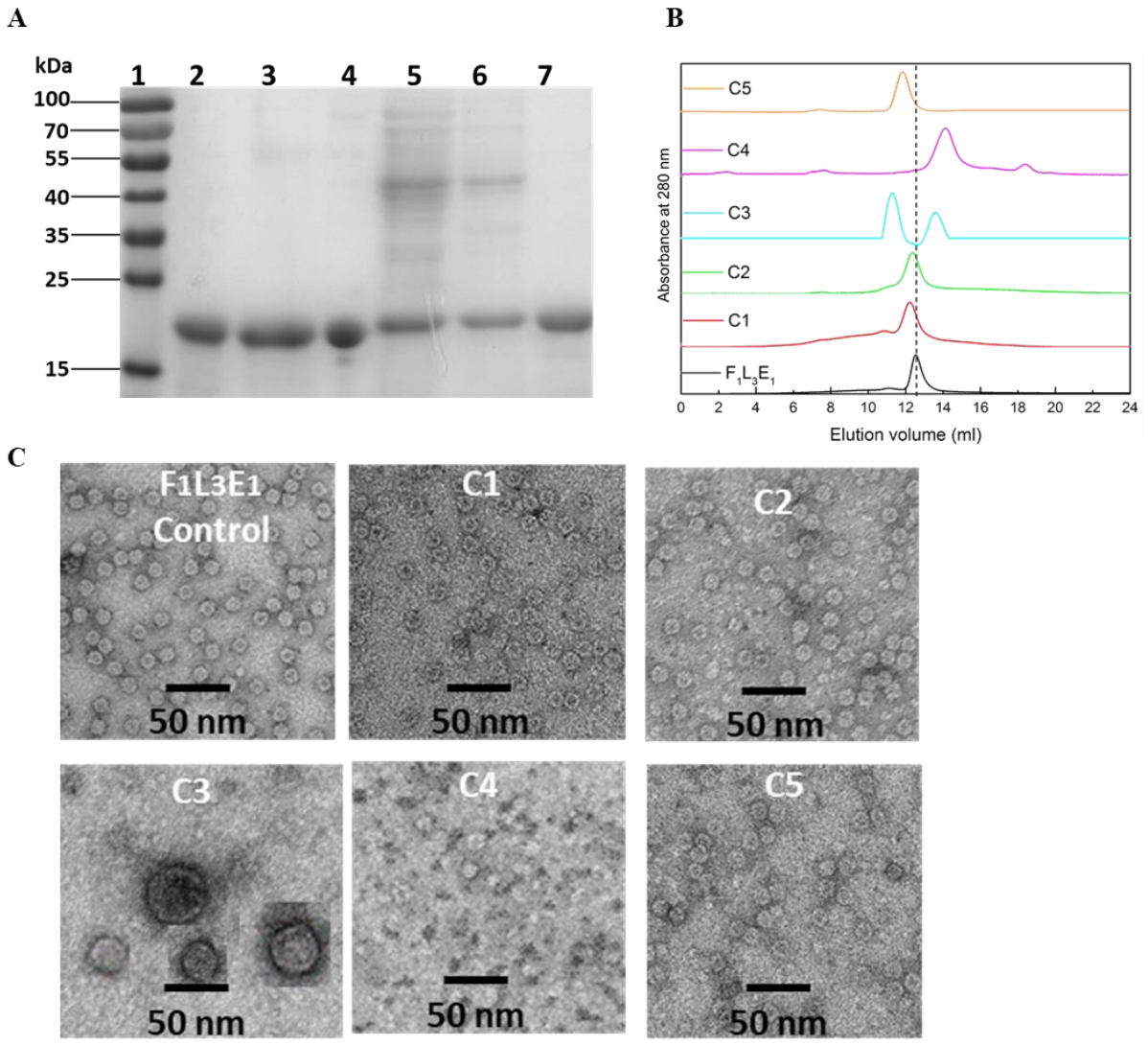


Figure 5.6 Characterization of $F_1L_3E_1$ and five (5) variants (C1, C2, C3, C4 and C5). A) SDS-PAGE analysis post-purification: lane 1, marker; lane 2, $F_1L_3E_1$; lane 3, C1; lane 4, C2; lane 5, C3; lane 6, C4; lane 7, C5. B) Size exclusion chromatography Multi Angle Scattering (SEC-MALS). C) TEM images. Proteins were prepared in 20 mM Phosphate buffer, pH 7 with concentration of 0.1 mg mL^{-1}.C3 image is formed by combining multiple C3 TEM images at the same scale in order to show deviation of particle size. Original TEM images are shown in Figure S5.2, Supporting Information.

Table 5.3 Summary comparison of particle size and molecular weight (MW) for $F_1L_3E_1$ and five (5) variants (C1, C2, C3, C4 and C5).

Protein	Assembled molecular weight (kDa)	Hydrodynamic radius from SEC-MALS (nm)	Elution volume from SEC (mL)	Particle diameter from DLS (nm)
$F_1L_3E_1$	542.5	6.62 (± 0.78 %)	12.53	15.69 (± 5.21)
C1	542.8	6.96 (± 0.73 %)	12.26	18.17 (± 5.52)
C2	544.1	7.33 (± 0.58 %)	12.3	18.17 (± 4.78)
C3	540.8	6.12 (± 3.00 %)	11.24; 13.38	11.70 (± 1.22)
C4	542.6	N/A	14.10	8.72 (± 2.51)
C5	542.0	7.72 (± 0.45 %)	11.90	21.04 (± 8.18)

5.3.3 Hydrophobic study by MDS and experiment

To determine the impact of mutation on protein surface hydrophobicity, the six proteins, $F_1L_3E_1$, C1, C2, C3, C4 and C5 were simulated in water environment first to predict hydrophobic solvent accessible surface area (SASA). The greater hydrophobic SASA, the more hydrophobic the protein [32,33]. Shown by Figure 5.7A and Table 5.4, average hydrophobic SASA for each variant was highly similar to the control ($F_1L_3E_1$: 827.5, C1: 807.1, C2: 848.0, C3: 843.6, C4: 809.4, and C5: 842 nm^2).

These 6 proteins were loaded to hydrophobic interaction column (HIC) to determine protein surface hydrophobicity by experiment. Proteins were eluted by reducing salt concentration. The shorter the retention time before appearance of major elution peak the weaker the protein binding to the column. Therefore, the less hydrophobic the protein. Because of low purity and significant proportion of aggregated and/or unfolded protein, experimental data for variants C3 and C4 were excluded from analysis.

Figure 5.7B and Table 5.4 demonstrated that there was no significant difference of retention time between variants (C1: 33.9; C2: 30.8 min; C5: 33.9 min) and the control ($F_1L_3E_1$: 34.0 min). For variants C3 and C4 (Figure S5.3, Supporting Information), a significant amount of protein could not be easily eluted from the column and could only be removed by the denaturant sodium hydroxide (1.0 M). This result can be caused by two (2) reasons: 1) the protein was too hydrophobic to be eluted by the elution buffer; 2) unstable/unfolded protein precipitated in HIC. Simulation data has shown that well-assembled variants should have similar protein hydrophobic surface area. In addition, proven by protein characterization study, variant C3 was likely to form clusters because of low structural stability, and variant C4 failed to assemble. Therefore, C3 and C4 are more likely to precipitate in HIC.

It is noticed that all variants replaced residues at Helix E of C-terminus. The C-terminus was located inside the ferritin cage [34]. Therefore, well-assembled protein surface hydrophobicity of each variant is similar to the control, as is illustrated by Figure 5.7C, even though the hydrophobicity of the C-terminus in each variant was different (Figure 5.7D). The hydrophobicities of the C-terminus in C1 and C2 were similar to that for $F_1L_3E_1$, however, the hydrophobicity of the C-terminus in C3 was significantly increased, shown by greater hydrophobic area (white) occupied than for $F_1L_3E_1$ in Figure 5.7D. The C-terminus of C4 was found to decrease native hydrophobicity significantly, shown by a larger hydrophilic area (blue) occupied than $F_1L_3E_1$. C5 replaced two hydrophobic residues with polar uncharged residues (Gln), and therefore decreased native C-terminus hydrophobicity. These changes all contribute to significant changes in protein stability.

Proven by simulation and experiment, well-assembled variants C1, C2 and C5 had negligible changes in protein surface hydrophobicity. Our previous work has shown that insertion of epitopes at the C-terminus of ferritin (inside ferritin cage) caused less changes to protein surface hydrophobicity compared with insertion at the N-terminus (ferritin surface) [1]. Research demonstrated by Calhoun *et al.* also found that mutation on the solution-accessible outer surface of the protein impacted more significantly on the stability than the inner hydrophobic core [35]. Protein stability is closely related to protein hydrophobicity [36] [37]. Therefore, this research indirectly proves that the mutation of the hydrophobic core impacts less on protein hydrophobicity compared with the one of outer surface. Consequently, C1, C2 and C5 had similar protein surface hydrophobicities to $F_1L_3E_1$.

For C3 and C4, because of lower purity levels compared with other variants and the control, there is not enough reliable evidence from experiment that validated simulation results

suggesting ‘C3 and C4 were similarly hydrophobic to $F_1L_3E_1$’. The limitation of simulation results is that, C3 and C4 were assumed to have similar assembled protein structure to $F_1L_3E_1$. However based on protein characterization results, C3 was found to be highly likely aggregate. C4 was not folded properly. Therefore, the hydrophobic SASA predicted by MDS for C3 and C4 were not accurate enough without consideration of actual protein structures. The C-terminus hydrophobic modelling (Figure 5.7D) for C3 and C4 indicate significant changes on native C-terminus hydrophobicity by flipping native hydrophobic interface, which may significantly affect correct protein folding and therefore protein structural stability.

It can be concluded from hydrophobicity study that, well-assembled variants (C1, C2 and C5) obtained similar protein surface hydrophobicity, explained by both simulation and experiment. However, the replacement of native hydrophobic residues in the C-terminus of engineered ferritin caused irreversible impact on protein folding and stability, therefore, variants C3 and C4 were unable to be characterized from this hydrophobic study.

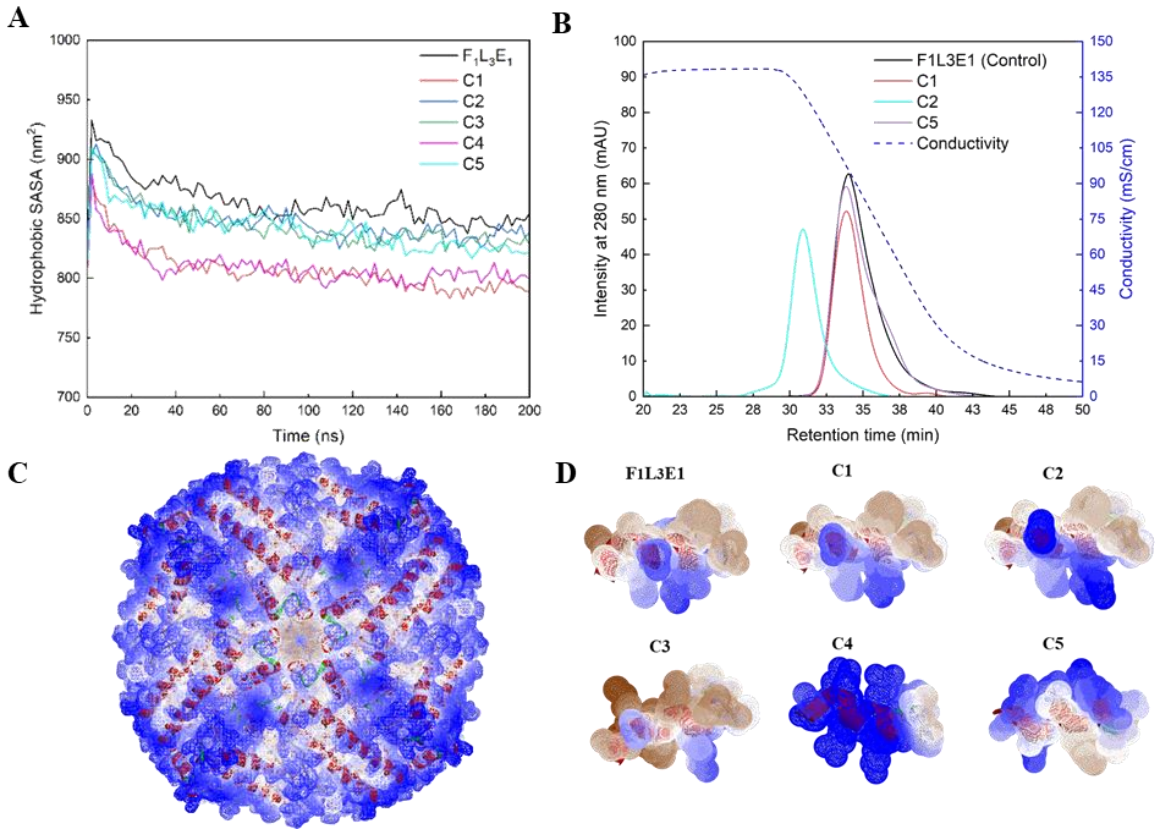


Figure 5.7 Hydrophobic study by MDS and experiment. A) Hydrophobic solvent accessible surface area (SASA) predicted by MDS. B) Hydrophobic interaction chromatography of variants. C) Hydrophobic and hydrophilic surface modelling for engineered ferritin. $F_1L_3E_1$ is used as an example. Hydrophilic residues are highlighted in blue. Hydrophobic residues are highlighted in white. D) Hydrophobic and hydrophilic surface modelling for C-terminus of $F_1L_3E_1$ and variants. Hydrophilic residues are highlighted in blue. Hydrophobic residues are highlighted in white. Neutral residues are highlighted in brown.

Table 5.4 Hydrophobicity- related parameters from simulations and experiments.

Molecule	Average hydrophobic SASA by simulation (nm^2)	Retention time by experiment (min)
$F_1L_3E_1$	827.5 (± 0.3)	34.0
C1	807.1 (± 17.1)	33.9
C2	848.0 (± 14.7)	30.8
C3	843.6 (± 15.9)	-
C4	809.4 (± 14.7)	-
C5	842.0 (± 18.5)	33.9

5.3.4 Thermal stability study by MDS and experiment

MDS prediction

Because C3 and C4 cannot assemble into correct conformations, they were excluded in this thermal stability study by both MDS and experiment. Four (4) assembled engineered ferritins ($F_1L_3E_1$, C1, C2, and C5) were simulated in water at room temperature 27 °C (300 K) and 150 °C (423 K), respectively, to compare protein conformational stability, compactness and flexibility in terms of root mean square deviation (RMSD), radius of gyration (R_g) and root mean square fluctuation (RMSF), respectively. At 150 °C, α-helix content was calculated to predict protein secondary structure change in response to thermal denaturation.

Figure 5.8 demonstrates comparisons of protein stability, compactness and flexibility between $F_1L_3E_1$ and variants (C1, C2 and C5) at 27 °C. The error analysis has shown the simulation converged with acceptable deviation (Table S5.5, Supporting Information). RMSD refers to protein structural stability. The greater RMSD, the less stable the protein structure is in MDS [38-40]. As is shown in Figure 5.8A, protein structural stability for C1, C2 and C5 was similar to $F_1L_3E_1$ in terms of similar variance of RMSD of C_α (< 0.1 nm). R_g is an indicator of protein compactness, which depends on both protein particle size and protein conformational stability. A low R_g value illustrates a compacted protein structure, therefore, greater protein stability [41-42]. Protein compactness for C1, C2 and C5 was similar to $F_1L_3E_1$, shown by similar variance of R_g of C_α in Figure 5.8B (< 0.1 nm). These results showed that protein structural and conformational stability for C1, C2 and C5 at 27 °C were similar to the control, $F_1L_3E_1$.

In order to determine the fluctuation of residues at α-helix of C-terminus in engineered ferritin, RMSF was determined for C_α of residues from aa 164 to aa 173. RMSF examines protein flexibility. A greater RMSF value suggests a more fluctuated movement of protein structure with likely lower stability [33], [43], [44]. Low average RMSF values indicate individual amino acid

residues that preserve protein stability [45]. Shown by Figure 5.8C, the RMSF range of $F_1L_3E_1$ was between 0.10 and 0.16 nm, while variants (C1: 0.10 to 0.22 nm, C2: 0.12 to 0.27 nm, and C5: 0.13 to 0.24 nm) exhibited greater flexibility than $F_1L_3E_1$. High fluctuation for all proteins can be noticed at residues positions 164 and 173, which are two end points of the α-helix. Hot spots predicted previously were aa 167, aa 171 and aa 172. At aa 167, variant C2 exhibited the greatest flexibility with a fluctuation of 0.18 nm, compared with others. Variant C1 exhibited highly similar RMSF as $F_1L_3E_1$ at aa 167. It can be concluded here that, at room temperature, variants (C1, C2 and C5) displayed very similar conformational and structural stability to the control, however, notably greater fluctuations were observed.

To predict protein stability against thermal denaturants, $F_1L_3E_1$, C1, C2, and C5 were heated at 150 °C for 200 ns by GROMACS. Shown by Figure 5.9, variants C1 and C5 exhibited similar RMSD to $F_1L_3E_1$, while variant C2 showed the most unstable protein conformational structure indicated by the greatest RMSD and R_g (Figure 5.9A and Figure 5.9B). Compared with Figure 5.8C, heating treatment results in a more fluctuating structure, shown by greater RMSF (Figure 5.9C). Helix content was calculated to demonstrate protein secondary structure change. As is indicated in Figure 5.9D, variants C1 (15.1 % reduced) and C5 (14.6 % reduced) resulted in slightly less helix content change compared with $F_1L_3E_1$ (13.8 % reduced). Secondary structure of variant C2 was changed more significantly against thermal denaturants (19.7 % reduced). It can be predicted here that, variant C2 was the least thermally stable. Variants C1 and C5 exhibited similar protein thermal stability to $F_1L_3E_1$ for heating at 150 °C for 200 ns.

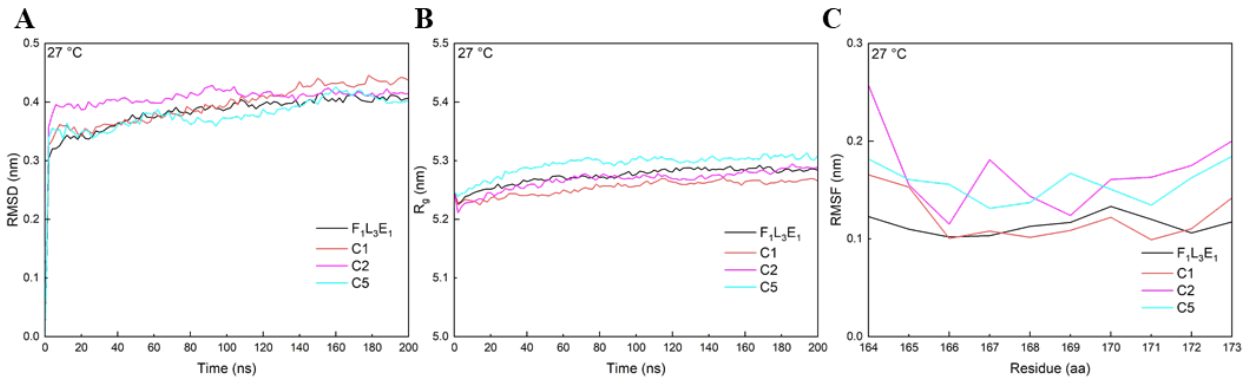


Figure 5.8 RMSD and R_g for assembled $F_1L_3E_1$, C1, C2 and C5 and RMSF for C-terminus from aa 164 to aa 173 using GROMACS at 27 °C for 200 ns. A) RMSD. B) R_g. C) RMSF for one of C-termini (aa 164 to aa 173) from assembled $F_1L_3E_1$, C1, C2 and C5. $F_1L_3E_1$: 0.10-0.16 nm; C1: 0.10-0.22 nm; C2: 0.12-0.27 nm; and C5: 0.13-0.24 nm.

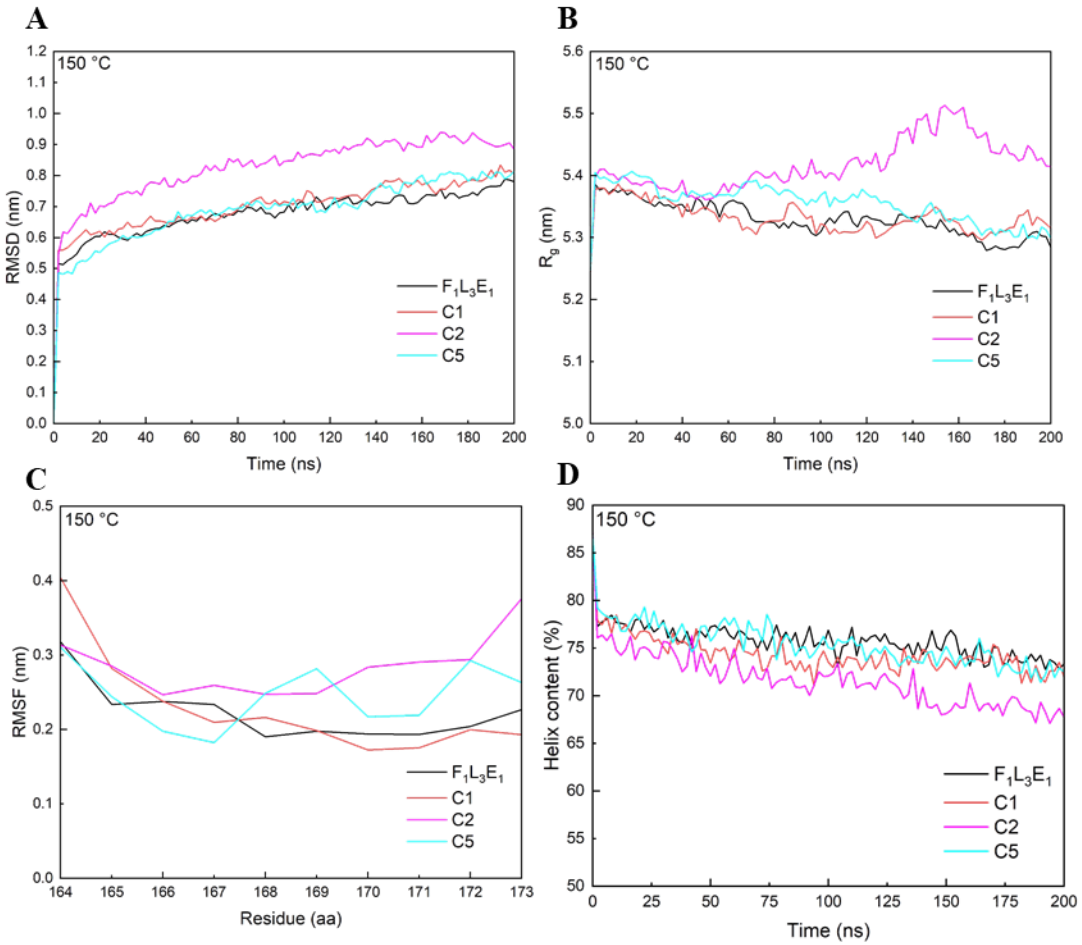


Figure 5.9 RMSD, R_g, helix content for assembled $F_1L_3E_1$, C1, C2 and C5 and RMSF for C-terminus from aa 164 to aa 173 using GROMACS at 150 °C for 200 ns. A) RMSD. B) R_g. C) RMSF of C-termini (aa 164 to aa 173) from assembled $F_1L_3E_1$, C1, C2 and C5. $F_1L_3E_1$: 0.19-0.32 nm; C1: 0.17-0.40 nm; C2: 0.25-0.38 nm; and C5: 0.18-0.31 nm. D) Reduction of helix content (%). $F_1L_3E_1$: 13.77%; C1: 15.07 %; C2: 19.70 %; and C5: 14.58 %.

Experimental study by DSC

DSC was used to characterize protein thermal stability by determining the midpoint temperature of the transition (T_m) and calorimetric enthalpy (ΔH) [3] [36]. T_m is also known as protein denaturing temperature [46]. For multi-domains proteins, such as engineered ferritins ($F_1L_3E_1$, C1, C2 and C5), there are multiple T_m values. This is because there are multiple intermediates during the thermal denaturing process. The intermediate with the greatest enthalpy change indicates the greatest protein conformational change. Therefore, in this case, protein denaturing temperature was defined as the temperature at the peak with the greatest change of enthalpy values. Shown by Figure 5.10 and Table 5.5, variants (C1, C2 and C5) exhibited less thermal stability than $F_1L_3E_1$. The denaturing temperature for the control ($F_1L_3E_1$) was 90.2 °C, while denaturing temperature values for C1, C2 and C5 were 74.9, 64.6 and 69.7 °C, respectively (Table 5.5).

By comparing results from both MDS and experiment, it was found that MDS predicted protein thermal stability well. These data have confirmed that variants C1, C2 and C5 have decreased native protein thermal stability compared with the control $F_1L_3E_1$. Variants C1 and C2 changed electrostatic interfaces by replacing hot spot residues at position aa 167 and aa 171 with uncharged residues (C1) or positively charged residues (C2). The destruction of native electrostatic interface was found to destabilize the protein. In the native C-terminus of $F_1L_3E_1$, negative charged residues Asp (D) and Glu (E) maintained ion pairs with positive charged residues Lys (K) and His (H) (Figure 5.1). These residues form strong hydrogen bonds and other non-specific electrostatic interactions to support protein folding and retain protein thermal stability [28]. Therefore, removal of native ion pairs and hydrogen bonds decreased native thermal stability. In addition, C2 was less thermally stable than C1, which was caused by greater binding free energy at hot spots aa 167 and aa 171 in C2 than C1.

C3, C4 and C5 are variants destroying the native hydrophobic interface. Altering hydrophobic interfaces affected more significantly protein folding and stability compared with altering electrostatic interfaces. The hydrophobic effect has been found to play a crucial role in protein folding and protein stability [47] [48]. The hydrophobic effect leads structures to pack nonpolar side chains in the protein interior to avoid contact with aqueous solution [49]. Research has already proven that substitutions of buried or partly buried nonpolar residues (hydrophobic or hydrophilic) in proteins changed thermal stability between the wild type and variants [49] [50] [51]. Variants C3 and C4 were unlikely to assume correct protein assembled structure, which has been proved in the protein characterization study. Therefore, no reliable thermal stability data could be generated. Variant C5 was formed by replacing non-polar residues at aa 165 and aa 168 with polar non-charged residues Gln, which destroyed the hydrophobic interface formed at 'a' and 'd' in a heptad repeat. Therefore, C5 was less thermally stable compared with $F_1L_3E_1$.

Overall, it can be concluded based on thermal stability study by MDS and experiments that, alteration of the electrostatic interface in the native C-terminus of $F_1L_3E_1$ destroyed native ion pairs and hydrogen bonds, therefore, decreased thermal stability was found in variants C1 and C2. C2 was less thermally stable than C1 because of greater binding free energy at hot spots aa 167 and aa 171. Alteration of the hydrophobic interface in the native C-terminus of $F_1L_3E_1$ destroyed native hydrophobic interactions between 'a' and 'd' in a heptad repeat, therefore, C5 was less thermally stable than $F_1L_3E_1$.

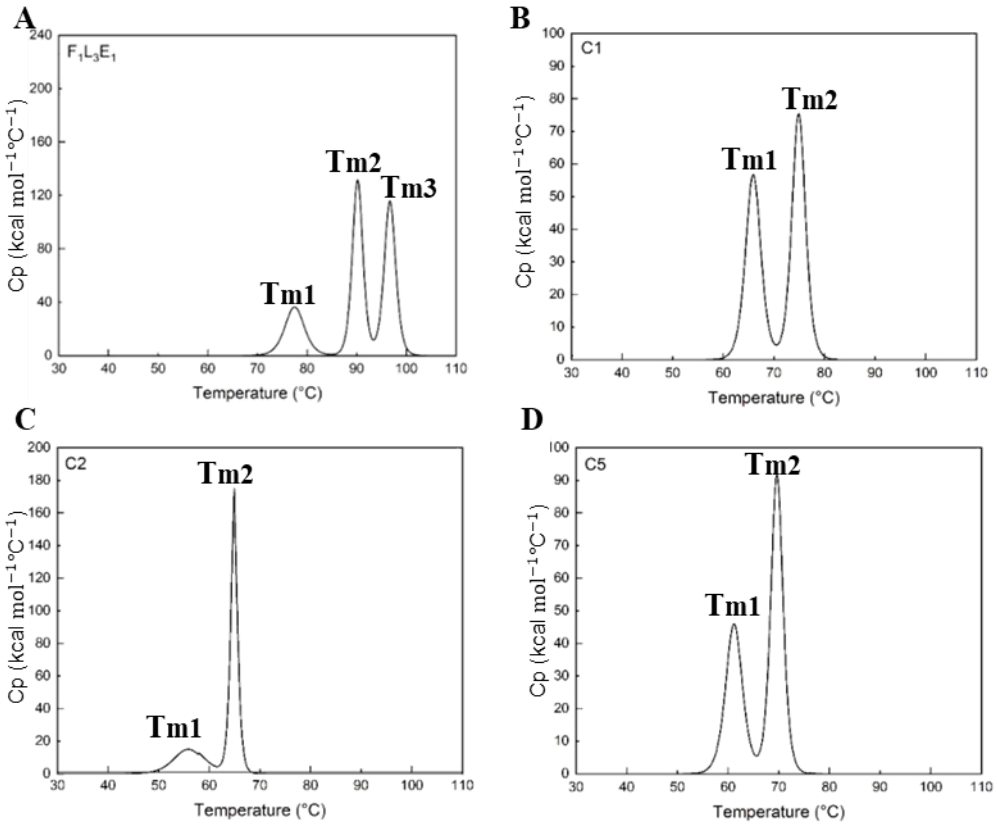


Figure 5.10 DSC thermogram for engineered ferritin and its variants. A) DSC thermogram for $F_1L_3E_1$ (control). B) DSC thermogram for C1. C) DSC thermogram for C2. D) DSC thermogram for C5.

Table 5.5 Disassembled temperature values and enthalpy change values for ferritin and its engineered variants obtained from DSC.

Protein	**Disassembled temperature T_m (°C)**			**Enthalpy change ΔH (kcal mol^{-1})**		
	T_{m1}	T_{m2}	T_{m3}	ΔH_1	ΔH_2	ΔH_3
$F_1L_3E_1$	77.5	**90.2**	96.2	189	**372**	355
C1	65.9	**74.9**	-	228	**270**	-
C2	55.8	**64.6**	-	114	**413**	-
C5	61.2	**69.7**	-	202	**293**	-

5.4 Conclusions

A comprehensive protein stability study for engineered ferritin $F_1L_3E_1$ and its engineered variants by MDS and experiment has shown that:

1. There are three potential hot spots affecting C-terminus stability predicted by MDS, aa 167, aa 171 and aa 172. Five variants were constructed by considering hot spots and electrostatic/hydrophobic interfaces. C1, substituted with non-charged residues; C2, substituted with positive-charged residues; C3, substituted with the highly hydrophobic residues; C4, substituted with the very hydrophilic residues; C5, substituted with polar non-charged residues to break the hydrophobic interface.
2. Change of the hydrophobic interface affected more significantly protein folding and stability compared with changes to the electrostatic interface.
3. Protein characterization study has shown that variants C1, C2 and C5 could be folded properly with good conformational stability. However, variants C3 and C4 introduced significant changes to the hydrophobic effect. Variant C3 exhibited a significant proportion of aggregates, and variant C4 was not folded properly. Variants C3 and C4 did were of insufficient quality for further stability studies.
4. Hydrophobic study by both MDS and experiment showed that variants C1, C2 and C5 displayed highly similar hydrophobicity to $F_1L_3E_1$. This is because the C-terminus is located inside of the ferritin cage. The substitution of residues had negligible effect on the protein surface. However, there were significant amount of variants unable to be eluted from HIC because of reduced protein stability.
5. Thermal stability study demonstrated that variants C1, C2 and C5 had decreased thermal stability compared with $F_1L_3E_1$.

This study of engineered ferritin C-terminus variants provides essential understanding that the α-helix of the C-terminus is significantly important in maintaining integrated protein stability. The molecular design of the C-terminus α-helix of engineered ferritin is essential to enhance protein stability, and therefore confidence in purification and production of safe and efficacious epitope-based chimeric vaccine. This particularly saves costs from production of unstable and inefficacious vaccines.

5.5 References

1 Qu, Y. *et al.* Stability of engineered ferritin nanovaccines investigated by combined molecular simulation and experiments. *The Journal of Physical Chemistry B* **125**, 3830-3842, doi:10.1021/acs.jpcb.1c00276 (2021).

2 Qu, Y. *et al.* Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine*, doi:https://doi.org/10.1016/j.vaccine.2021.07.021 (2021).

3 Nemtseva, E. V. *et al.* Experimental approach to study the effect of mutations on the protein folding pathway. *PLoS One* **14**, e0210361-e0210361, doi:10.1371/journal.pone.0210361 (2019).

4 Choe, S. E. *et al.* Differential stabilization of two hydrophobic cores in the transition state of the villin 14T folding reaction. *Journal of Molecular Biology* **304**, 99-115, doi:10.1006/jmbi.2000.4190 (2000).

5 Lindberg, M. O. *et al.*. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4083-4088, doi:10.1073/pnas.0508863103 (2006).

6 Matthews, J. M. & Fersht, A. R. Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry* **34**, 6805-6814, doi:10.1021/bi00020a027 (1995).

7 Pilipczuk, J. *et al.* Role of the disulfide bond in stabilizing and folding of the fimbrial protein DraE from uropathogenic Escherichia coli. *Journal of Biological Chemistry* **292**, 16136-16149, doi:10.1074/jbc.M117.785477 (2017).

8 Viguera, A. R. *et al.* The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *Journal of Molecular Biology* **247**, 670-681, doi:10.1006/jmbi.1994.0171 (1995).

9 Ingrassia, R. *et al.* Mutations of ferritin H chain C-terminus produced by nucleotide insertions have altered stability and functional properties. *Journal of Biochemistry* **139**, 881-885, doi:10.1093/jb/mvj101 (2006).

10 Fan, R. *et al.* A helix swapping study of two protein cages. *Biochemistry* **48**, 5623-5630, doi:10.1021/bi900387t (2009).

11 Zhang, Y. & Orner, B. P. Self-assembly in the ferritin nano-cage protein superfamily.

International Journal of Molecular Sciences **12**, 5406-5421, doi:10.3390/ijms12085406 (2011).

12 Luzzago, A. & Cesareni, G. Isolation of point mutations that affect the folding of the H chain of human ferritin in E.coli. *The EMBO Journal* **8**, 569-576, doi:10.1002/j.1460-2075.1989.tb03411.x (1989).

13 Lupas, A. N. *et al.* The structure and topology of α-helical coiled coils. *Subcellular Biochemistry* **82**, 95-129, doi:10.1007/978-3-319-49674-0_4 (2017).

14 Paliakasis, C. D. & Kokkinidis, M. Relationships between sequence and structure for the four-alpha-helix bundle tertiary motif in proteins. *Protein Engineering* **5**, 739-748, doi:10.1093/protein/5.8.739 (1992).

15 Lindahl, E. *et al.* GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **7**, 306-317, doi:10.1007/s008940100045 (2001).

16 Showalter, S. A. & Brüschweiler, R. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB Force Field. *Journal of Chemical Theory and Computation* **3**, 961-975, doi:10.1021/ct7000045 (2007).

17 Yang, B. *et al.* Molecular docking and molecular dynamics (MD) simulation of human anti-complement factor H (CFH) antibody Ab42 and CFH polypeptide. *International journal of molecular sciences* **20**, 2568, doi:10.3390/ijms20102568 (2019).

18 Kumari, R. *et al.* g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling* **54**, 1951-1962, doi:10.1021/ci500020m (2014).

19 Li, J. *et al.* Molecular basis of the initial platelet adhesion in arterial thrombosis: molecular dynamics simulations. *Journal of Molecular Graphics and Modelling* **37**, 49-58, doi:10.1016/j.jmgm.2012.04.002 (2012).

20 Zhang, L. *et al.* Molecular energetics in the capsomere of virus-like particle revealed by molecular dynamics simulations. *The Journal of Physical Chemistry B* **117**, 5411-5421, doi:10.1021/jp311170w (2013).

21 Zhang, L. & Sun, Y. Biomimetic design of platelet adhesion inhibitors to block integrin α2β1-collagen interactions: Construction of an affinity binding model. *Langmuir* **30**, 4725-4733, doi:10.1021/la404599s (2014).

22 Durowoju, I. B. *et al.* Differential scanning calorimetry - a method for assessing the thermal stability and conformation of protein antigen. *Journal of Visualized Experiments*, 55262, doi:10.3791/55262 (2017).

23 Lafont, V. *et al.* Protein–protein recognition and interaction hot spots in an antigen–antibody complex: Free energy decomposition identifies “efficient amino acids”. *Proteins: Structure, Function, and Bioinformatics* **67**, 418-434, doi:https://doi.org/10.1002/prot.21259 (2007).

24 Woods, C. J. *et al.* Rapid decomposition and visualisation of protein–ligand binding free energies by residue and by water. *Faraday Discussions* **169**, 477-499, doi:10.1039/C3FD00125C (2014).

25 Lee, B.-R. *et al.* Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. *Scientific Reports* **6**, 35182, doi:10.1038/srep35182 (2016).

26 Chi, E. Y. *et al.* Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Science* **12**, 903-913, doi:10.1110/ps.0235703 (2003).

27 Berrill, A. *et al.* in *Peptide and Protein Delivery* (ed Chris Van Der Walle) 313-339 (Academic Press, 2011).

28 Zhou, H.-X. & Pang, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews* **118**, 1691-1741, doi:10.1021/acs.chemrev.7b00305 (2018).

29 Giovambattista, N. *et al.* Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proceedings of the National Academy of Sciences* **105**, 2274, doi:10.1073/pnas.0708088105 (2008).

30 Baldwin, R. L. Energetics of protein folding. *Journal of Molecular Biology* **371**, 283-301, doi:10.1016/j.jmb.2007.05.078 (2007).

31 Nick Pace, C. *et al.* Forces stabilizing proteins. *FEBS Letters* **588**, 2177-2184, doi:10.1016/j.febslet.2014.05.006 (2014).

32 Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports* **7**, 44651, doi:10.1038/srep44651 (2017).

33 Kumar, C. V. *et al.* Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A. *Advances in Bioinformatics* **2014**, 502618-502618, doi:10.1155/2014/502618 (2014).

34 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

35 Calhoun, J. R. *et al.* Computational design and characterization of a monomeric helical dinuclear metalloprotein. *Journal of molecular biology* **Vol.334**, 1101-1115 (2003).

36 Strub, C. *et al.* Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry* **5**, 9, doi:10.1186/1471-2091-5-9 (2004).

37 Malleshappa Gowder, S. *et al.* Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *The Scientific World Journal* **2014**, 7, doi:10.1155/2014/971258 (2014).

38 Sviben, D. *et al.* Investigation of the thermal shift assay and its power to predict protein and virus stabilizing conditions. *Journal of Pharmaceutical and Biomedical Analysis* **161**, 73-82, doi:https://doi.org/10.1016/j.jpba.2018.08.017 (2018).

39 Chen, Z. *et al.* Molecular dynamics simulation of barnase: contribution of noncovalent intramolecular interaction to thermostability. *Mathematical Problems in Engineering* **2013**, 504183, doi:10.1155/2013/504183 (2013).

40 Manjunath, K. & Sekar, K. Molecular dynamics perspective on the protein thermal stability: A case study using SAICAR synthetase. *Journal of Chemical Information and Modeling* **53**, 2448-2461, doi:10.1021/ci400306m (2013).

41 Arnittali, M. *et al.* Structure of biomolecules through molecular dynamics simulations. *Procedia Computer Science* **156**, 69-78, doi:https://doi.org/10.1016/j.procs.2019.08.181 (2019).

42 Pikkemaat, M. G. *et al.* Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering, Design and Selection* **15**, 185-192, doi:10.1093/protein/15.3.185 (2002).

43 Zhao, Y. *et al.* Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. *PLoS One* **10**, e0124377, doi:10.1371/journal.pone.0124377 (2015).

44 Burton, B. *et al.* A computational investigation on the connection between dynamics properties of ribosomal proteins and ribosome assembly. *PLoS Computational Biology* **8**, e1002530-e1002530, doi:10.1371/journal.pcbi.1002530 (2012).

45 Abdulazeez, S. Molecular simulation studies on B-cell lymphoma/leukaemia 11A (BCL11A). *American Journal of Translational Research* **11**, 3689-3697 (2019).

46 Cao, X. M. *et al.* Effects of protein and phosphate buffer concentrations on thermal denaturation of lysozyme analyzed by isoconversional method. *Bioengineered* **7**, 235-240, doi:10.1080/21655979.2016.1197629 (2016).

47 Camilloni, C. *et al.* Towards a structural biology of the hydrophobic effect in protein

folding. *Scientific Reports* **6**, 28285, doi:10.1038/srep28285 (2016).

48 Islam, M. M. *et al.* Hydrophobic surface residues can stabilize a protein through improved water–protein interactions. *The FEBS Journal* **286**, 4122-4134, doi:https://doi.org/10.1111/febs.14941 (2019).

49 Prevost, M. *et al.* Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96----Ala mutation in barnase. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10880-10884, doi:10.1073/pnas.88.23.10880 (1991).

50 Matsumura, M. *et al.* Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* **334**, 406-410, doi:10.1038/334406a0 (1988).

51 Matouschek, A. *et al.* Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-126, doi:10.1038/340122a0 (1989).

5.6 Supporting Information

5.6.1 Supporting figures

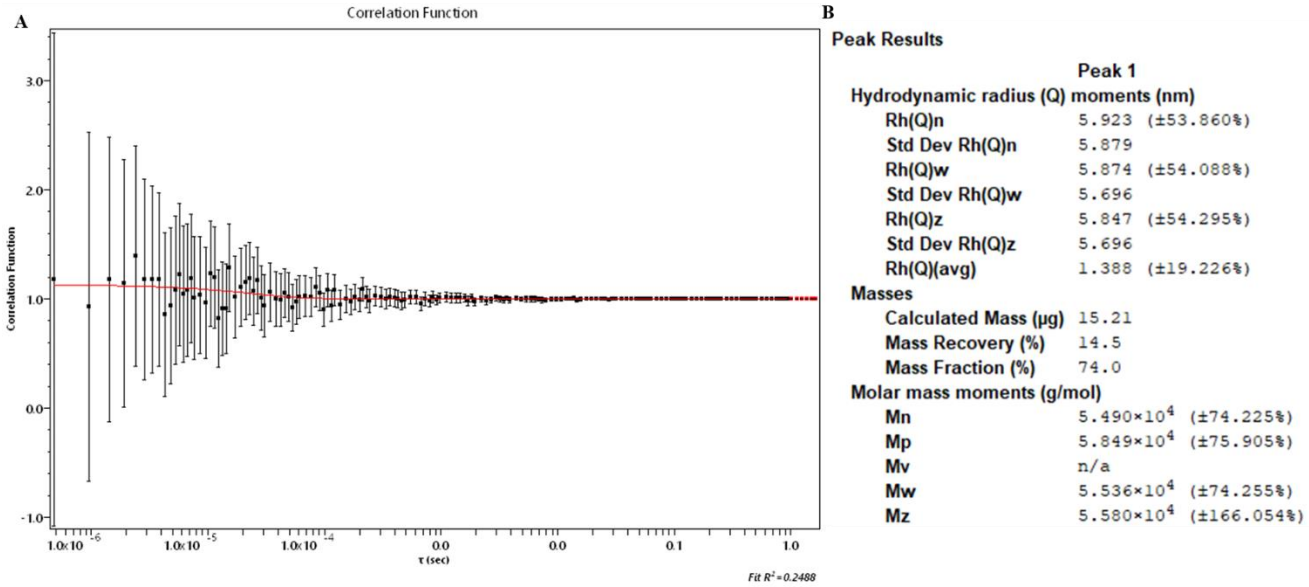


Figure S5.1 SEC-MALS results for C4. A) Correlation function. R2 is 0.2488. B) Results summary from SEC-MALS.

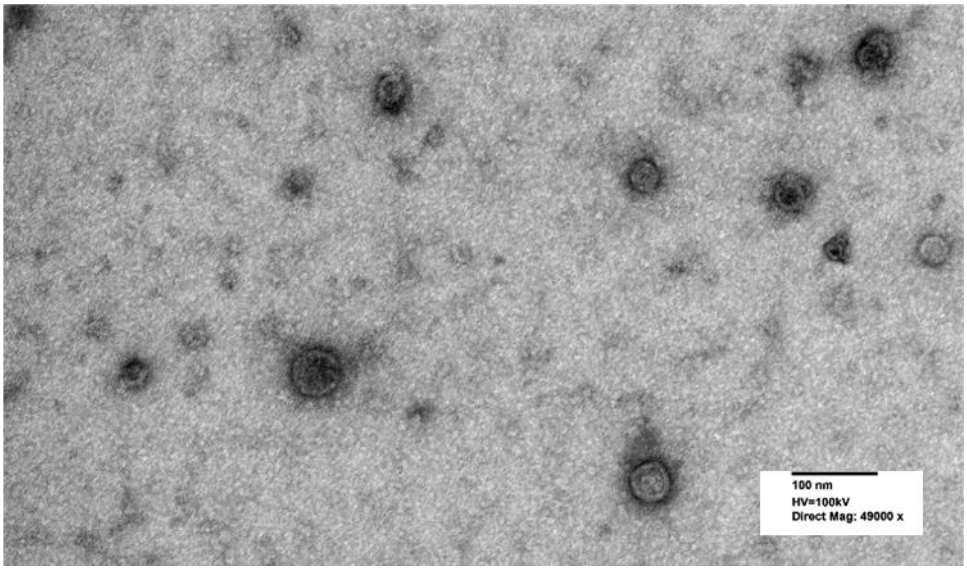


Figure S5.2 Original TEM image for C3 (20 mM Phosphate buffer, pH 7, protein concentration was 0.1 mg mL^{-1}).

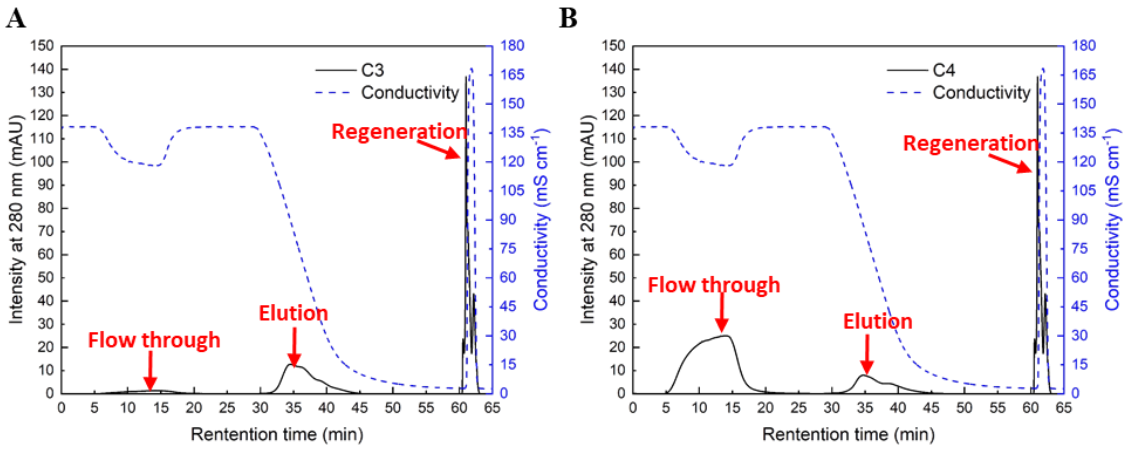


Figure S5.3 Hydrophobic interaction chromatography of variants C3 and C4. Flow through peak refers to protein that cannot bind to the column. Elution peak refers to protein that is eluted out from the column applying reducing salt gradient. Regeneration peak refers to protein that strongly binds to the column and can only being eluted out by introducing sodium hydroxide (1M).

5.6.2 Supporting tables

Table S5.1 Constructed PDB files by Discovery Studio.

PDB file name	Represented structure name
$F_1 L_3E_1$	The control, ferritin inserted with EBNA1 epitope at C-terminus with three residues of linkers (GGS)
C1	F1 L3E1 variant
C2	F1 L3E1 variant
C3	F1 L3E1 variant
C4	F1 L3E1 variant
C5	F1 L3E1 variant
CC	Two cut C-termini of $F_1 L_3E_1$
CCC1	Two cut C-termini of C1
CCC2	Two cut C-termini of C2
CCC3	Two cut C-termini of C3
CCC4	Two cut C-termini of C4
CCC5	Two cut C-termini of C5

Table 5.2 Cubic box for each structure generated from GROMACS.

PDB file name	Cubic box dimensions (nm*nm*nm)
$F_1L_3E_1$	20*20*20
C1	20*20*20
C2	20*20*20
C3	20*20*20
C4	20*20*20
C5	20*20*20

Table S5.3 Amino acid sequence of variants.

Name	Amino acid Sequence
Ferritin-EBNA1 (F1L3E1) **(Ferritin with inserted EBNA1 epitope at the C-terminus)** ***Control***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSH EEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNESGGSHPVGEADYFEY
C1 ***No charge***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQS HEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAQYLFQKHTLGDSDNESGGSHPVGEADYFEY
C2 ***Same charge***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQS HEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLARYLFRKHTLGDSDNESGGSHPVGEADYFEY
C3 ***Most hydrophobic***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQS HEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGVIEIVIDKHTLGDSDNESGGSHPVGEADYFEY
C4 ***Most hydrophilic***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQS HEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGQNEQNQDKHTLGDSDNESGGSHPVGEADYFEY
C5 ***Break L and Y combination***	TTASTSQVRQNYHQDSEAAINRQINLELYASYVYLSMSYYFDRDDVALKNFAKYFLHQSH EEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGQAEQLFDKHTLGDSDNESGGSHPVGEADYFEY

Table S5.4 Purification methods for control ($F_1L_3E_1$) and its variants (C1, C2, C3, C4 and C5).

Protein	Step 1: Pre-purification	Step 2: HIC	Step 3: Desalting	Step 4: IEC membrane
$F_1L_3E_1$	5M NaCl (1:5) 1M NaAc-HAc (pH4.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF	N/A	N/A
C1	5M NaCl (1:5) 1M NaAc-HAc (pH5.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF	N/A	N/A
C2	5M NaCl (1:5) 50 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF	Desalt to 20 mM PB pH 7.5	Load to Sartobind Q Elute proteins using buffer (2M NaCl, pH7.5)
C3	5M NaCl (1:5) 50 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF	Desalt to 20 mM PB pH 7.5	Load to Sartobind Q Elute proteins using buffer (2M NaCl, pH7.5)
C4	50 °C for 10 min	N/A	Desalt to 20 mM PB pH 7.5	Load to Sartobind Q Elute proteins using buffer (2M NaCl, pH7.5)
C5	5M NaCl (1:5) 1M NaAc-HAc (pH5.5) (1:10) 60 °C for 10 min	Dilute supernatant from step 1 with 1.2M AS, 100 mM PB, pH6.5 Load samples to Butyl FF	N/A	N/A

Table S5.5 Error analysis for MDS at 150 °C.

Protein	RMSD (mean ± SD) (nm)	R_g (mean ± SD) (nm)
$F_1L_3E_1$	0.67 (± 0.091)	5.33 (± 0.028)
C1	0.69 (± 0.100)	5.33 (± 0.024)
C2	0.70 (± 0.096)	5.42 (± 0.042)
C5	5.36 (± 0.031)	5.36 (± 0.031)

5.6.3 Other supporting information not included in online publication

Figure S5.4 indicated that all five (5) variants achieved soluble expression under expression condition: 37 °C, 200 rpm, 4 h.

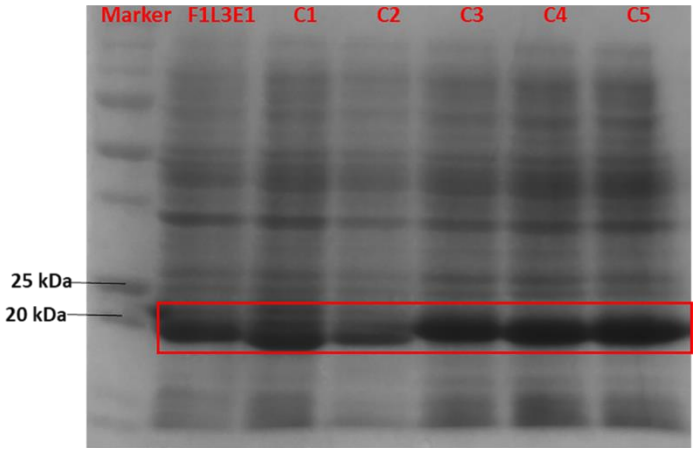


Figure S5.4 SDS PAGE image of soluble expression for $F_1L_3E_1$ and 5 variants (C1, C2, C3, C4 and C5).

CHAPTER 6 IMMUNOGENICITY STUDY OF ENGINEERED FERRITINS WITH C- AND N-TERMINUS INSERTION OF EPSTEIN-BARR NUCLEAR ANTIGEN 1 EPITOPE

Statement of Authorship

Title of Paper	Immunogenicity Study of Engineered Ferritins with C- and N- terminus insertion of Epstein-Barr nuclear antigen 1 epitope
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Qu, Y.; Zhang, B.; Wang, Y.; Yin, S.; Pederick, J. L.; Bruning, J. B.; Sun, Y.; Middelberg, A.; Bi, J. Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* 2021, 39, 4830-4841.

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Designed experiments, performed experiments; analysed data; writing manuscript		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30/07/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Bingyang Zhang		
Contribution to the Paper	Designed experiment, performed experiment		
Signature		Date	30/07/2021

Name of Co-Author	Yingli Wang		
Contribution to the Paper	Designed experiment, performed experiment, analysed data		
Signature		Date	09/09/2021

Name of Co-Author	Shuang Yin		
Contribution to the Paper	Performed experiments		
Signature		Date	30/07/2021

Name of Co-Author	Jordan L. Pederick		
Contribution to the Paper	Designed experiment, performed experiment, proof read the manuscript		
Signature		Date	30/07/2021

Name of Co-Author	John B. Bruning		
Contribution to the Paper	Designed experiment, proof read the manuscript		
Signature		Date	30/07/2021

Name of Co-Author	Yan Sun		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	20/09/2021

Name of Co-Author	Anton Middelberg		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/21

Name of Co-Author	Jingxiu Bi		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/2021

Immunogenicity study of engineered ferritins with C- and N- terminus insertion of Epstein-Barr nuclear antigen 1 epitope

Yiran Qu, Bingyang Zhang, Yingli Wang, Shuang Yin, Jordan L. Pederick, John B. Bruning, Yan Sun, Anton Middelberg, Jingxiu Bi*

Y. Qu, B. Zhang, S. Yin, A/Prof. Dr. J. Bi

School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide
SA 5005 (Australia)

Prof. Dr. Y. Wang

Shanxi University of Chinese Medicine, Shanxi, China

J. L. Pederick, Dr J. B. Bruning

Institute for Photonics and Advanced Sensing, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia

Department of Molecular and Cellular Biology, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia

Prof. Dr. Y. Sun
Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072 (China)

Prof. Dr. A. Middelberg

Division of Research and Innovation, The University of Adelaide, Adelaide, South Australia 5005, Australia

Vaccine 2021, 39, 34, 4830-4841 /doi.org/10.1016/j.vaccine.2021.07.021

Abstract

Human ferritin heavy chain, an example of a protein nanoparticle, has recently been used as a vaccine delivery platform. Human ferritin has advantages of uniform architecture, robust thermal and chemical stabilities, and good biocompatibility and biodegradation. There is however a lack of understanding about the relationship between insertion sites in ferritin (N-terminus and C-terminus) and the corresponding humoral and cell-mediated immune responses. To bridge this gap, we utilized an Epstein-Barr Nuclear Antigen 1 (EBNA1) epitope as a model to produce engineered ferritin-based vaccines E1F1 (N-terminus insertion) and F1E1 (C-terminus insertion) for the prevention of Epstein-Barr virus (EBV) infections. X-ray crystallography confirmed the relative positions of the N-terminus insertion and C-terminus insertion. For N-terminus insertion, the epitopes were located on the exterior surface of ferritin, while for C-terminus insertion, the epitopes were inside the ferritin cage. Based on the results of antigen-specific antibody titers from *in-vivo* tests, we found that there was no obvious difference on humoral immune responses between N-terminus and C-terminus insertion. We also evaluated splenocyte proliferation and memory lymphocyte T cell differentiation. Both results suggested C-terminus insertion produced a stronger proliferative response and cell-mediated immune response than N-terminus insertion. C-terminus insertion of EBNA1 epitope was also processed more efficiently by dendritic cells (DCs) than N-terminus insertion. This provides new insight into the relationship between the insertion site and immunogenicity of ferritin nanoparticle vaccines.

KEYWORDS: Vaccine, ferritin nanocage, Epstein-Barr virus, Epstein-Barr nuclear antigen 1, Virus-like particles

6.1 Introduction

In the history of vaccine development, the first generation of vaccines used live-attenuated strains of a pathogen or inactivated killed pathogen, both of which have known limitations. The live-attenuated vaccine strains impose high risks of reversion into disease-causing viruses generating more virulent strains. The inactivated killed vaccine pathogens require multiple dosage administrations and adjuvants because of weak immune reactions [1]. There has been a rapid development of innovative vaccine delivery platforms, such as nanoparticle systems and proteins [2].

Nanoparticle proteins self-assemble from multiple protein subunits into hollow structures with desirable features that may overcome these limitations: high stability, biocompatibility and biodegradability [3]. Human ferritin heavy chain is an example of a protein nanoparticle. It is a single particle made up of 24 equivalent monomers, giving inner and outer diameters of 8 and 12 nm respectively [4]. As shown in Figure 6.1A, human ferritin heavy chain provides inner cavity by the hollow structure. Various properties of protein nanoparticles have been characterized, revealing very robust thermal and chemical stabilities [1], good ability to present antigens and expose immunogens [5-8], low toxicity, high biocompatibility and good biodegradability [9]. These properties present ferritin as a desirable nanoparticle vaccine platform. Each monomer of ferritin has 3 commonly used insertion sites; the N-terminus, C-terminus and a flexible loop region (Figure 6.1B). N-terminus insertion is considered as epitope insertion site in early research exploring ferritin as a vaccine delivery platform. For example, research conducted by Kanekiyo *et al.* [6] and Sliepen *et al.* [5] both selected the N-terminus as the insertion site of epitopes. Up to date, to see if other insertions would result a better dendritic cells (DCs) processing [10], some researchers started selecting the C-terminus or loop region for epitope

insertion [8,11-14]. There are still limited studies investigating the potential relationship between the insertion site and corresponding immunogenicity. In this study, we focus on investigation of the immunogenicity between N-terminus and C-terminus insertion and in particular the ability of ferritin to accommodate a selected epitope.

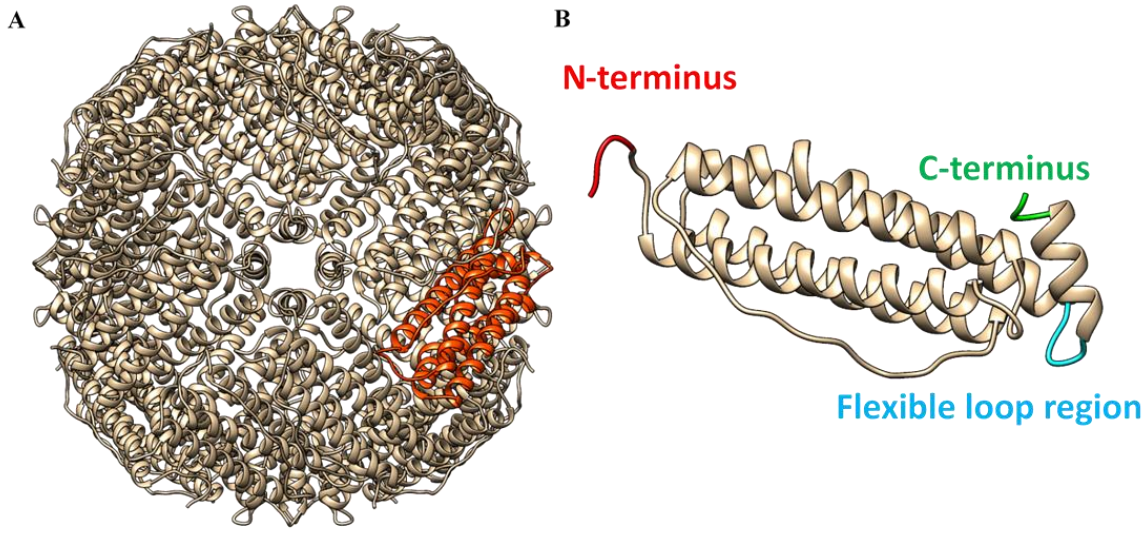


Figure 6.1 Human ferritin heavy chain mimic diagram by the program Chimera. A) The native ferritin assembly structure (PDB ID: 1FHA). Orange label represents ferritin monomer. B) Ferritin monomer. Red label represents N-terminus insertion site; Green label represents C-terminus insertion site; Blue represents flexible loop region insertion site.

Epstein-Barr nuclear antigen 1 (EBNA1) was used as the model for exploring an EBNA1-associated vaccine against Epstein-Barr virus (EBV) infection. EBV is one of the human herpesvirus types [15]. It infects B lymphocytes and some epithelial cells and causes various diseases including Burkitt's lymphoma, Hodgkin's disease, lymphomas and lymphoproliferative diseases [16]. More than 90 % of human adult populations are infected by EBV and approximately 1.5 % of all cancers in the world are associated with EBV infection [15]. The co-infection of EBV and other viruses also increases the risk of cancer. For example, EBV and human papillomavirus (HPV) are related to 38 % of all virus-associated cancers [17,18], such as cervical cancer, breast cancer, prostate cancer and lung cancer [17]. Unfortunately, therapy treatment for EBV infections is still under development with a lack of licensed vaccines for

EBV infection [15]. EBNA1 is expressed in all EBV-carrying tumours, which acts as a marker to differentiate the virus-associated cancer cells from normal cells. EBNA1 is essential for viral genome maintenance and controlling viral gene expression. As a result, virus is dependent on EBNA1 to persist [19]. As such, the development of an EBNA1 vaccine has the potential to prevent EBV infection and EBV related cancer.

In this study, we inserted an EBNA1 epitope (HPVGEADYFEY), which was found as the main immunogenic domain of EBNA1 [20], at the N-terminus or C-terminus of ferritin with 15 residues of soft linker (GGSGGGGSGGGGSGG) between the antigen and the ferritin. Our main purpose was to develop engineered ferritin vaccines with different insertion sites (N-terminus and C-terminus) to compare their potential immune response against EBV infection. To achieve this aim, engineered ferritin vaccines, EBNA1-ferritin (E1F1) (N-terminus insertion) and ferritin-EBNA1 (F1E1) (C-terminus insertion), were first successfully expressed and purified. Characterization experiments were also performed to confirm purity, size and structure of E1F1 and F1E1. To determine if both E1F1 and F1E1 could induce potent immune responses, mice were then injected with either engineered ferritin vaccine to determine antigen-specific IgG titers, proliferation indices, and lymphocyte T cells differentiation. These results will be used as a preliminary immunology study of E1F1 and F1E1 to test the potential on if they are able to induce potent immune responses after stimulating by the short EBNA1 peptide. Therefore, the successful outcomes will build confidence on further immunity experiments design including stimulating by the full length of EBNA1 antigen along with virus neutralization assay. Our study provides new understanding regarding the relationship between epitope insertion sites of ferritin and corresponding immunogenicity, which might provide

guidance for the future development of ferritin-based vaccines that illicit strong immune responses.

6.2 Materials and methods

6.2.1 Plasmid generation, expression and purification of recombinant ferritins

The pET 30a vector was used to construct plasmids harbouring genes of two types of EBNA1 fused ferritins, E1F1 and F1E1, inserted between the *Nde I* and *BamH I* restriction sites. The epitope was selected as EBNA1 (HPVGEADYFEY). A fifteen-residue linker (GGSGGGGSGGGGSGG) was inserted between the epitope and ferritin in both ferritin complexes. The plasmids were transformed into *E. coli* BL21 (DE3). The expression process of the two engineered ferritin constructs was the same. A single bacterial colony transformed with pET 30a containing E1F1 or F1E1 was inoculated in 50 mL LB medium supplemented with kanamycin (100 μg mL^{-1}), and incubated overnight at 37 °C in an orbital mixer incubator (Ratek), shaken at 180 rpm. A 10 mL of exponentially growing bacterial culture was then transferred to 500 mL of LB medium supplemented with kanamycin (100 μg mL^{-1}). The culture was grown at 37 °C and shaken at 200 rpm until OD_{600} reached 0.8, followed by induction with 1 mM IPTG at 37 °C for 4 h. Cells were collected by centrifugation at 13,751 xg. Cell pellets were resuspended in lysis buffer (20 mM PBS, 2 mM EDTA, pH 7.0) and lysed by mechanical disruption. The bacterial lysate was centrifuged at 19,802 xg for 30 min to remove cell debris and supernatant was harvested for protein purification.

The protein purification processes for E1F1 and F1E1 were the same, except they differed in the chromatographic resin used. First, clarified supernatant was diluted with Milli-Q water to a final protein concentration of 4 mg mL^{-1}. NaCl and NaAc-HAc were added to a final

concentration of 1 M and 100 mM respectively and the pH was adjusted to 5.0 to the diluted protein solutions and the mixtures underwent thermal precipitation at 60 °C for 10 min. Centrifugation (19,802 xg at 4°C for 10 min) was conducted to remove precipitated impurities. The resultant supernatant pH was adjusted to 6.5 by using 1 M phosphate buffer (PBS; Na_2HPO_4-NaH_2PO4), pH 8.0, and 1 M ammonium sulfate was added for purification by hydrophobic interaction chromatography (HIC). HIC was performed on an AKTA pure system (GE Healthcare, USA), with a HiTrap Octyl FF column (GE Healthcare, USA) used for E1F1, and a HiTrap Butyl FF column (GE Healthcare, USA) used for F1E1. The equilibration buffer was 100 mM PBS and 1 M ammonium sulfate, pH 6.5. The elution buffer was 20 mM PBS, pH 6.5. Briefly, samples were loaded onto columns equilibrated with equilibration buffer. After washing the column for 5 CV with equilibration buffer, a linear gradient elution of 0 -100 % elution buffer in 3 CV was conducted and the second eluted peak was collected. Absorbance at 280 and 260 nm were recorded. Finally, collected fractions from HIC were buffer-exchanged to 20 mM PBS, pH 7.0 using a HiTrap desalting column (GE Healthcare, USA). Protein expression and purity were analysed by SDS-PAGE.

6.2.2 Protein characterization

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Analysis of protein expression and purification was performed under 5 % (w/v) SDS spacer gel and 12 % (w/v) SDS separation gel. Samples were diluted with the same amount of 5× loading buffer (6 % (v/v) 1 M Tris-HCl pH 6.8, 25% (v/v) glycerol, 1 mg mL^{-1} bromophenol blue, 2 % (v/v) SDS and 5 % (v/v) β-mercaptoethanol). Treated samples were heated at 100 °C for 10 min and were separated by electrophoresis (Bio-Rad). The gel was stained with 0.25 % (w/v) Coomassie R-250 for 60 min and the background was washed using a washing buffer

(10 % (v/v) ethanol and 10 % (v/v) acetic acid) for 60 min.

Size exclusion chromatography coupled with multiple-angle laser light scattering (SEC-MALS)

SEC-MALS was performed to determine hydrodynamic radius of purified engineered ferritins, E1F1 and F1E1. Superose 6 10/300 GL (GE Healthcare, USA) was connected to High Performance Liquid Chromatography (HPLC) (Shimazdu, Japan) coupled with Wyatt Optilab refractive index (RI) and Wyatt DAWN MALLS detector (USA). Equilibration buffer was 20 mM PBS, pH 7.0. Protein concentration was 5 mg mL^{-1}. Loading volume was 50 µl. The output chromatogram was processed by Origin 2021 to calculate integrated area for each peak. Purity level was estimated using equation (6.1).

$$\text{Purity (\%)} = \frac{A_{\text{Target protein}}}{A_{\text{Total protein}}} \quad (6.1)$$

where $A_{\text{Target protein}}$ is the integrated area of target protein peak and $A_{\text{Total protein}}$ is the sum of integrated area of target protein peak and impure protein peaks.

Mass spectrometry

Purified protein samples were analysed using an Agilent 6230 TOF mass spectrometer (MS) equipped with Agilent 1260 LC system. Instrument was tuned at 3200 m/z mass range in positive ion mode. Using ESI source, 1 mL of the samples was injected into MS via the LC auto sampler and eluted without chromatographic separation at a flow rate of 0.5 mL min^{-1} using isocratic elution of 50 % buffer B, where buffer A is 0.1 % formic acid in water and buffer B is 0.1 % formic acid in acetonitrile. Mass spectra were acquired under the control of MassHunter Workstation software, LC-MS Data Acquisition (version B.08.00, Agilent technologies, USA).

Transmission electron microscopy (TEM)

E1F1 and F1E1 samples were diluted to a concentration of 0.1 mg mL^{-1}. Carbon-coated copper grids were incubated face down on 10 µL of sample for 10 min. The grid was washed 3 times in droplets of Milli-Q water, and then negatively stained in 2 % (v/v) uranyl acetate for 2 min. Excess liquid was removed with filter paper. Grids were analysed with a Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

Protein crystallization and data processing

For crystallization, previously purified E1F1 and F1E1 were dialysed to 10mM Tris pH 7.5 at concentrations of 5 mg mL^{-1} and 10 mg mL^{-1} respectively. Crystallization trials were completed using the PEGION, Index and Crystal Screen sparse matrix screens (Hampton Research) with 96 well Intelliplates (Art Robbins) in a sitting drop format at 16 ºC. Wells contained 80 µl of reservoir solution, with drops consisting of 1 µL ferritin construct and 1 µL of reservoir solution. Conditions for crystallization differed for the two constructs. For E1F1, crystals were obtained in a reservoir solution of 0.1 M ammonium citrate tribasic pH 7.0 with 12 % PEG 3350. F1E1 crystals were obtained in a reservoir solution of 0.1M HEPES pH 7.0 with 30 % Jeffamine M-600 pH 7.0. In both cases, crystals formed within 1 week.

Data collection, processing and refinement

Crystals of E1F1 and F1E1 were transferred to Paratone-N for cryoprotection and flash-frozen in liquid nitrogen. Diffraction images were collected at the MX1 beamline of the Australian Synchrotron [21]. Data were integrated and indexed using XDS [22], and scaled and merged using Aimless (CCP4) [23]. The structures of E1F1 and F1E1 were solved by molecular replacement with Phaser, using a previously solved structure of heavy chain ferritin (PDB: 6J4A) with solvent and ligands removed [24]. Solutions from Phaser underwent multiple rounds of refinement in Phenix, followed by multiple rounds of model building in Coot, until R-factors

converged [25,26]. Both structures possess good overall geometry, with >98 % of residues favoured in the Ramachandran plot and no outliers. All residues from aa. 5 – 176 were modelled, except for aa. 88 – 96, which were unable to be completely modelled for half of the chains of E1F1, and all chains of F1E1 due to ambiguous electron density. Detailed crystallographic statistics for E1F1 and F1E1 were recorded in Table 6.2.

Structure comparison and visualisation

Structural superposition was completed by sequence independent alignment in PyMol [27]. Figures for visualizing protein structure were generated using PyMol. Simulated annealing 2Fo-Fc composite omit maps were generated in Phenix.

6.2.3 Endotoxin concentration determination

Pierce^{TM} Chromogenic Endotoxin Quant Kit (Thermofisher, USA) was used to determinate endotoxin concentration of each protein sample (E1F1 and F1E1). High Standards (0.10, 0.25, 0.50 and 1.00 EU mL^{-1}) of *E. coli* endotoxin standard solution was prepared accordingly. Lyophilized amebocyte lysate and chromogenic substrate were reconstituted with endotoxin-free water (EFW) of 1.7 mL and 3.4 mL, respectively. Protein samples (E1F1 and F1E1) with protein concentration of 1 mg mL^{-1} in 20 mM PB buffer, pH 7.0, were diluted with EFW to protein concentration of 0.1 mg mL^{-1}.

The endotoxin-free 96-well plate was pre-equilibrated in a heating block at 37 °C. 50 µL of endotoxin standard dilutions, blank (EFW), and diluted protein samples (E1F1 and F1E1) were added per well. Each sample, blank and standards were run in triplicate. The plate was kept at 37 °C, and 50 µL of the reconstituted amebocyte lysate reagent was added per well. Once the amebocyte lysate reagent has been added to the plate well, the plate was removed from the

heater and mixed by tapping 10 times on the side of the plate. The plate was then incubated at 37 °C for 12 min. After exactly time 12 min, 100 µL of pre-warmed reconstituted chromogenic substrate solution was added per well. Once the substrate solution was added, the plate was removed from the heater and mixed gently by tapping 10 times to facilitate mixing. The plate was incubated at 37 °C for 6 min. After that, 50 µL of stop solution (25 % acetic acid) was added per well. Once the stop solution was added, the plate was removed from the plate heater and mixed by tapping 10 times on the side of the plate. The optical density (OD) at 405 nm was read via microplate reader (BioTek, USA). The average absorbance of the blank replicates was subtracted from the average absorbance of all individual standards and sample replicates to calculate mean absorbance. The standard curve (linear regression) was prepared by plotting the average blank-corrected absorbance for each standard on the y-axis against the corresponding endotoxin concentration in EU mL^{-1} on the x-axis. The coefficient of determination, r^2 was required to be ≥ 0.98. The endotoxin concentration of each protein sample was determined using the formulated standard curve. The final endotoxin concentration of each protein sample was after multiplying by 10 times dilution.

6.2.4 Mice immunisation

All animal experiments were approved by the Medical Ethics Committee of ShanXi University of Chinese Medicine (Approval Number 2019LL137).

Three groups (n=6/treatment) of 6 to 8 week-old female BALB/c mice were immunised intraperitoneally three times (Day 0, 14, 28) with 6 µg of EBNA1 short peptide epitope without adjuvant (Group 1: control), purified E1F1 (1 mg mL^{-1}) and F1E1 (1 mg mL^{-1}) in 200 µL PBS (Group 2 and 3: samples) (equivalent mass amount of EBNA1 epitope) mixed with an equal volume amount of endotoxin free aluminum hydroxide (0.8 mg mL^{-1}). Two extra groups

(n=6/treatment) of mice were treated with 200 µL PBS (negative control) and 100 µg of Ovalbumin (OVA), respectively, used for check if the system and animal models were working. The serum was collected at 14 and 28 days after first time immunization and stored at -80 °C.

6.2.5 Enzyme-linked immunosorbent assay (ELISA)

EBNA1 epitope (HPVGEADYFEY) was adsorbed overnight at 4 °C in 96-well plates at 10 µg mL^{-1} in 50 mM sodium carbonate buffer pH 9.6. After being blocked with PBS containing 1 % bovine serum albumin (BSA) for 2 h at 37 °C, the plate was incubated with serial dilutions of mouse sera, collected at 14, 28 days after the first immunization from each group, for 2 h at 37 °C. The plate was washed three times with PBS containing 0.1 % Tween-20. A goat anti-mouse IgG horseradish peroxidase-conjugated antibody was applied at a 1:5000 dilution, and 200 µL were added to the plate. After being incubated at 37 °C for 2h, the plates were washed, and 3,3′,5,5′-Tetramethylbenzidine (TMB) Single-Component Substrate solution was added for colour development. After 10 min, 1 M sulphuric acid was added to stop the colour development. After 10 min, the plates were analysed using a microplate reader (Perlong, China) at wavelength 450 nm. Endpoint titers were defined as the highest serum dilution that resulted in an absorbance value two times greater than that of negative control sera derived from the non-immunized mice.

6.2.6 Antigen-specific IgG isotype

The procedure for antigen-specific IgG isotype followed ELISA method as outlined in section 2.5. EBNA1 short peptide (HPVGEADYFEY) was absorbed overnight at 4 °C in 96-well plates at 10 µg mL^{-1} in 50 mM sodium carbonate buffer at pH 9.6. After being blocked with phosphate-buffered saline (PBS) containing 1 % bovine serum albumin (BSA) for 2 h at 37 °C, the plate was incubated with 100 µL diluted mouse sera (same dilution of end point) prepared

before for ELISA, for 2 h at 37 °C. The plate was washed three times with PBS containing 0.1 % Tween-20. Antibodies goat anti-mouse IgG1 and IgG2a were applied at 1:1000 dilution, and 100 μL were added to the plate. A rabbit anti-goat IgG horseradish peroxidase-conjugated antibody was applied at a 1:5000 dilution, and 200 μL were added to the plate. The remained procedure was exactly same as section 6.2.5.

6.2.7 T-cell proliferation assay

Murine splenocytes were harvested using red blood cell lysing buffer (Sigma). Single-cell suspensions were prepared (4 x 10^6 cells mL^{-1}) and cultured in RPMI 1640 (Gibco, Germany) with a short EBNA1 epitope at 20 μg mL^{-1}. The plates were then incubated for 48 h before the addition of 10 μL of CCK-8 (Dojindo, Japan) per well. The absorbance at 450 nm was recorded in a microplate reader (Perlong, China) after 4 h incubation at 37 °C. The results were presented as proliferation indexes (PI) using the following equation (6.2) and equation (6.3).

$$Proliferation\ rate = \frac{A(sample) - A_0(control)}{A_0(control)} \quad (6.2)$$

A(sample): A_{450} nm of the sample group with CCK-8 and stimulator (EBNA1 epitope, E1F1 or F1E1)
A_0(control): A_{450} nm of the control group with CCK-8 only

$$Proliferation\ Index = \frac{Proliferation\ rate\ (sample)}{Proliferation\ rate\ (PBS)} \quad (6.3)$$

6.2.8 Lymphocyte activation

Splenocytes from immunized mice were cultured for 60 h at 37°C with and without EBNA1 epitope before staining with FITC CD4 Monoclonal antibody, PerCP-Cy 5.5 CD8a Monoclonal antibody, APC-eFluor 780 CD19 Monoclonal antibody and APC CD25 Monoclonal antibody (Thermo fisher, USA) at 4 °C. After washing with PBS, fluorescence patterns were analysed using a FACSCanto II (BD Bioscience, USA). The percentage of CD25/CD4 T cells, CD25/CD8 T cells and CD25/CD19 B cells were recorded and analysed using FACSCanto II

software (BD Bioscience, USA).

The procedure to detect memory lymphocyte T cells differentiation was same as above. After the gate for CD25 was circled, gates for CD44 and CD62L were circled in $CD4^+$ T and $CD8^+$ T cells, respectively.

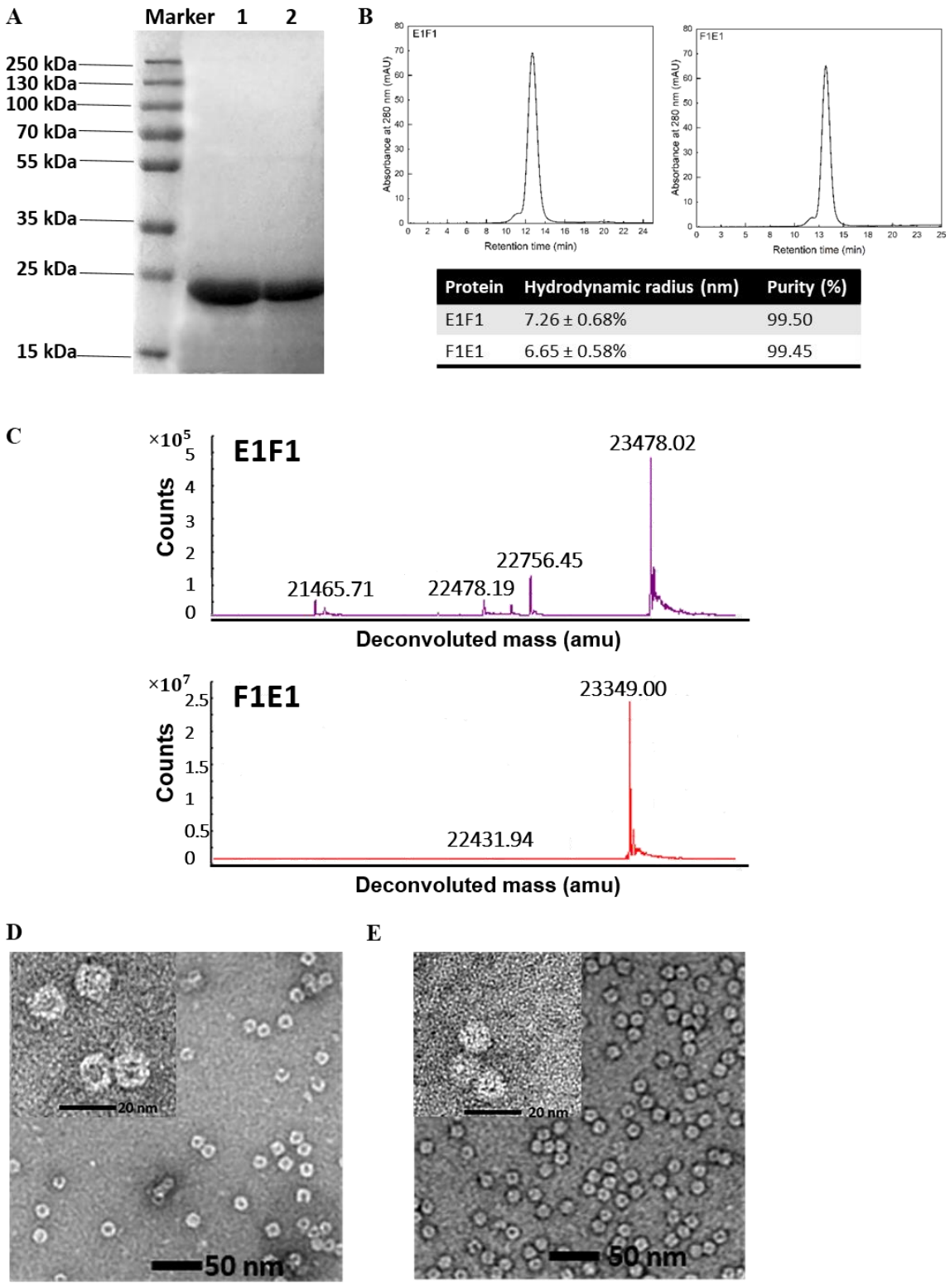
6.2.9 Statistical analysis

Data generated from *in vivo* immunology studies were presented in Mean ± Standard deviation (SD). GraphPad Prism 9 was applied to determine statistical significance of results in antigen-specific antibodies titers, T-cell proliferation assay and lymphocyte activation following One-way ANOVA method. P-value < 0.05 (*) was considered as significant. P-value < 0.01 (**) was moderately significant. P-value < 0.001 (***) was highly significant.

6.3 Results

6.3.1 Characterization of engineered ferritins

The production procedures (including expression and purification) for engineered ferritins were general by applying advantages of native ferritins, such as robust thermal and chemical stabilities. Characterization of engineered ferritins are significant to check if they meet quality criteria in terms of satisfied purity and well-assembled structures, before performing animal experiments. It also helps reveal the potential structural differences between E1F1 and F1E1, in order to build connections with further immunology results. The molecular weight for monomer of E1F1 and F1E1 is 23.35 kDa based on amino acid sequence estimation by ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss). E1F1 and F1E1 showed the same molecular weight by SDS-PAGE as expected (Figure 6.2A). From the mass spectrum, the molecular weight for the monomers of E1F1 and F1E1 were 23.48 and 23.35 kDa, respectively (Figure 6.2C), matching the expected size of the engineered ferritins with the EBNA1 epitope intact. The slight difference in mass of E1F1 and F1E1 seems like matching the start methionine (M). It may suggest that, for F1E1, M was removed in *E. coli* because of similarity with bacterial ferritin. The hydrodynamic radius measured by SEC-MALS indicated a particle size difference between E1F1 and F1E1, with sizes 7.26 and 6.65 nm, respectively (Figure 6.2B). These sizes were consistent with previous measured hydrodynamic radii of ferritin nanoparticles (6.7 nm) [28]. Both SDS-PAGE and SEC results confirmed that purity of E1F1 and F1E1, which were 99.50 % and 99.45 %, respectively (Figure 6.2A, Figure 6.2B). It indicates that less than 0.6 % impurity is process-related proteins (for example, host cell protein (HCP) of *E. coli*) and product-related proteins (for example, dissociated or denature HFn). The high purity of protein sample for vaccine performance evaluation is critical.



Protein	Hydrodynamic radius (nm)	Purity (%)
E1F1	7.26 ± 0.68%	99.50
F1E1	6.65 ± 0.58%	99.45

Figure 6.2 Characterization of E1F1 and F1E1. A) SDS PAGE analysis post purification: Marker: protein marker; lane 1: E1F1; lane 2: F1E1. B) Hydrodynamic radius measurements and purity level estimation from size exclusion chromatography with multiple-angle laser light scattering (SEC-MALS). C) Mass spectra of E1F1 and F1E1. The molecular weights for the E1F1 and F1E1 monomers were determined as 23.48 and 23.35 kDa, respectively. D) TEM image of E1F1. E) TEM image of F1E1.

The endotoxin determination experiment was conducted to ensure the immunology result and vaccine performance validation is accurate. The endotoxin standard curve was shown in Figure S6.1 (Supporting Information). The equation was valid with R-squared greater than 0.98. Shown by Table 6.1, the endotoxin concentration in the injection protein sample for E1F1 and F1E1 were 9.67 and 9.56 EU per mg of protein, respectively. They are much less than the endotoxin limit of 152 EU mg^{-1} protein for preclinical research animal models. The detail of endotoxin limits for preclinical research animal models is calculated and shown in Table S6.1, which is based on the threshold pyrogenic human dose of 5 EU kg^{-1} [29].

Table 6.1 Endotoxin concentration of protein samples.

Protein sample name	E1F1	F1E1
Average OD (405 nm)	1.35	1.33
Endotoxin concentration by 10x dilution (EU mL^{-1})	0.967	0.956
Actual endotoxin concentration (EU mL^{-1})	9.67	9.56
Total protein injection (mg)	0.2	0.2
Total Endotoxin amount per dose (EU mg^{-1} protein)	9.67	9.56

Purified proteins E1F1 and F1E1 both presented well-assembled structures based on TEM images (Figure 6.2D, Figure 6.2E). To verify the location of epitopes for different insertion site, the structures of E1F1 and F1E1 were solved by x-ray crystallography to 2.8 and 2.2 Å resolution, respectively (Table 6.2). Both engineered ferritin constructs displayed the same nanoparticle assembly and monomer structure compared to native heavy chain ferritin (Figure 6.3A). No electron density was observed for the EBNA1 epitope and Gly-Ser linker in either structure (Figure 6.3B). As the structural integrity was verified by mass spectrometry, this is likely due to the flexibility of the insert. As for native heavy chain ferritin, the N-terminus and C-terminus of each ferritin monomer projected towards the exterior and interior of the

nanoparticle, respectively (Figure 6.3C). This supports E1F1 presenting the EBNA1 epitope on the surface of the nanoparticle, while the epitope is contained within the ferritin cage for F1E1 and is consistent with the larger hydrodynamic radius measured for the E1F1 construct.

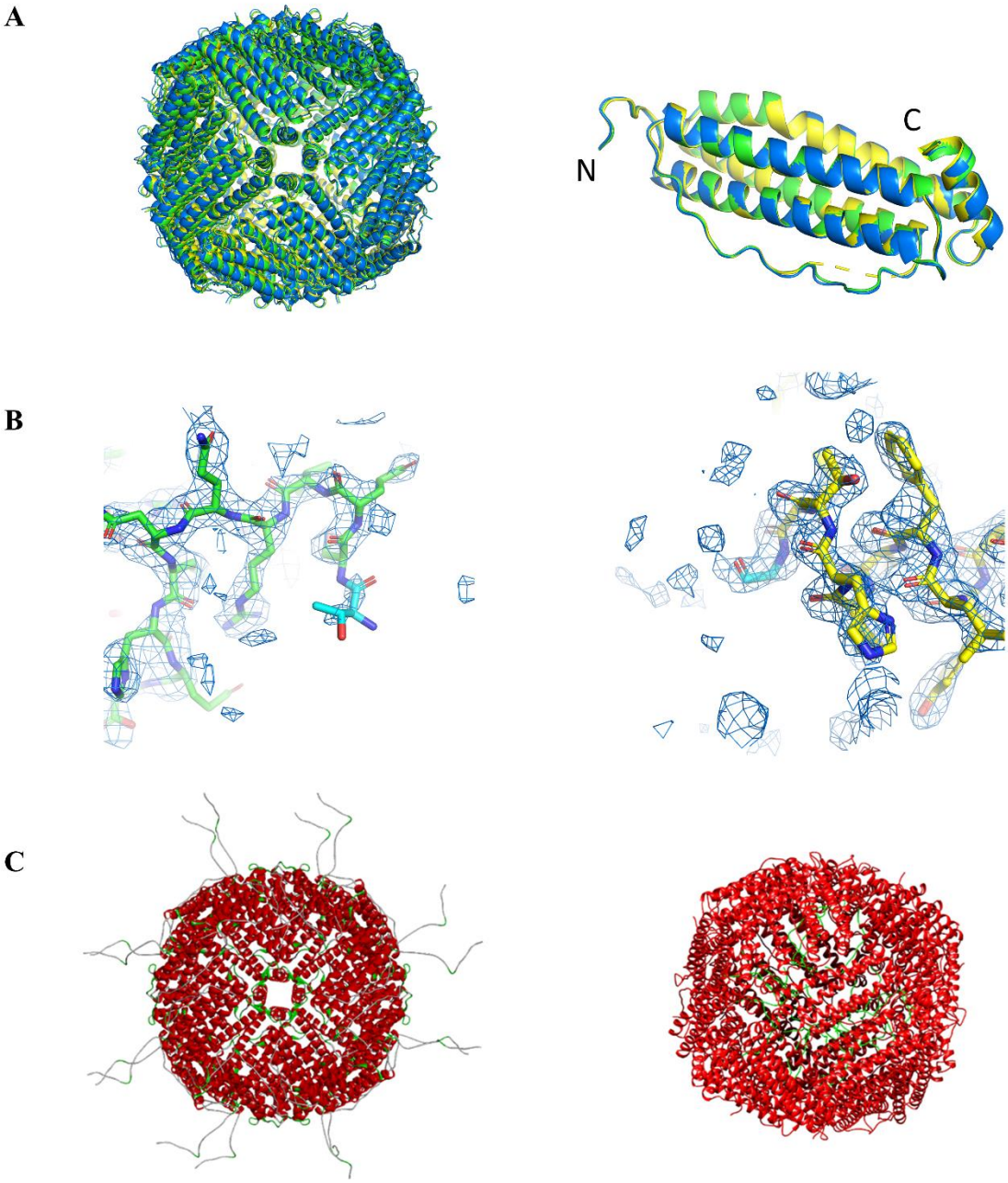


Figure 6.3 Structural determination of E1F1 and F1E1. A) Structural superposition of the nanoparticle assembly (left) and a single ferritin monomer (right) for E1F1 (green; RMSD = 0.33) and F1E1 (yellow; RMSD = 0.40) with native heavy chain ferritin (blue, PDB:1FHA). B) Simulated annealing 2Fo-Fc composite omit maps (blue mesh, 1.5σ) for the N-terminus of E1F1 (green, Chain A; left) and C-terminus of F1E1 (yellow, chain A; right). No electron density was observed for the EBNA1 epitope for either construct. Terminal residues are shown in cyan. The presented model is representative of each monomer forming the E1F1 and F1E1 nanoparticles. C) E1F1 (left) and F1E1 (right) modelled with the EBNA1 epitope.

Table 6.2 Crystallographic statics for structures of E1F1 and F1E1.

	E1F1 (PDB: 7KE5)	**F1E1 (PDB: 7KE3)**
Wavelength (Å)	0.9537	0.9537
Resolution range (Å)	48.3 - 2.8 (2.9 - 2.8) [a]	48.72 - 2.2 (2.279 - 2.2) [a]
Space group	P 4_2 2_1 2	P 4_2 2_1 2
Unit cell		
a, b, c (Å)	219.323 219.323 148.556	217.882 217.882 147.203
α, β, γ (°)	90 90 90	90 90 90
Total reflections	1766537 (126783)	4871002 (477929)
Unique reflections	85428 (8054)	177897 (17576)
Multiplicity	20.7 (15.7)	27.4 (27.2)
Completeness (%)	91.01 (89.44)	99.02 (98.17)
Mean I/sigma(I)	7.48 (2.53)	8.86 (1.94)
Wilson B-factor	28.80	24.55
R-merge	0.2514 (0.8948)	0.3341 (2.091)
R-meas	0.2579 (0.9243)	0.3404 (2.131)
R-pim	0.05599 (0.2261)	0.06503 (0.4079)
CC1/2	0.998 (0.975)	0.999 (0.911)
Reflections used in refinement	81188 (7863)	176417 (17273)
Reflections used for R-free	4127 (396)	8842 (873)
R-work	0.2726 (0.3180)	0.2408 (0.3245)
R-free	0.3141 (0.3808)	0.2709 (0.3522)
Number of non-hydrogen atoms	16844	17042
macromolecules	16745	16263
ligands	64	68
solvent	35	711
Protein residues	2064	2001
RMS(bonds)	0.009	0.007
RMS(angles)	0.99	0.85
Ramachandran favoured (%)	98.58	98.72
Ramachandran allowed (%)	1.42	1.28
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	2.02	1.22
Clashscore	3.81	2.50
Average B-factor	27.79	26.48
macromolecules	27.80	26.23
ligands	29.31	33.32
solvent	21.05	31.71

[a] Statistics for the highest resolution shell are shown in parentheses.

6.3.2 Evaluation of antigen-specific antibodies titers

To confirm the immunogenicity of purified engineered ferritin constructs, an *in vivo* model was applied. The measurement of antigen-specific IgG titers is an indicator of immunogenicity induced by a vaccine. Compared to the EBNA1 epitope, both E1F1 and F1E1 showed stronger humoral immunogenicity after the second and third immunizations indicated by greater IgG titers (Figure 6.4A). The immune response was improved significantly after the third immunization. There was no obvious difference between E1F1 and F1E1 regarding humoral immunogenicity. To identify the immunization pathway towards humoral immune response or cell-mediated immune response, the ratio between IgG2a and IgG1 was measured [30]. If IgG2a/IgG1 is greater than 1, the immune response tends to be Th1 biased, which is related to a cell-mediated response. If this ratio is less than 1, the immune response tends to be Th2 biased, indicating humoral immune response [31]. Th1 mainly secretes cytokines including IL-12, TNF-α and IFN-γ, which activate a cytotoxic T lymphocyte (CTL) response that is necessary for the control of a variety of bacterial and viral infections [32]. Th2 mainly secretes cytokines including IL-4, IL-5 and IL-6 to activate a humoral immune response and promote the antibody-dependent cell-mediated cytotoxicity (ADCC) system [33]. As shown in Figure 6.4B, the IgG2a/IgG1 ratio induced by EBNA1 short epitope without adjuvant was slightly greater than 1, suggesting similar level of humoral and cell-mediated immune response strength. However, for both E1F1 and F1E1, which were administered with aluminum hydroxide adjuvant, the ratio of IgG2a/IgG1 was <1, suggesting a humoral immune response was dominant for this case.

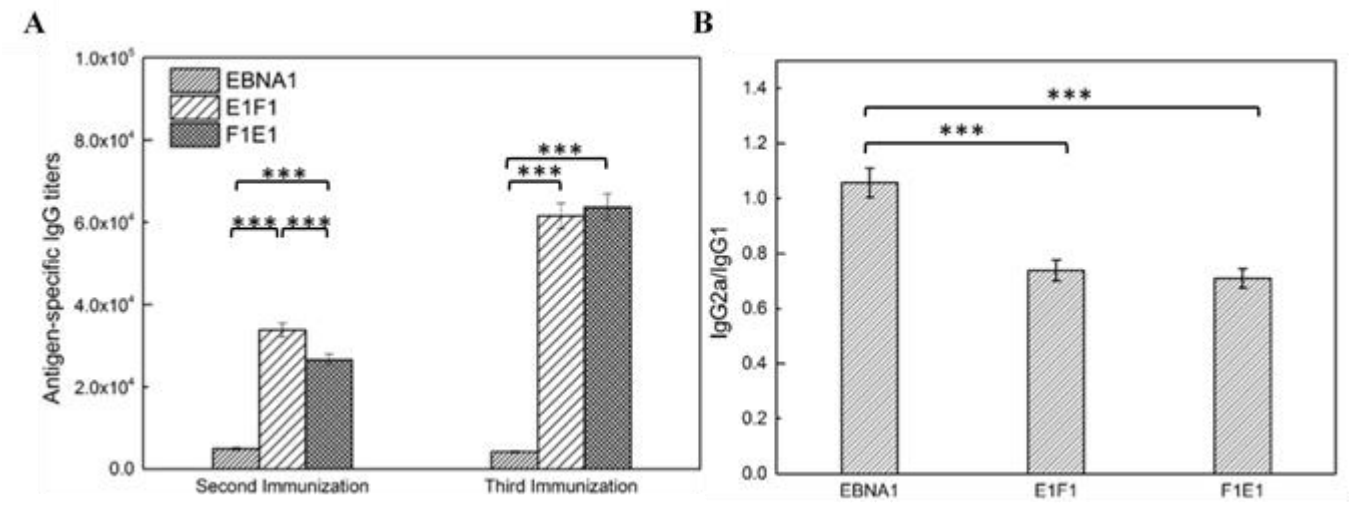


Figure 6.4 Production of antigen-specific antibodies in the sera of BALB/c mice. The dosage of groups was 6 μg EBNA1. (A) The antigen-specific IgG titers at day 14 (second immunization) and day 28 (third immunization). (B) The ratio of IgG2a/IgG1 at day 28. Data are expressed as mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001), where n represents the number of mice from each group.

6.3.3 T-cell proliferation assay

There are specific receptors on the surface of B cells and T cells that recognize antigens. After being stimulated by specific antigen, in this case EBNA1 epitope , the lymphocytes perform mitosis [34]. The level of splenocyte proliferative responses gives an indication of the strength of the immune response induced by a vaccine, which is measured as proliferative index (PI) [35]. The PI for EBNA1 epitope, E1F1 and F1E1 were 1.2, 1.5 and 2.3, respectively (Figure 6.5). There was a significant difference in PI depending on the insertion site of EBNA1 epitope, with F1E1 producing a greater PI than E1F1 (***p<0.001), indicative of a stronger immune response for C-terminus insertion.

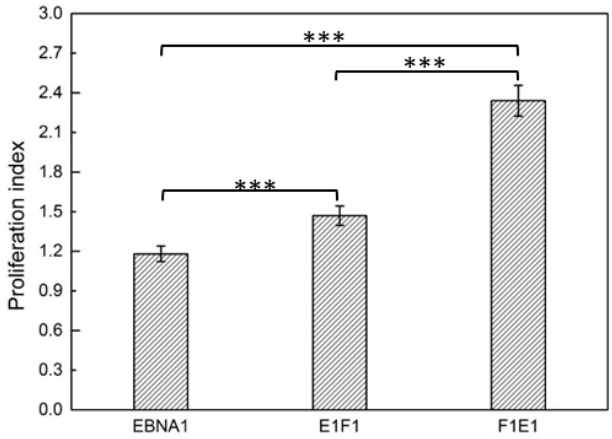


Figure 6.5 Proliferative responses of splenocytes responding to antigen simulation *ex vivo*. Splenocytes were harvested 10 d after the third immunization and restimulated *ex vivo* with antigen EBNA1 (20 μg mL^{-1}). Splenocytes proliferation was measured using CCK-8 kit, and the proliferation index was calculated. (EBNA1: 1.18; E1F1: 1.47; F1E1: 2.34) Data are expressed as mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001), where n represents number of mice from each group.

6.3.4 Lymphocyte activation

CD25 is regarded as an activation marker for B cells and T cells [36], which represents the activation level for B cells and T cells after stimulating with antigen, such as the EBNA1 epitope. $CD4^+$ T cells are T helper cells that play an important role in instigating adaptive immune responses [37]. $CD8^+$ T cells, like $CD4^+$ helper T cells, are crucial for immune defense against intracellular pathogens [38]. After recognition by the antigen and activation, $CD8^+$ T cells will secrete cytokines, mainly TNF-α and IFN-γ, which are related to the cell-mediated immune response. CD19 is a biomarker for B cells, which is important to maintain the balance between humoral and antigen-induced response [39]. By measuring the relative number of $CD4^+$ T cells, $CD8^+$ T cells and $CD19^+$ B cells, the strength of humoral immune response and cell-mediated immune response could be compared. As shown in Figure 6.6, after activation, cellular levels for $CD4^+$ T cells, $CD8^+$ T cells and $CD19^+$ B cells induced by E1F1 and F1E1 were significantly increased compared to before activation (***p<0.001). EBNA1 short peptide vaccine showed negligible activation level (Figure 6.6A and Figure 6.6C). There were approximately 2 times

$CD4^+$ T cells (Figure 6.6A) and 6 times $CD8^+$ T cells (Figure 6.6B) more found in engineered ferritin groups than EBNA1 short peptide, which means activation by EBNA1 was more effective for engineered ferritins with adjuvant groups than short peptide without adjuvant. There were no obvious differences found between E1F1 and F1E1 for $CD4^+$ T cells and $CD19^+$ B cells after activation (Figure 6.6A and Figure 6.6C). Amount of $CD8^+$ T cells after activation for F1E1 was obviously 11 % greater than E1F1 (**p<0.01) (Figure 6.6B). This result further supports that C-terminus insertion induced stronger cell-mediated immune responses than N-terminus insertion.

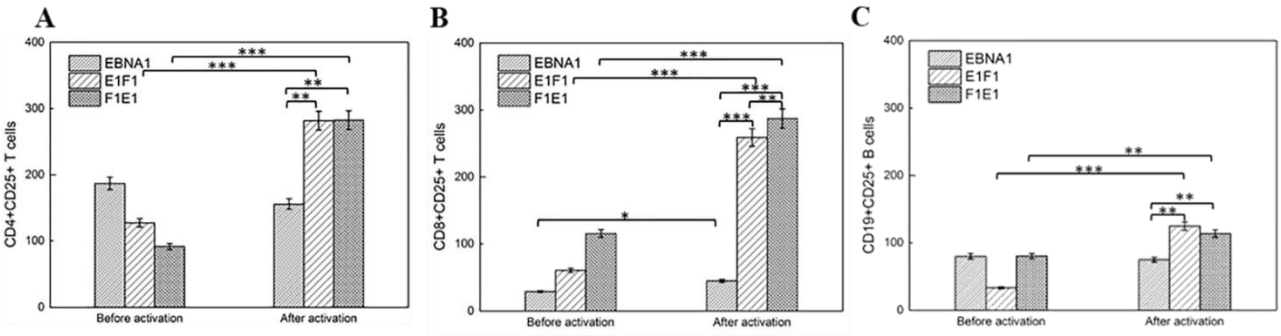


Figure 6.6 Frequencies of CD25+ in CD4+, CD8+ T cells and CD19+ B cells were determined by flow cytometry. T cells amount were count for each group before and after activation by short EBNA1 peptide (500 µL, 20 µg mL^{-1}). Data are expressed as mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001), where n represents the number of mice from each group. A) Frequencies of $CD25^+$ in $CD4^+$. B) Frequencies of $CD25^+$ in $CD8^+$ T cells. C) Frequencies of $CD25^+$ in $CD19^+$ B cells.

6.3.5 Memory lymphocyte T cells differentiation

One of the purposes of vaccination is to acquire immunologic memory by maintaining memory T cells ($CD4^+$ and $CD8^+$) to introduce fast and effective actions on the antigen-specific immune response [40]. Memory T cells are divided into effector T cells (T_{EM}) and central memory T cells (T_{CM}) [41]. The differences between T_{EM} and T_{CM} are: the start-up for T_{EM} is faster compared to T_{CM} but with shorter lifetime. The process that T_{EM} and T_{CM} can be transferred to each other is

activated by the stimulator (short EBNA1 peptide). For a strong immune response, cells amount for T_{CM} before the stimulation and T_{EM} after the stimulation, are expected to be significantly high. The expression of CD44 and CD62L population indicates if memory T cells are T_{CM} or T_{EM} [42]. T_{CM} highly express both CD44 and CD62L ($CD44^{hi}$ $CD62L^{hi}$), while T_{EM} are restricted to highly express CD44 only ($CD44^{hi}$ $CD62L^{low}$) [43]. It was observed that before stimulation by EBNA1 epitope, the number of $CD4^+$ and $CD8^+$ T_{CM} for mice immunized with F1E1 was >1000 cells, 2 times greater than immunization with both EBNA1 epitope (~300 cells) and E1F1 (~ 500 cells) (Figure 6.7B and Figure 6.7C). After stimulation by EBNA1 epitope, there was a decreasing trend of T_{CM} found for both two groups, because $CD4^+$ and $CD8^+$ T_{CM} were converted to T_{EM}. Overall, the number of T_{EM} for E1F1 and F1E1 was 2 times greater than for EBNA1 epitope (Figure 6.7D and Figure 6.7E). While E1F1 and F1E1 caused similar $CD4^+$ T_{EM} levels after stimulation (Figure 6.7D), F1E1 induced 9 % greater number of $CD8^+$ effector memory T cells than E1F1 (*$p<0.05$) (Figure 6.7E). Hence, C-terminus insertion produces a stronger cell-mediated immune response than N-terminus insertion, because greater number of T_{CM} before stimulation and T_{EM} after the stimulation were found in F1E1 than E1F1.

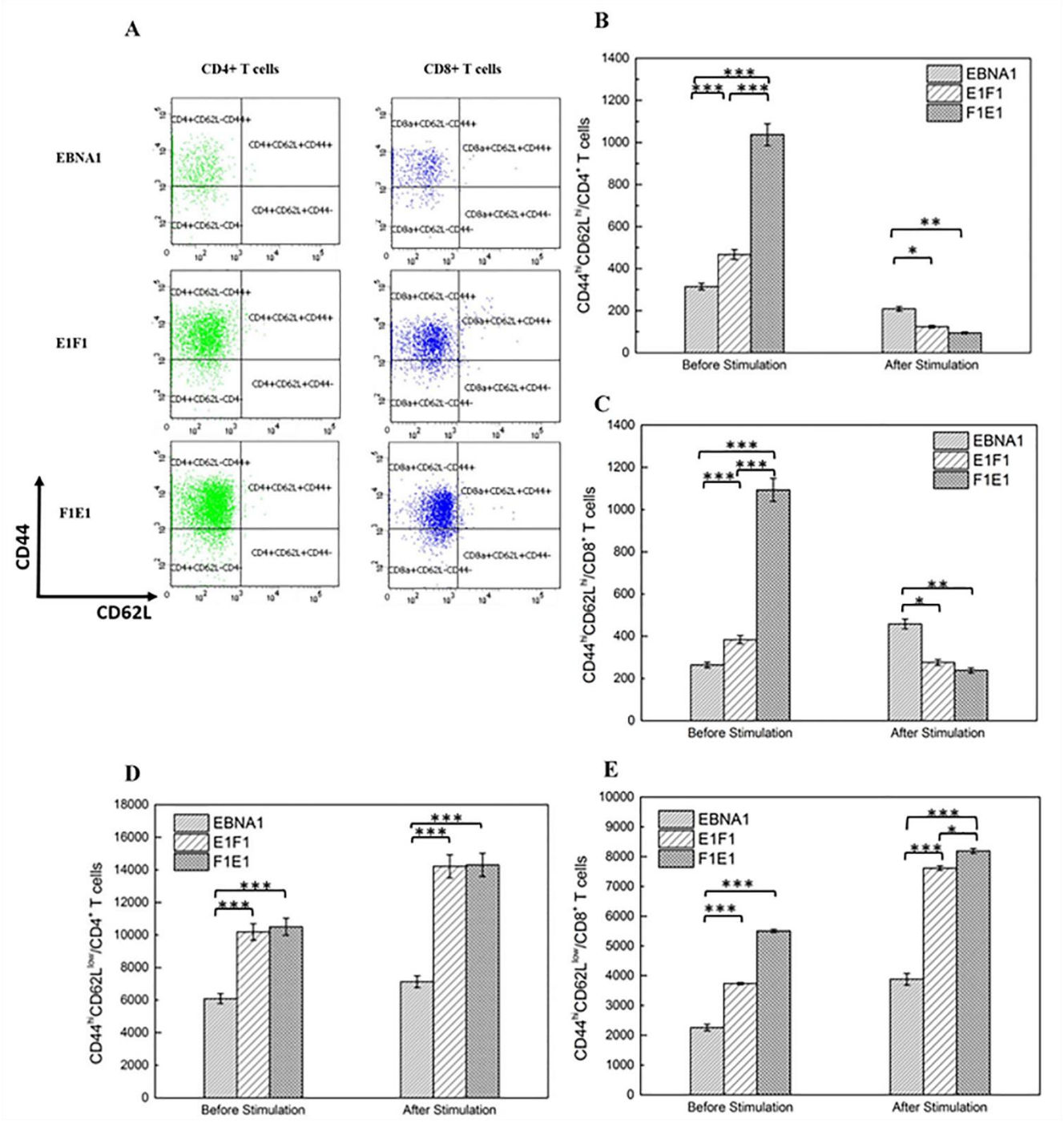


Figure 6.7 Frequencies of central (CD44hiCD62Lhi)/ effector (CD44hiCD62Llow) memory CD4+ and CD8+ T cells. Mice were immunized three times. Splenocytes were harvested on day 38 after the first immunization and stimulated *ex vivo* with antigen EBNA1 peptide (500 µL, 20 µg mL^{-1}) for 60 hours. The frequency of $CD44^{hi}$ $CD62L^{hi}/CD4^{+}$ T cells, $CD44^{hi}$ $CD62L^{low}/CD4^{+}$ T cells, $CD44^{hi}$ $CD62L^{hi}/CD8^{+}$ T cells, $CD44^{hi}$ $CD62L^{low}/CD8^{+}$ T cells were measured by flow cytometry. FACS plots in A) are representative of the mean percentages of six mice in each group. Data are expressed as mean ± SEM (n=6) (*$p<0.05$; **$p<0.01$; ***$p<0.001$), where n represents the number of mice from each group. A) FACS plots. B) Frequencies of central ($CD44^{hi}CD62L^{hi}$) memory $CD4^{+}$ T cells. C) Frequencies of central ($CD44^{hi}CD62L^{hi}$) memory $CD8^{+}$ T cells. D) Frequencies of effector ($CD44^{hi}CD62L^{low}$) memory $CD4^{+}$ T cells. E) Frequencies of effector ($CD44^{hi}CD62L^{low}$) memory $CD8^{+}$ T cells.

6.4 Discussion

This study presents the development of engineered ferritin vaccines against EBV, cooperation with aluminum hydroxide adjuvant. Compared to the EBNA1 short peptide without adjuvant, both N-terminus insertion and C-terminus insertion induced potent humoral and cell-mediated immune responses. The proliferative response was enhanced after applying ferritin as the vaccine carrier. This all confirmed that human ferritin heavy chain is eligible to be treated as a vaccine carrier. This is possibly caused by the molecular weight difference and addition of adjuvant. Ferritin has uniform architecture with 24 subunits of total molecular weight around 504 kDa. The molecular weight for the short EBNA1 epitope is 1.3 kDa. Hence, the ferritin vaccine platform presenting antigens provides a larger structure than for the EBNA1 epitope structure, explaining the greater induction of humoral and cell-mediated immune responses [44]. This also shows the limitations of peptide vaccine. Peptide vaccine induces weak immune responses and is found to induce fairy weak immune responses even with adjuvant [45-47]. Therefore, peptide vaccine require carriers to counterbalance the low-molecular nature and low efficiency.

The dominant immune response for the EBNA1 epitope was found to be different to the engineered ferritins groups. The ratio of IgG2a/IgG1 for the short peptide group was slightly greater than 1. For engineered ferritins with adjuvant, the ratio is lowered and the dominant immune response changed from cell-mediated to humoral immune response. The reason may be that, for engineered ferritin groups, aluminum hydroxide adjuvant was added to increase the vaccine efficacy and enhance the immunity [48]. Aluminum adjuvants induce Th2 responses only, which mostly activate humoral immune response, and they are unable to promote cytotoxic T cell-mediated immunity [49-51]. EBNA1 epitope without addition of aluminum adjuvants, can

elicit Th1-type $CD4^+$ and $CD8^+$ T cells [52,53], which are mostly related to cell-mediated immune response in principle. As a result, cell-mediated immune response is dominant for EBNA1 peptide group, while humoral immune response is dominant for engineered ferritins groups.

In this study, we inserted the antigen at either N-terminus or C-terminus of ferritin to produce E1F1 and F1E1 vaccines, respectively. Structural characterization of the engineered ferritins revealed that for N-terminus insertion EBNA1 epitope was presented on the exterior surface of the ferritin, while C-terminus insertion placed the epitopes into the interior cavity of ferritin. It was found that there was no obvious difference in humoral immune responses for N-terminus insertion and C-terminus insertion. From previous research by Wang *et al*. [8], it was more likely that when epitopes were inserted outside the ferritin cage (N-terminus insertion and the flexible loop region), it was easier to induce greater antibody titers than inserting antigens inside the ferritin cage (C-terminus insertion). Nonetheless, this finding does not really accord with our findings here. We found that C-terminus insertion induces a comparable antibody titer to N-terminus insertion. It may be that many factors result in different behaviours, including different adjuvant types used and different epitopes applied. Hence, our findings suggest that presenting epitopes outside or inside the nanoparticles will not directly determine a stronger or weaker humoral immune response.

Desirable vaccines also achieve antigen-specific cellular response in addition to humoral response. Both N-terminus and C-terminus insertion produced antigen-specific cellular response after stimulation with EBNA1 epitope. Dendritic cells (DCs), which are the most important antigen-presenting cells that initiate and direct adaptive immune responses [54], orderly

phagocytosed these engineered ferritin nanoparticles carrying EBNA1 epitope [10]. Following that, the processed EBNA1 epitopes within endosomes were successfully displayed on the surface of DCs with the major histocompatibility complexes (MHCs) MHCI and MHC II, and hence induced EBNA1 epitope specific $CD4^+$ and $CD8^+$ T cells via their T cell receptors (TCRs), shown in Figure 6.8. C-terminus insertion showed stronger proliferative responses than N-terminus insertion. Meanwhile, from the evaluation of memory T cells, C-terminus insertion presented a high number of central memory T cells (before stimulation) and effector memory T cells (after stimulation). These results suggest that inserting epitopes at the C-terminus of ferritins may be processed more efficiently by DCs than N-terminus insertions. It is inferred that the positions of antigen and protein conformations result in different antigens-processing pathways by DCs. The possible reason may be the structural differences between E1F1 and F1E1. For N-terminus insertion, the epitopes and 15 units of soft linkers were fully presented outside the ferritin cage. The role of soft linkers ensures the flexibility of epitopes [55] in order to avoid unfavourable interactions between the epitopes and ferritin. On the other hand, soft linkers are not rigid structures and can be easily broken away from the protein nanocage because of changes in the surrounding environment [56]. Antigens may thus be released from ferritin. Compared to N-terminus insertion, epitopes and linkers are protected inside the ferritin cage for C-terminus insertion. On this account, the effective antigen density for C-terminus insertion would be greater than N-terminus insertion due to the protection afforded by the cage. There would be more effective antigen processed by DCs for C-terminus insertion than N-terminus insertion.

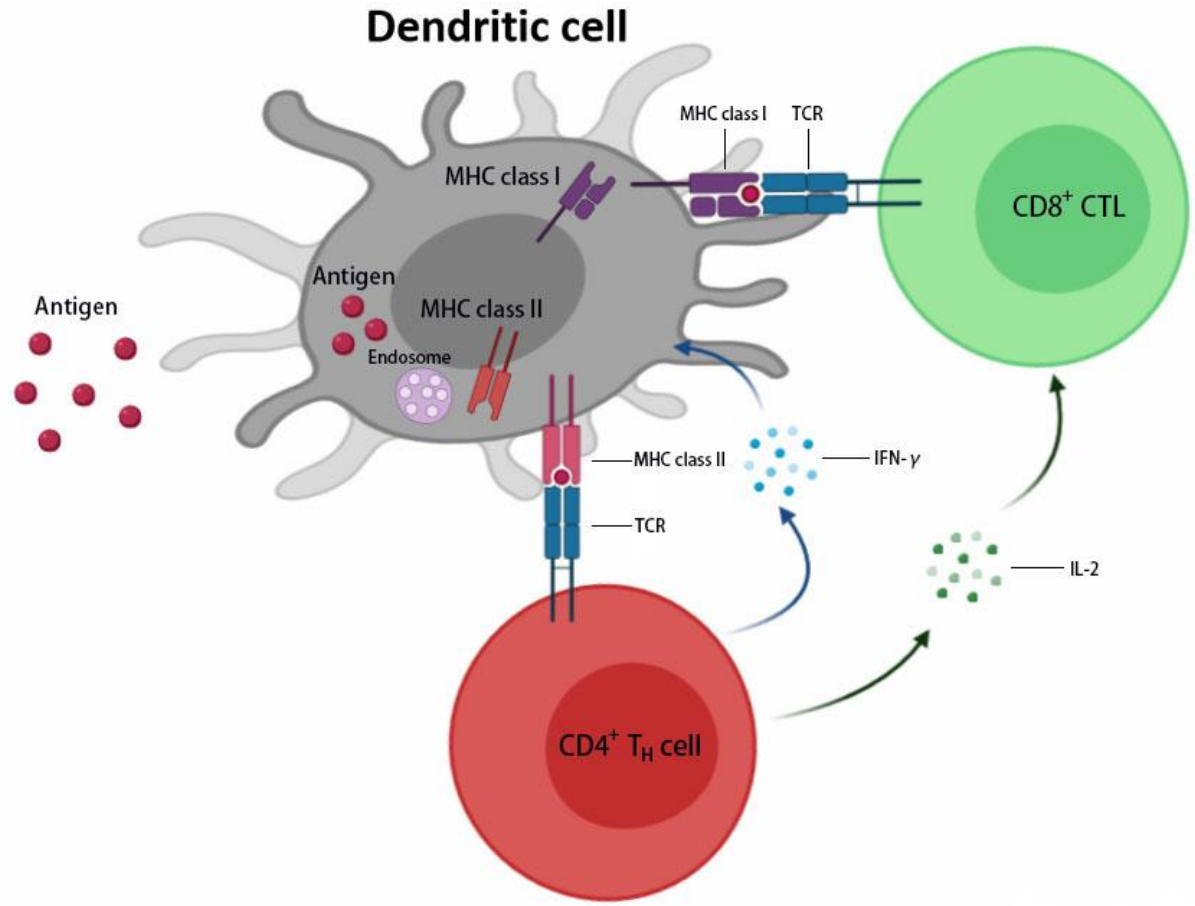


Figure 6.8 Dendritic cell processing diagram by Biorender. Antigen presentation is mediated by MHC class I molecules, and the class II molecules, found on the surface of antigen-presenting cells (APCs), such as dendritic cell (DC). MHC class I and class II molecules deliver short peptides to the cell surface in a similar manner. $CD8^+$ T cells recognize peptides on the surface of MHC I molecules while $CD4^+$ T cells recognize the ones on MHC II molecules. T_H cells produce interferon gamma (IFN-γ) that is critical for innate and adaptive immunity against viral. Activated $CD4^+$ T cells and $CD8^+$ T cells are also the major sources of IL-2.

6.5 Conclusions

In summary, we successfully produced engineered ferritin-based nanoparticles for the prevention of EBV infections. N-terminus and C-terminus locations were selected as the insertion site for the short peptide EBNA1 epitope. N-terminus insertion and C-terminus insertion presented totally different protein structures. N-terminus insertion presented antigen outside the nanoparticle cage, while C-terminus insertion protected antigens inside the cage. This structural difference significantly affects the corresponding immunogenicity, especially in terms of cell-mediated immune response. C-terminus insertion presented stronger proliferative responses and cell-mediated immune response than N-terminus insertion. Our study will be used as a preliminary immunology investigation as it has been proven that, E1F1 and F1E1 are able to induce potent humoral and cell-mediated immune response after stimulating by the short EBNA1 peptide. Our next-step plan will further test associated immune response after stimulating by full-length EBNA1 antigen and also perform virus neutralization assay to further prove the potential of E1F1 and F1E1 in the prevention of EBV virus. In future application, the type of adjuvant and the insertion site for specific antigens in the ferritin will be extremely important to be considered to improve the immune response induced by engineered ferritin vaccines.

6.6 References

1 López-Sagaseta *et al.* Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

2 Zhao, L. *et al.* Nanoparticle vaccines. *Vaccine* **32**, 327-337, doi:https://doi.org/10.1016/j.vaccine.2013.11.069 (2014).

3 Diaz, D. *et al.* Bioengineering Strategies for Protein-Based Nanoparticles. *Genes (Basel)* **9**, 370, doi:10.3390/genes9070370 (2018).

4 Lee, B.-R. *et al.* Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. *Scientific Reports* **6**, 35182, doi:10.1038/srep35182 https://www.nature.com/articles/srep35182#supplementary-information (2016).

5 Sliepen, K. *et al.* Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. *Retrovirology* **12**, 82, doi:10.1186/s12977-015-0210-4 (2015).

6 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102, doi:10.1038/nature12202 https://www.nature.com/articles/nature12202#supplementary-information (2013).

7 Li, C. Q. *et al.* Ferritin nanoparticle technology... A new platform for antigen presentation and vaccine development. *Industrial Biotechnology* **2**, 143-147, doi:10.1089/ind.2006.2.143 (2006).

8 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

9 Zhen, Z. *et al.* Ferritins as nanoplatforms for imaging and drug delivery. *Expert Opinion Drug Delivery* **11**, 1913-1922, doi:10.1517/17425247.2014.941354 (2014).

10 Han, J.-A. *et al.* Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine: Nanotechnology, Biology and Medicine* **10**, 561-569, doi:https://doi.org/10.1016/j.nano.2013.11.003 (2014).

11 Lee, J.-L. *et al.* Functional expression and production of human H-ferritin in Pichia pastoris. *Biotechnology Letters* **25**, 1019-1023, doi:10.1023/a:1024193104858 (2003).

12 Lee, J.-H. *et al.* Proteinticle Engineering for Accurate 3D Diagnosis. *ACS Nano* **7**,

10879-10886, doi:10.1021/nn404325t (2013).

13 Chul, K. K. *et al.* Enhanced *in vivo* tumor detection by active tumor cell targeting using multiple tumor receptor-binding peptides presented on genetically engineered human ferritin nanoparticles. *Small* **12**, 4241-4253, doi:10.1002/smll.201600917 (2016).

14 Guo, J. *et al.* Efficient expression of recombinant human heavy chain ferritin (FTH1) with modified peptides. *Protein Expression and Purification* **131**, 101-108, doi:https://doi.org/10.1016/j.pep.2016.06.003 (2017).

15 Farrell, P. J. Epstein–Barr virus and cancer. *Annual Review of Pathology: Mechanisms of Disease* **14**, 29-53, doi:10.1146/annurev-pathmechdis-012418-013023 (2019).

16 Pattle, S. B. & Farrell, P. J. The role of Epstein–Barr virus in cancer. *Expert Opinion on Biological Therapy* **6**, 1193-1205, doi:10.1517/14712598.6.11.1193 (2006).

17 Shi, Y. *et al.* Co-infection of Epstein-Barr virus and human papillomavirus in human tumorigenesis. *Chinese Journal of Cancer* **35**, 16-16, doi:10.1186/s40880-016-0079-1 (2016).

18 Mesri, Enrique A. *et al.* Human viral oncogenesis: A cancer hallmarks analysis. *Cell Host & Microbe* **15**, 266-282, doi:https://doi.org/10.1016/j.chom.2014.02.011 (2014).

19 Wilson, J. B. *et al.* EBNA1: Oncogenic activity, immune evasion and biochemical functions provide targets for novel therapeutic strategies against Epstein-Barr virus-associated cancers. *Cancers* **10**, 30, doi:10.3390/cancers10040109 (2018).

20 Destro, F. *et al.* Proteasome inhibitors induce the presentation of an Epstein-Barr virus nuclear antigen 1-derived cytotoxic T lymphocyte epitope in Burkitt's lymphoma cells. *Immunology* **133**, 105-114, doi:10.1111/j.1365-2567.2011.03416.x (2011).

21 McPhillips, T. M. *et al.* Blu-Ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation* **9**, 401-406, doi:10.1107/s0909049502015170 (2002).

22 Kabsch, W. XDS. *Acta Crystallographica Section D Biological Crystallography* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

23 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**, 235-242, doi:10.1107/s0907444910045749 (2011).

24 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674, doi:10.1107/s0021889807021206 (2007).

25 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Journal of Applied Crystallography* **60**, 2126-2132, doi:10.1107/s0907444904019158 (2004).

26 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Journal of Applied Crystallography* **66**, 213-221, doi:10.1107/s0907444909052925 (2010).

27 DeLano, W. L. (New York, 2012).

28 Butts, C. A. *et al.* Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **47**, 12729-12739, doi:10.1021/bi8016735 (2008).

29 Malyala, P. & Singh, M. Endotoxin limits in formulations for preclinical research. *Journal of Pharmaceutical Sciences* **97**, 2041-2044, doi:10.1002/jps.21152 (2008).

30 Hammarström, L. & Smith, C. I. IgG subclass changes in response to vaccination. *Monographs in Allergy* **19**, 241-252 (1986).

31 Lubeck, M. D. *et al.* The interaction of murine IgG subclass proteins with human monocyte Fc receptors. The *Journal of Immunology* **135**, 1299-1304 (1985).

32 Williams, M. A. & Bevan, M. J. Effector and memory CTL differentiation. *Annual Review of Immunology* **25**, 171-192, doi:10.1146/annurev.immunol.25.022106.141548 (2007).

33 Poston, R. N. & Morgan, R. S. Interactions between soluble IgG, complement and cells in lymphocyte and monocyte ADCC. *Immunology* **50**, 461-469 (1983).

34 Paul, W. E. & Seder, R. A. Lymphocyte responses and cytokines. *Cell* **76**, 241-251, doi:https://doi.org/10.1016/0092-8674(94)90332-8 (1994).

35 Datta, S. & Sarvetnick, N. Lymphocyte proliferation in immune-mediated diseases. *Trends in Immunology* **30**, 430-438, doi:https://doi.org/10.1016/j.it.2009.06.002 (2009).

36 Driesen, J., Popov, A. & Schultze, J. L. CD25 as an immune regulatory molecule expressed on myeloid dendritic cells. *Immunobiology* **213**, 849-858, doi:https://doi.org/10.1016/j.imbio.2008.07.026 (2008).

37 Lederman, S. *et al.* Identification of a novel surface protein on activated CD4+ T cells that induces contact-dependent B cell differentiation (help). *Journal of Experimental Medicine* **175**, 1091-1101, doi:10.1084/jem.175.4.1091 (1992).

38 Hoyer, S. *et al.* Concurrent interaction of DCs with CD4(+) and CD8(+) T cells improves secondary CTL expansion: It takes three to tango. *European Journal of Immunology* **44**, 3543-3559, doi:10.1002/eji.201444477 (2014).

39 Wang, K. *et al.* CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. *Experimental Hematology and Oncology* **1**, 36-36, doi:10.1186/2162-3619-1-36 (2012).

40 Krzych, U. *et al.* Memory T cells maintain protracted protection against malaria.

Immunology Letters **161**, 189-195, doi:10.1016/j.imlet.2014.03.011 (2014).

41 Sallusto, F. *et al.* Central memory and effector memory T Cell subsets: Function, generation, and maintenance. *Annual Review of Immunology* **22**, 745-763, doi:10.1146/annurev.immunol.22.012703.104702 (2004).

42 Sckisel, G. D. *et al.* Differential phenotypes of memory CD4 and CD8 T cells in the spleen and peripheral tissues following immunostimulatory therapy. *Journal for ImmunoTherapy of Cancer* **5**, 33, doi:10.1186/s40425-017-0235-4 (2017).

43 Kaech, S. M. *et al.* Effector and memory T-cell differentiation: implications for vaccine development. *Nature Reviews Immunology* **2**, 251-262, doi:10.1038/nri778 (2002).

44 Pati, R. *et al.* Nanoparticle vaccines against infectious diseases. *Frontiers in Immunology* **9**, 2224-2224, doi:10.3389/fimmu.2018.02224 (2018).

45 Skwarczynski, M. & Toth, I. Peptide-based synthetic vaccines. *Chemical Science Journal* **7**, 842-854, doi:10.1039/c5sc03892h (2016).

46 Slingluff, C. L., Jr. The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer J* **17**, 343-350, doi:10.1097/PPO.0b013e318233e5b2 (2011).

47 Purcell, A. W. *et al.* More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery* **6**, 404-414, doi:10.1038/nrd2224 (2007).

48 HogenEsch, H. *et al.* Optimizing the utilization of aluminum adjuvants in vaccines: you might just get what you want. *npj Vaccines* **3**, 51, doi:10.1038/s41541-018-0089-x (2018).

49 Bomford, R. *et al.* The Control of the antibody isotype response to recombinant human immunodeficiency virus gp120 antigen by adjuvants. *AIDS Research and Human Retroviruses* **8**, 1765-1771, doi:10.1089/aid.1992.8.1765 (1992).

50 Comoy, E. E. *et al. In viv*o induction of type 1 and 2 immune responses against protein antigens. *International Immunology* **9**, 523-531, doi:10.1093/intimm/9.4.523 (1997).

51 Jiao, X.-d. *et al.* Comparative study of the effects of aluminum adjuvants and Freund's incomplete adjuvant on the immune response to an Edwardsiella tarda major antigen. *Vaccine* **28**, 1832-1837, doi:https://doi.org/10.1016/j.vaccine.2009.11.083 (2010).

52 Bickham, K. *et al.* EBNA1-specific CD4+ T cells in healthy carriers of Epstein-Barr virus are primarily Th1 in function. *Journal of Clinical Investigation* **107**, 121-130, doi:10.1172/JCI10209 (2001).

53 Münz, C. Epstein-barr virus nuclear antigen 1: from immunologically invisible to a promising T cell target. *Journal of Experimental Medicine* **199**, 1301-1304,

doi:10.1084/jem.20040730 (2004).

54 Cohn, L. & Delamarre, L. Dendritic cell-targeted vaccines. *Frontiers in Immunology* **5**, 255-255, doi:10.3389/fimmu.2014.00255 (2014).

55 Papaleo, E. *et al.* The role of protein loops and linkers in conformational dynamics and allostery. *Chemical Reviews* **116**, 6391-6423, doi:10.1021/acs.chemrev.5b00623 (2016).

56 Wriggers, W. *et al.* Control of protein functional dynamics by peptide linkers. *Peptide Science* **80**, 736-746, doi:10.1002/bip.20291 (2005).

6.7 Supporting information

6.7.1 Supporting figures

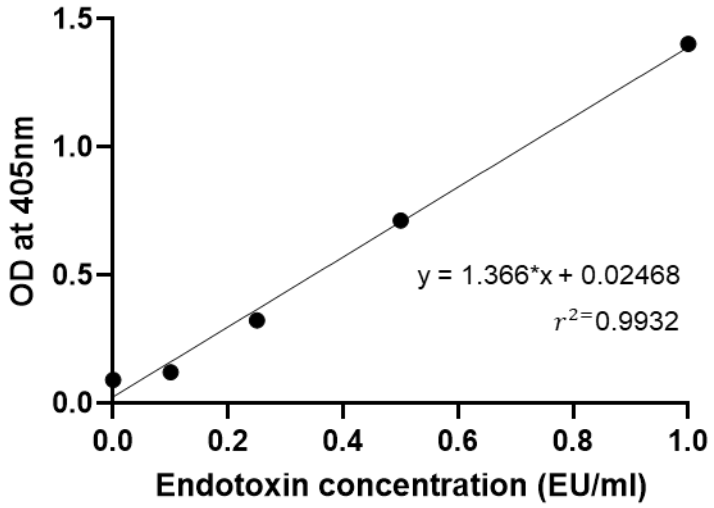


Figure S6.1 Standard curve for endotoxin determination experiment. The equation for this standard curve was shown as above.

6.7.2 Supporting tables

Table S6.1 Calculation of endotoxin limit for the animal model mouse.

Protein sample name	E1F1	F1E1
Dose (mg)	0.2	0.2
Dose in mg h^{-1}	0.0006	0.0006
Weight of mouse (kg)	0.018	0.018
M* (mg kg^{-1} h^{-1})	0.033	0.033
K (EU kg^{-1})	5	5
Endotoxin limit K/M (EU mg^{-1})	152	152

*M is the maximum dose of product per kg of body weight in a single hour period, which is based on the injection frequency of 0.2 mg per 14 days; K is the threshold human pyrogenic dose of endotoxin per kg of body weight; K/M is the endotoxin limit for mouse with unit of EU per mg of proteins.

6.7.3 Other supporting information not included in online publication

Figure S6.2 showed successful construction of E1F1 and F1E1 protein crystals by protein crystallization.

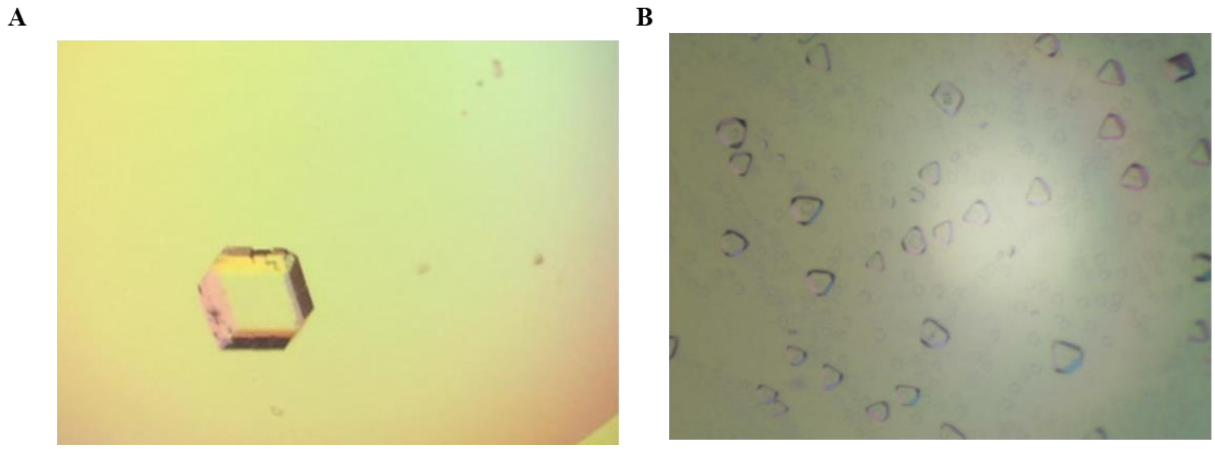


Figure S6.2 Protein crystals for E1F1 and F1E1. A) E1F1 crystal construct (200-300 μm). B) F1E1 crystal construct (50 μm).

Figure S6.3 showed the detailed gating information of cytometry for groups EBNA1, E1F1 and F1E1.

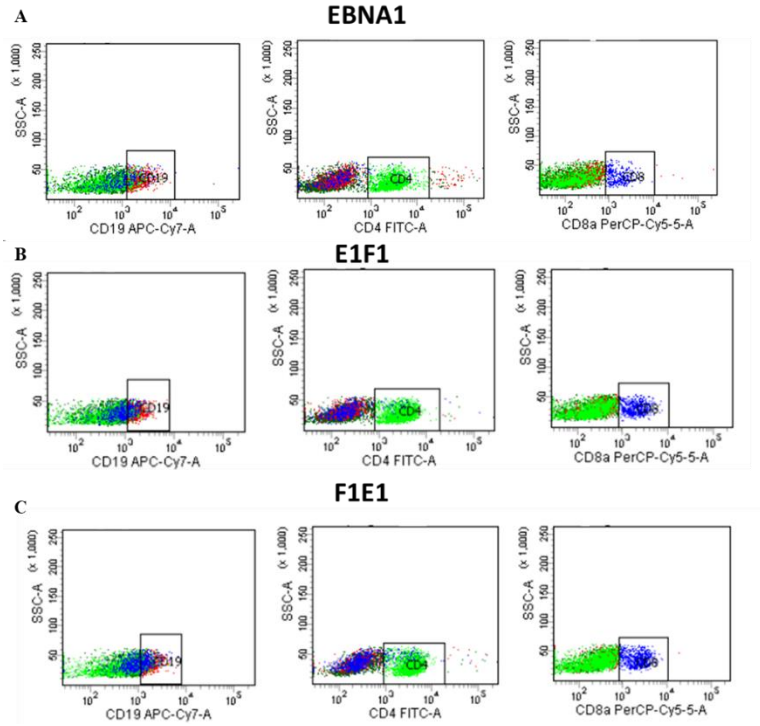


Figure S6.3 Flow cytometry gates showing CD19, CD4 and CD8. CD19 is shown in red; CD 4 is shown in green; and CD8 is shown in blue. A) EBNA1 short peptide. B) E1F1. C) F1E1.

Publication included in Chapter 6

Qu, Y. *et al.* Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* **39**, 4830-4841, doi:https://doi.org/10.1016/j.vaccine.2021.07.021 (2021).

Vaccine 39 (2021) 4830–4841

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope

Yiran Qu [a], Bingyang Zhang [a], Yingli Wang [b], Shuang Yin [a], Jordan L. Pederick [c,d], John B. Bruning [c,d], Yan Sun [e], Anton Middelberg [f], Jingxiu Bi [a,*]

[a] School of Chemical Engineering and Advanced Materials, Faculty of Engineering, Computer and Mathematical Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[b] Shanxi University of Chinese Medicine, Shanxi, China
[c] Institute for Photonics and Advanced Sensing, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia
[d] Department of Molecular and Cellular Biology, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia
[e] Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
[f] Division of Research and Innovation, The University of Adelaide, Adelaide, South Australia 5005, Australia

ARTICLE INFO

Article history:
Received 6 November 2020
Received in revised form 2 July 2021
Accepted 8 July 2021
Available online 18 July 2021

Keywords:
Vaccine
Ferritin nanocage
Epstein-Barr virus
Epstein-Barr nuclear antigen 1
Virus-like particles

ABSTRACT

Human ferritin heavy chain, an example of a protein nanoparticle, has recently been used as a vaccine delivery platform. Human ferritin has advantages of uniform architecture, robust thermal and chemical stabilities, and good biocompatibility and biodegradation. There is however a lack of understanding about the relationship between insertion sites in ferritin (N-terminus and C-terminus) and the corresponding humoral and cell-mediated immune responses. To bridge this gap, we utilized an Epstein-Barr Nuclear Antigen 1 (EBNA1) epitope as a model to produce engineered ferritin-based vaccines E1F1 (N-terminus insertion) and F1E1 (C-terminus insertion) for the prevention of Epstein-Barr virus (EBV) infections. X-ray crystallography confirmed the relative positions of the N-terminus insertion and C-terminus insertion. For N-terminus insertion, the epitopes were located on the exterior surface of ferritin, while for C-terminus insertion, the epitopes were inside the ferritin cage. Based on the results of antigen-specific antibody titers from *in-vivo* tests, we found that there was no obvious difference on humoral immune responses between N-terminus and C-terminus insertion. We also evaluated splenocyte proliferation and memory lymphocyte T cell differentiation. Both results suggested C-terminus insertion produced a stronger proliferative response and cell-mediated immune response than N-terminus insertion. C-terminus insertion of EBNA1 epitope was also processed more efficiently by dendritic cells (DCs) than N-terminus insertion. This provides new insight into the relationship between the insertion site and immunogenicity of ferritin nanoparticle vaccines.

1. Introduction

In the history of vaccine development, the first generation of vaccines used live-attenuated strains of a pathogen or inactivated killed pathogen, both of which have known limitations. The live-attenuated vaccine strains impose high risks of reversion into disease-causing viruses generating more virulent strains. The inactivated killed vaccine pathogens require multiple dosage administrations and adjuvants because of weak immune reactions [1]. There has been a rapid development of innovative vaccine delivery platforms, such as nanoparticle systems and proteins [2].

Nanoparticle proteins self-assemble from multiple protein subunits into hollow structures with desirable features that may overcome these limitations: high stability, biocompatibility and biodegradability [3]. Human ferritin heavy chain is an example of a protein nanoparticle. It is a single particle made up of 24 equivalent monomers, giving inner and outer diameters of 8 and 12 nm respectively [4]. As shown in Fig. 1A, human ferritin heavy chain provides inner cavity by the hollow structure. Various properties of protein nanoparticles have been characterized, revealing very robust thermal and chemical stabilities [1], good ability to present antigens and expose immunogens [5,6,7,8], low toxicity, high biocompatibility and good biodegradability [9]. These properties present ferritin as a desirable nanoparticle vaccine platform. Each monomer of ferritin has 3 commonly used insertion sites; the N-

* Corresponding author.
E-mail address: jingxiu.bi@adelaide.edu.au (J. Bi).

https://doi.org/10.1016/j.vaccine.2021.07.021
0264-410X/

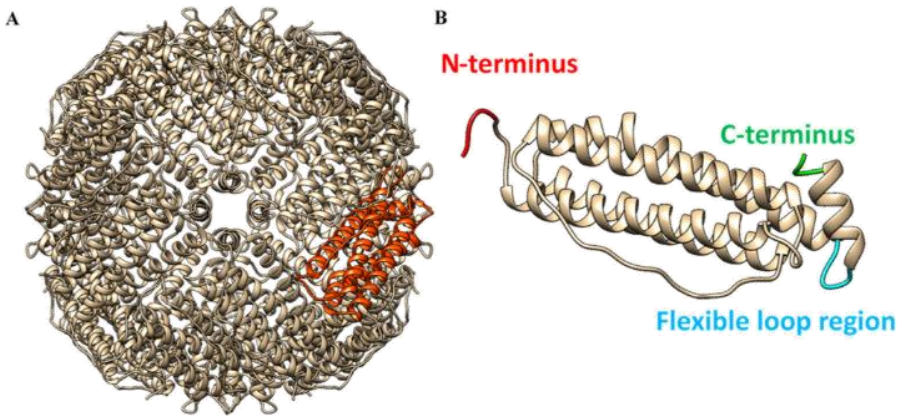


Fig. 1. Human ferritin heavy chain mimic diagram by the program Chimera. A) The native ferritin assembly structure (PDB ID: 1FHA). Orange label represents ferritin monomer. B) Ferritin monomer. Red label represents N-terminus insertion site; Green label represents C-terminus insertion site; Blue label represents flexible loop region insertion site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

terminus, C-terminus and a flexible loop region (Fig. 1B). N-terminus insertion is considered as epitope insertion site in early research exploring ferritin as a vaccine delivery platform. For example, research conducted by Kanekiyo *et al.* [6] and Sliepen *et al.* [5] both selected the N-terminus as the insertion site of epitopes. Up to date, to see if other insertions would result a better dendritic cells (DCs) processing [10], some researchers started selecting the C-terminus or loop region for epitope insertion [8,11,12,13,14]. There are still limited studies investigating the potential relationship between the insertion site and corresponding immunogenicity. In this study, we focus on investigation of the immunogenicity between N-terminus and C-terminus insertion and in particular the ability of ferritin to accommodate a selected epitope.

Epstein-Barr nuclear antigen 1 (EBNA1) was used as the model for exploring an EBNA1-associated vaccine against Epstein-Barr virus (EBV) infection. EBV is one of the human herpesvirus types [15]. It infects B lymphocytes and some epithelial cells and causes various diseases including Burkitt's lymphoma, Hodgkin's disease, lymphomas and lymphoproliferative diseases [16]. More than 90% of human adult populations are infected by EBV and approximately 1.5% of all cancers in the world are associated with EBV infection [15]. The co-infection of EBV and other viruses also increases the risk of cancer. For example, EBV and human papillomavirus (HPV) are related to 38% of all virus-associated cancers [17,18], such as cervical cancer, breast cancer, prostate cancer and lung cancer [17]. Unfortunately, therapy treatment for EBV infections is still under development with a lack of licensed vaccines for EBV infection [15]. EBNA1 is expressed in all EBV-carrying tumours, which acts as a marker to differentiate the virus-associated cancer cells from normal cells. EBNA1 is essential for viral genome maintenance and controlling viral gene expression. As a result, virus is dependent on EBNA1 to persist [19]. As such, the development of an EBNA1 vaccine has the potential to prevent EBV infection and EBV related cancer.

In this study, we inserted an EBNA1 epitope (HPVGEADYFEY), which was found as the main immunogenic domain of EBNA1 [20], at the N-terminus or C-terminus of ferritin with 15 residues of soft linker (GGSGGGGSGGGGSGG) between the antigen and the ferritin. Our main purpose was to develop engineered ferritin vaccines with different insertion sites (N-terminus and C-terminus) to compare their potential immune response against EBV infection. To achieve this aim, engineered ferritin vaccines, EBNA1-ferritin (E1F1) (N-terminus insertion) and ferritin-EBNA1 (F1E1) (C-terminus insertion), were first successfully expressed and purified. Characterization experiments were also performed to confirm purity, size and structure of E1F1 and F1E1. To determine if both E1F1 and F1E1 could induce potent immune responses, mice were then injected with either engineered ferritin vaccine to determine antigen-specific IgG titers, proliferation indices, and lymphocyte T cells differentiation. These results will be used as a preliminary immunology study of E1F1 and F1E1 to test the potential on if they are able to induce potent immune responses after stimulating by the short EBNA1 peptide. Thus, the successful outcomes will build confidence on further immunity experiments design including stimulating by the full length of EBNA1 antigen along with virus neutralization assay. Our study provides new understanding regarding the relationship between epitope insertion sites of ferritin and corresponding immunogenicity, which might provide guidance for the future development of ferritin-based vaccines that illicit strong immune responses.

2. Material & methods

2.1. Plasmid generation, expression and purification of recombinant ferritins

The pET 30a vector was used to construct plasmids harboring genes of two types of EBNA1 fused ferritins, E1F1 and F1E1, inserted between the *Nde I* and *BamH I* restriction sites. The epitope was selected as EBNA1 (HPVGEADYFEY). A fifteen-residue linker (GGSGGGGSGGGGSGG) was inserted between the epitope and ferritin in both ferritin complexes. The plasmids were transformed into *Escherichia coli* (*E. coli*) BL21 (DE3). The expression process of the two engineered ferritin constructs was the same. A single bacterial colony transformed with pET 30a containing E1F1 or F1E1 was inoculated in 50 ml LB medium supplemented with kanamycin (100 μg/ml), and incubated overnight at 37 °C in an orbital mixer incubator (Ratek), shaken at 180 rpm. A 10 ml of exponentially growing bacterial culture was then transferred to 500 ml of LB medium supplemented with kanamycin (100 μg/ml). The culture was grown at 37 °C and shaken at 200 rpm until OD_{600} reached 0.8, followed by induction with 1 mM IPTG at 37 °C for 4 h. Cells were collected by centrifugation at 13,751 xg. Cell pellets were resuspended in lysis buffer (20 mM PBS, 2 mM EDTA, pH 7.0) and lysed by mechanical disruption. The bacterial lysate was centrifuged at 19,802 xg for 30 min to remove cell debris and supernatant was harvested for protein purification.

The protein purification processes for E1F1 and F1E1 were the same, except they differed in the chromatographic resin used. First, clarified supernatant was diluted with Milli-Q water to a final protein concentration of 4 mg/ml. NaCl and NaAc-HAc were added to a final concentration of 1 M and 100 mM respectively and the pH was adjusted to 5.0 to the diluted protein solutions and the mixtures underwent thermal precipitation at 60 °C for 10 min. Centrifugation (19,802 xg at 4 °C for 10 min) was conducted to remove precipitated impurities. The resultant supernatant pH was adjusted to 6.5 by using 1 M phosphate buffer (PBS; Na_2HPO_4-NaH_2PO4), pH 8.0, and 1 M ammonium sulfate was added for purification by hydrophobic interaction chromatography (HIC). HIC was performed on an AKTA pure system (GE Healthcare, USA), with a HiTrap Octyl FF column (GE Healthcare, USA) used for E1F1, and a HiTrap Butyl FF column (GE Healthcare, USA) used for F1E1. The equilibration buffer was 100 mM PBS and 1 M ammonium sulfate, pH 6.5. The elution buffer was 20 mM PBS, pH 6.5. Briefly, samples were loaded onto columns equilibrated with equilibration buffer. After washing the column for 5 CV with equilibration buffer, a linear gradient elution of 0%-100 % elution buffer in 3 CV was conducted and the second eluted peak was collected. Absorbance at 280 and 260 nm were recorded. Finally, collected fractions from HIC were buffer-exchanged to 20 mM PBS, pH 7.0 using a HiTrap desalting column (GE Healthcare, USA). Protein expression and purity were analyzed by SDS-PAGE.

2.2. Protein characterization

2.2.1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Analysis of protein expression and purification was performed under 5% (w/v) SDS spacer gel and 12% (w/v) SDS separation gel. Samples were diluted with the same amount of 5 × loading buffer (6% (v/v) 1 M Tris-HCl pH 6.8, 25% (v/v) glycerol, 1 mg/ml bromophenol blue, 2% (v/v) SDS and 5% (v/v) β-mercaptoethanol). Treated samples were heated at 100 °C for 10 min and were separated by electrophoresis (Bio-Rad). The gel was stained with 0.25% (w/v) Coomassie R-250 for 60 min and the background was washed using a washing buffer (10% (v/v) ethanol and 10% (v/v) acetic acid) for 60 min.

2.2.2. Size exclusion chromatography coupled with multiple-angle laser light scattering (SEC-MALS)

SEC-MALS was performed to determine hydrodynamic radius of purified engineered ferritins, E1F1 and F1E1. Superose 6 10/300 GL (GE Healthcare, USA) was connected to High Performance Liquid Chromatography (HPLC) (Shimazdu, Japan) coupled with Wyatt Optilab refractive index (RI) and Wyatt DAWN MALLS detector (USA). Equilibration buffer was 20 mM PBS, pH 7.0. Protein concentration was 5 mg/ml. Loading volume was 50 μL. The output chromatogram was processed by Origin 2021 to calculate integrated area for each peak. Purity level was estimated using Eq. (1).

$$\text{Purity}\ (\%) = \frac{A_{\text{Target protein}}}{A_{\text{Total protein}}} \quad (1)$$

where $A_{\text{Target protein}}$ is the integrated area of target protein peak and $A_{\text{Total protein}}$ is the sum of integrated area of target protein peak and impure protein peaks.

2.3. Mass spectrometry

Purified protein samples were analyzed using an Agilent 6230 TOF mass spectrometer (MS) equipped with Agilent 1260 LC system. Instrument was tuned at 3200 m/z mass range in positive ion mode. Using ESI source, 1 ml of the samples was injected into MS via the LC auto sampler and eluted without chromatographic separation at a flow rate of 0.5 ml/min using isocratic elution of 50% buffer B, where buffer A is 0.1% formic acid in water and buffer B is 0.1% formic acid in acetonitrile. Mass spectra were acquired under the control of MassHunter Workstation software, LC-MS Data Acquisition (version B.08.00, Agilent technologies, USA).

2.4. Transmission electron microscopy (TEM)

E1F1 and F1E1 samples were diluted to a concentration of 0.1 mg/ml. Carbon-coated copper grids were incubated face down on 10 μL of sample for 10 min. The grid was washed 3 times in droplets of Milli-Q water, and then negatively stained in 2% (v/v) uranyl acetate for 2 min. Excess liquid was removed with filter paper. Grids were analyzed with a Philips CM100 transmission electron microscope operated at 100 kV (Field Electron and Ion Company, USA).

2.5. Protein crystallization and data processing

For crystallization, previously purified E1F1 and F1E1 were dialyzed to 10 mM Tris pH 7.5 at concentrations of 5 mg/ml and 10 mg/ml respectively. Crystallization trials were completed using the PEGION, Index and Crystal Screen sparse matrix screens (Hampton Research) with 96 well Intelliplates (Art Robbins) in a sitting drop format at 16 °C. Wells contained 80 μL of reservoir solution, with drops consisting of 1 μL ferritin construct and 1 μL of reservoir solution. Conditions for crystallization differed for the two constructs. For E1F1, crystals were obtained in a reservoir solution of 0.1 M ammonium citrate tribasic pH 7.0 with 12% PEG 3350. F1E1 crystals were obtained in a reservoir solution of 0.1 M HEPES pH 7.0 with 30% Jeffamine M-600 pH 7.0. In both cases, crystals formed within 1 week.

2.6. Data collection, processing and refinement

Crystals of E1F1 and F1E1 were transferred to Paratone-N for cryoprotection and flash-frozen in liquid nitrogen. Diffraction images were collected at the MX1 beamline of the Australian Synchrotron [21]. Data were integrated and indexed using XDS [22], and scaled and merged using Aimless (CCP4) [23]. The structures of E1F1 and F1E1 were solved by molecular replacement with Phaser, using a previously solved structure of heavy chain ferritin (PDB: 6J4A) with solvent and ligands removed [24]. Solutions from Phaser underwent multiple rounds of refinement in Phenix, followed by multiple rounds of model building in Coot, until R-factors converged [25,26]. Both structures possess good overall geometry, with >98% of residues favored in the Ramachandran plot and no outliers. All residues from aa. 5 – 176 were modelled, except for aa. 88 – 96, which were unable to be completely modelled for half of the chains of E1F1, and all chains of F1E1 due to ambiguous electron density. Detailed crystallographic statistics for E1F1 and F1E1 were recorded in Table 2.

2.7. Structure comparison and visualization

Structural superposition was completed by sequence independent alignment in PyMol [27]. Figures for visualizing protein structure were generated using PyMol. Simulated annealing 2Fo-Fc composite omit maps were generated in Phenix.

2.8. Endotoxin concentration determination

Pierce™ Chromogenic Endotoxin Quant Kit (Thermofisher, USA) was used to determinate endotoxin concentration of each protein sample (E1F1 and F1E1). High Standards (0.10, 0.25, 0.50 and

1.00 EU/ml) of *E. coli* endotoxin standard solution was prepared accordingly. Lyophilized amebocyte lysate and chromogenic substrate were reconstituted with endotoxin-free water (EFW) of 1.7 ml and 3.4 ml, respectively. Protein samples (E1F1 and F1E1) with protein concentration of 1 mg/ml in 20 mM PB buffer, pH 7.0, were diluted with EFW to protein concentration of 0.1 mg/ml.

The endotoxin-free 96-well plate was pre-equilibrated in a heating block at 37 °C. 50 μL of endotoxin standard dilutions, blank (EFW), and diluted protein samples (E1F1 and F1E1) were added per well. Each sample, blank and standards were run in triplicate. The plate was kept at 37 °C, and 50 μL of the reconstituted amebocyte lysate reagent was added per well. Once the amebocyte lysate reagent has been added to the plate well, the plate was removed from the heater and mixed by tapping 10 times on the side of the plate. The plate was then incubated at 37 °C for 12 min. After exactly time 12 min, 100 μL of pre-warmed reconstituted chromogenic substrate solution was added per well. Once the substrate solution was added, the plate was removed from the heater and mixed gently by tapping 10 times to facilitate mixing. The plate was incubated at 37 °C for 6 min. After that, 50 μL of stop solution (25% acetic acid) was added per well. Once the stop solution was added, the plate was removed from the plate heater and mixed by tapping 10 times on the side of the plate. The optical density (OD) at 405 nm was read via microplate reader (BioTek, USA). The average absorbance of the blank replicates was subtracted from the average absorbance of all individual standards and sample replicates to calculate mean absorbance. The standard curve (linear regression) was prepared by plotting the average blank-corrected absorbance for each standard on the y-axis against the corresponding endotoxin concentration in EU/ml on the x-axis. The coefficient of determination, r^2 was required to be ≥0.98. The endotoxin concentration of each protein sample was determined using the formulated standard curve. The final endotoxin concentration of each protein sample was after multiplying by 10 times dilution.

2.9. Mice immunization

All animal experiments were approved by the Medical Ethics Committee of ShanXi University of Chinese Medicine (Approval Number 2019LL137).

Three groups (n = 6/treatment) of 6 to 8 week-old female BALB/c mice were immunized intraperitoneally three times (Day 0, 14, 28) with 6 μg of EBNA1 short peptide epitope without adjuvant (Group 1: control), purified E1F1 (1 mg/ml) and F1E1 (1 mg/ml) in 200 μL PBS (Group 2 and 3: samples) (equivalent mass amount of EBNA1 epitope) mixed with an equal volume amount of endotoxin free aluminum hydroxide (0.8 mg/ml). Two extra groups (n = 6/treatment) of mice were treated with 200 μL PBS (negative control) and 100 μg of Ovalbumin (OVA), respectively, used for check if the system and animal models were working. The serum was collected at 14 and 28 days after first time immunization and stored at −80 °C.

2.10. Enzyme-linked immunosorbent assay (ELISA)

EBNA1 epitope (HPVGEADYFEY) was adsorbed overnight at 4 °C in 96-well plates at 10 μg/ml in 50 mM sodium carbonate buffer pH 9.6. After being blocked with PBS containing 1% bovine serum albumin (BSA) for 2 h at 37 °C, the plate was incubated with serial dilutions of mouse sera, collected at 14, 28 days after the first immunization from each group, for 2 h at 37 °C. The plate was washed three times with PBS containing 0.1% Tween-20. A goat anti-mouse IgG horseradish peroxidase-conjugated antibody was applied at a 1:5000 dilution, and 200 μL were added to the plate. After being incubated at 37 °C for 2 h, the plates were washed, and 3,3′,5,5′-Tetramethylbenzidine (TMB) Single-Component Substrate solution was added for color development. After 10 min, 1 M sulphuric acid was added to stop the color development. After 10 min, the plates were analyzed using a microplate reader (Perlong, China) at wavelength 450 nm. Endpoint titers were defined as the highest serum dilution that resulted in an absorbance value two times greater than that of negative control sera derived from the non-immunized mice.

2.11. Antigen-specific IgG isotype

The procedure for antigen-specific IgG isotype followed ELISA method as outlined in Section 2.10. EBNA1 short peptide (HPVGEADYFEY) was absorbed overnight at 4 °C in 96-well plates at 10 μg/ml in 50 mM sodium carbonate buffer at pH 9.6. After being blocked with phosphate-buffered saline (PBS) containing 1% bovine serum albumin (BSA) for 2 h at 37 °C, the plate was incubated with 100 μL diluted mouse sera (same dilution of end point) prepared before for ELISA, for 2 h at 37 °C. The plate was washed three times with PBS containing 0.1% Tween-20. Antibodies goat anti-mouse IgG1 and IgG2a were applied at 1:1000 dilution, and 100 μL were added to the plate. A rabbit anti-goat IgG horseradish peroxidase-conjugated antibody was applied at a 1:5000 dilution, and 200 μL were added to the plate. The remained procedure was exactly same as Section 2.10.

2.12. T-cell proliferation assay

Murine splenocytes were harvested using red blood cell lysing buffer (Sigma). Single-cell suspensions were prepared (4×10^6 cells/ml) and cultured in RPMI 1640 (Gibco, Germany) with a short EBNA1 epitope at 20 μg/ml. The plates were then incubated for 48 h before the addition of 10 μL of CCK-8 (Dojindo, Japan) per well. The absorbance at 450 nm was recorded in a microplate reader (Perlong, China) after 4 h incubation at 37 °C. The results were presented as proliferation indexes (PI) using the following equations (Eq. (2) and Eq. (3)).

$$Proliferation\ \ rate = \frac{A(sample) - A_0(control)}{A_0(control)} \quad (2)$$

A(sample): A_{450} nm of the sample group with CCK-8 and stimulator (EBNA1 epitope, E1F1 or F1E1)
A_0(control): A_{450} nm of the control group with CCK-8 only

$$Proliferation\ \ Index = \frac{Proliferation\ \ rate\ \ (sample)}{Proliferation\ \ rate\ \ (PBS)} \quad (3)$$

2.13. Lymphocyte activation

Splenocytes from immunized mice were cultured for 60 h at 37 °C with and without EBNA1 epitope before staining with FITC CD4 Monoclonal antibody, PerCP-Cy 5.5 CD8a Monoclonal antibody, APC-eFluor 780 CD19 Monoclonal antibody and APC CD25 Monoclonal antibody (Thermo fisher, USA) at 4 °C. After washing with PBS, fluorescence patterns were analyzed using a FACSCanto II (BD Bioscience, USA). The percentage of CD25/CD4 T cells, CD25/CD8 T cells and CD25/CD19 B cells were recorded and analyzed using FACSCanto II software (BD Bioscience, USA).

The procedure to detect memory lymphocyte T cells differentiation was same as above. After the gate for CD25 was circled, gates for CD44 and CD62L were circled in $CD4^+$ T and $CD8^+$ T cells, respectively.

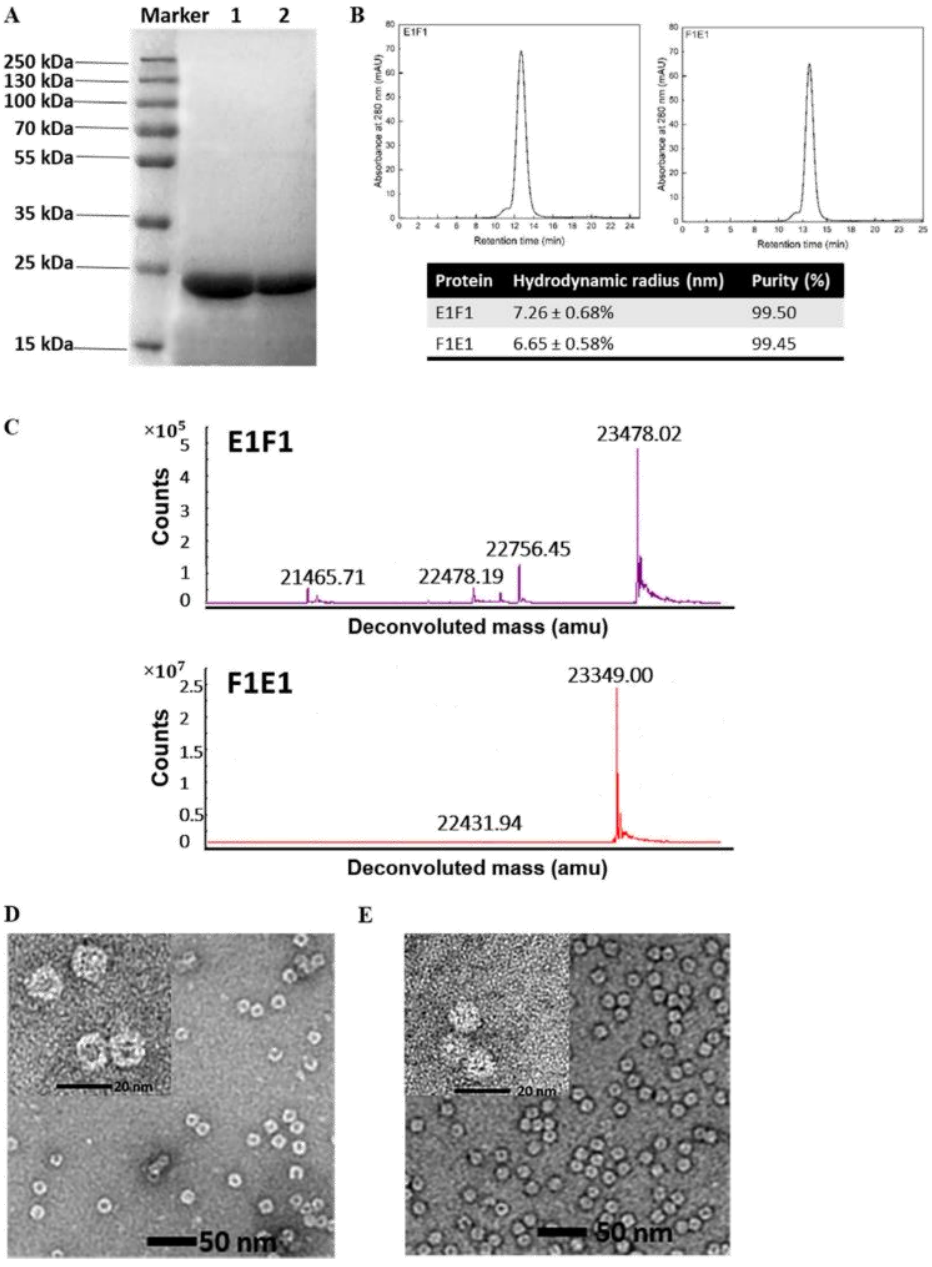
2.14. Statistical analysis

Data generated from in vivo immunology studies were presented in Mean Standard deviation (SD). GraphPad Prism 9 was applied to determine statistical significance of results in antigen-specific antibodies titers, T-cell proliferation assay and lymphocyte activation following One-way ANOVA method. P-value < 0.05 (*) was considered as significant. P-value < 0.01 (**) was moderately significant. P-value < 0.001 (***) was highly significant.

3. Results

3.1. Characterization of engineered ferritins

The production procedures (including expression and purification) for engineered ferritins were general by applying advantages of native ferritins, such as robust thermal and chemical stabilities. Characterization of engineered ferritins are significant to check if they meet quality criteria in terms of satisfied purity and well-



Protein	Hydrodynamic radius (nm)	Purity (%)
E1F1	7.26 ± 0.68%	99.50
F1E1	6.65 ± 0.58%	99.45

Fig. 2. Characterization of E1F1 and F1E1. (A) SDS PAGE analysis post purification: Marker: protein marker; lane 1: E1F1; lane 2: F1E1. (B) Hydrodynamic radius measurements and purity level estimation from size exclusion chromatography with multiple-angle laser light scattering (SEC-MALS). (C) Mass spectra of E1F1 and F1E1. The molecular weights for the E1F1 and F1E1 monomers were determined as 23.48 and 23.35 kDa, respectively. (D) TEM image of E1F1. (E) TEM image of F1E1.

assembled structures, before performing animal experiments. It also helps reveal the potential structural differences between E1F1 and F1E1, in order to build connections with further immunology results. The molecular weight for monomer of E1F1 and F1E1 is 23.35 kDa based on amino acid sequence estimation by ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss). E1F1 and F1E1 showed the same molecular weight by SDS-PAGE as expected (Fig. 2A). From the mass spectrum, the molecular weight for the monomers of E1F1 and F1E1 were 23.48 and 23.35 kDa, respectively (Fig. 2C), matching the expected size of the engineered ferritins with the EBNA1 epitope intact. The slight difference in mass of E1F1 and F1E1 seems like matching the start methionine (M). It may suggest that, for F1E1, M was removed in *E. coli* because of similarity with bacterial ferritin. The hydrodynamic radius measured by SEC-MALS indicated a particle size difference between E1F1 and F1E1, with sizes 7.26 and 6.65 nm, respectively (Fig. 2B). These sizes were consistent with previous measured hydrodynamic radii of ferritin nanoparticles (6.7 nm) [28]. Both SDS-PAGE and SEC results confirmed that purity of E1F1 and F1E1, which were 99.50% and 99.45%, respectively (Fig. 2A, B). It indicates that less than 0.6% impurity is process-related proteins (for example, host cell protein (HCP) of *E. coli*) and product-related proteins (for example, dissociated or denature HFn). The high purity of protein sample for vaccine performance evaluation is critical.

The endotoxin determination experiment was conducted to ensure the immunology result and vaccine performance validation is accurate. The endotoxin standard curve was shown in Fig. S1 (Supporting Information). The equation was valid with R-squared greater than 0.98. Shown by Table 1, the endotoxin concentration in the injection protein sample for E1F1 and F1E1 were 9.67 and 9.56 EU per mg of protein, respectively. They are much less than the endotoxin limit of 152 EU/mg protein for preclinical research animal models. The detail of endotoxin limits for preclinical research animal models is calculated and shown in Table S1, which is based on the threshold pyrogenic human dose of 5 EU/kg [29].

Purified proteins E1F1 and F1E1 both presented well-assembled structures based on TEM images (Fig. 2D, E). To verify the location of epitopes for different insertion site, the structures of E1F1 and F1E1 were solved by x-ray crystallography to 2.8 and 2.2 Å resolution, respectively (Table 2). Both engineered ferritin constructs displayed the same nanoparticle assembly and monomer structure compared to native heavy chain ferritin (Fig. 3A). No electron density was observed for the EBNA1 epitope and Gly-Ser linker in either structure (Fig. 3B). As the structural integrity was verified by mass spectrometry, this is likely due to the flexibility of the insert. As for native heavy chain ferritin, the N-terminus and C-terminus of each ferritin monomer projected towards the exterior and interior of the nanoparticle, respectively (Fig. 3C). This supports E1F1 presenting the EBNA1 epitope on the surface of the nanoparticle, while the epitope is contained within the ferritin cage for F1E1 and is consistent with the larger hydrodynamic radius measured for the E1F1 construct.

Table 1
Endotoxin concentration of protein samples.

Protein sample name	E1F1	F1E1
Average OD (405 nm)	1.35	1.33
Endotoxin concentration by 10× dilution (EU/ml)	0.967	0.956
Actual endotoxin concentration (EU/ml)	9.67	9.56
Total protein injection (mg)	0.2	0.2
Total Endotoxin amount per dose (EU/mg protein)	9.67	9.56

Table 2
Crystallographic statistics for structures of E1F1 and F1E1.

	E1F1 (PDB: 7KE5)	F1E1 (PDB: 7KE3)
Wavelength (Å)	0.9537	0.9537
Resolution range (Å)	48.3–2.8 (2.9–2.8) [a]	48.72–2.2 (2.279–2.2) [a]
Space group	P 4_2 2_1 2	P 4_2 2_1 2
Unit cell		
a, b, c (Å)	219.323 219.323	217.882 217.882
α, β, γ (°)	148.556	147.203
	90 90 90	90 90 90
Total reflections	1,766,537 (126783)	4,871,002 (477929)
Unique reflections	85,428 (8054)	177,897 (17576)
Multiplicity	20.7 (15.7)	27.4 (27.2)
Completeness (%)	91.01 (89.44)	99.02 (98.17)
Mean I/sigma(I)	7.48 (2.53)	8.86 (1.94)
Wilson B-factor	28.80	24.55
R-merge	0.2514 (0.8948)	0.3341 (2.091)
R-meas	0.2579 (0.9243)	0.3404 (2.131)
R-pim	0.05599 (0.2261)	0.06503 (0.4079)
CC1/2	0.998 (0.975)	0.999 (0.911)
Reflections used in refinement	81,188 (7863)	176,417 (17273)
Reflections used for R-free	4127 (396)	8842 (873)
R-work	0.2726 (0.3180)	0.2408 (0.3245)
R-free	0.3141 (0.3808)	0.2709 (0.3522)
Number of non-hydrogen atoms	16,844	17,042
macromolecules	16,745	16,263
ligands	64	68
solvent	35	711
Protein residues	2064	2001
RMS(bonds)	0.009	0.007
RMS(angles)	0.99	0.85
Ramachandran favoured (%)	98.58	98.72
Ramachandran allowed (%)	1.42	1.28
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	2.02	1.22
Clashscore	3.81	2.50
Average B-factor	27.79	26.48
macromolecules	27.80	26.23
ligands	29.31	33.32
solvent	21.05	31.71

[a] Statistics for the highest resolution shell are shown in parentheses.

3.2. Evaluation of antigen-specific antibodies titers

To confirm the immunogenicity of purified engineered ferritin constructs, an in vivo model was applied. The measurement of antigen-specific IgG titers is an indicator of immunogenicity induced by a vaccine. Compared to the EBNA1 epitope, both E1F1 and F1E1 showed stronger humoral immunogenicity after the second and third immunizations indicated by greater IgG titers (Fig. 4A). The immune response was improved significantly after the third immunization. There was no obvious difference between E1F1 and F1E1 regarding humoral immunogenicity. To identify the immunization pathway towards humoral immune response or cell-mediated immune response, the ratio between IgG2a and IgG1 was measured [30]. If IgG2a/IgG1 is greater than 1, the immune response tends to be Th1 biased, which is related to a cell-mediated response. If this ratio is less than 1, the immune response tends to be Th2 biased, indicating humoral immune response [31]. Th1 mainly secretes cytokines including IL-12, TNF-α and IFN-γ, which activate a cytotoxic T lymphocyte (CTL) response that is necessary for the control of a variety of bacterial and viral infections [32]. Th2 mainly secretes cytokines including IL-4, IL-5 and IL-6 to activate a humoral immune response and promote the antibody-dependent cell-mediated cytotoxicity (ADCC) system [33]. As shown in Fig. 4B, the IgG2a/IgG1 ratio induced by EBNA1 short epitope without adjuvant was slightly greater than 1, suggesting similar level of humoral and cell-mediated immune

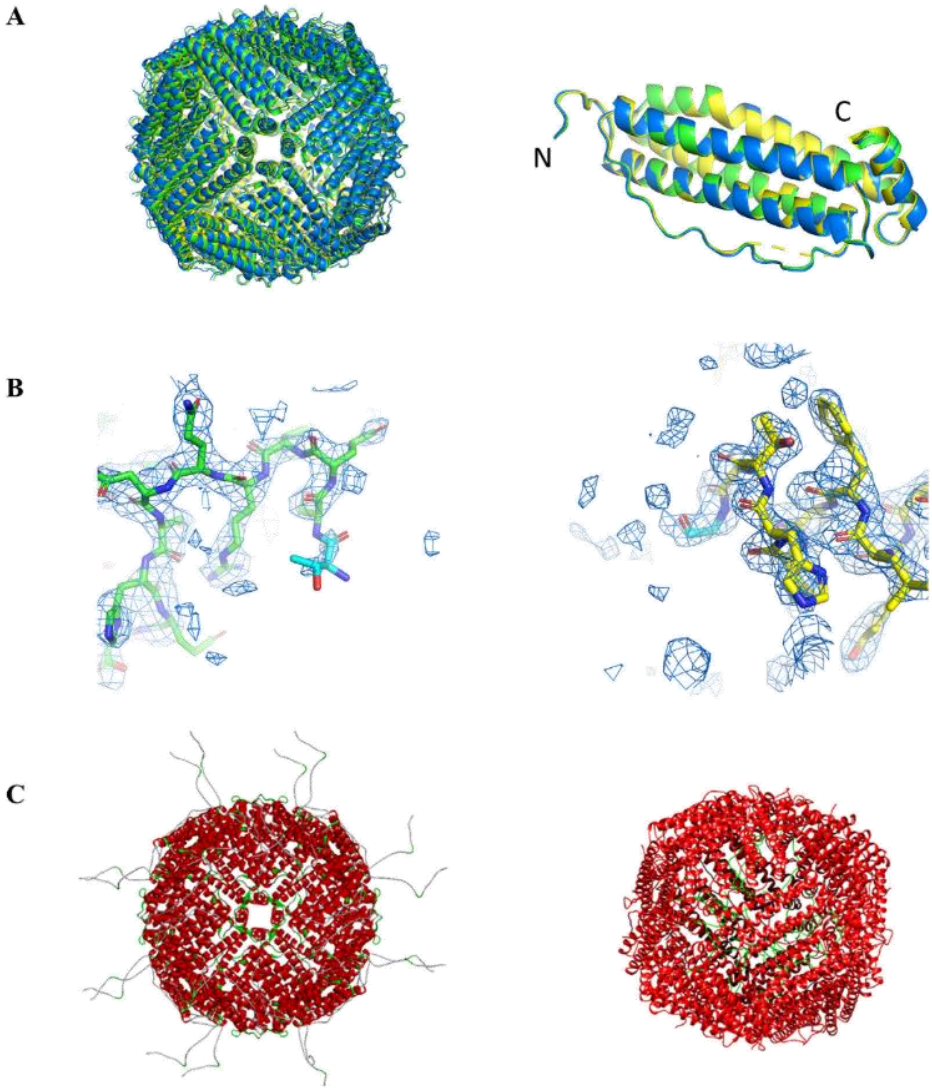


Fig. 3. Structural determination of E1F1 and F1E1 (A) Structural superposition of the nanoparticle assembly (left) and a single ferritin monomer (right) for E1F1 (green; RMSD = 0.33) and F1E1 (yellow; RMSD = 0.40) with native heavy chain ferritin (blue, PDB:1FHA). (B) Simulated annealing 2Fo-Fc composite omit maps (blue mesh, 1.5σ) for the N-terminus of E1F1 (green, Chain A; left) and C-terminus of F1E1 (yellow, Chain A; right). No electron density was observed for the EBNA1 epitope for either construct. Terminal residues are shown in cyan. The presented model is representative of each monomer forming the E1F1 and F1E1 nanoparticles. (C) E1F1 (left) and F1E1 (right) modelled with the EBNA1 epitope. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

response strength. However, for both E1F1 and F1E1, which were administered with aluminum hydroxide adjuvant, the ratio of IgG2a/IgG1 was < 1, suggesting a humoral immune response was dominant for this case.

3.3. T-cell proliferation assay

There are specific receptors on the surface of B cells and T cells that recognize antigens. After being stimulated by specific antigen, in this case EBNA1 epitope, the lymphocytes perform mitosis [34]. The level of splenocyte proliferative responses gives an indication of the strength of the immune response induced by a vaccine, which is measured as proliferative index (PI) [35]. The PI for EBNA1 epitope, E1F1 and F1E1 were 1.2, 1.5 and 2.3, respectively (Fig. 5). There was a significant difference in PI depending on the insertion site of EBNA1 epitope, with F1E1 producing a greater PI than E1F1 (***$p < 0.001$), indicative of a stronger immune response for C-terminus insertion.

3.4. Lymphocyte activation

CD25 is regarded as an activation marker for B cells and T cells [36], which represents the activation level for B cells and T cells after stimulating with antigen, such as the EBNA1 epitope. $CD4^+$ T cells are T helper cells that play an important role in instigating adaptive immune responses [37]. $CD8^+$ T cells, like $CD4^+$ helper T cells, are crucial for immune defense against intracellular pathogens [38]. After recognition by the antigen and activation, $CD8^+$ T cells will secrete cytokines, mainly TNF-α and IFN-γ, which are related to the cell-mediated immune response. CD19 is a biomar-

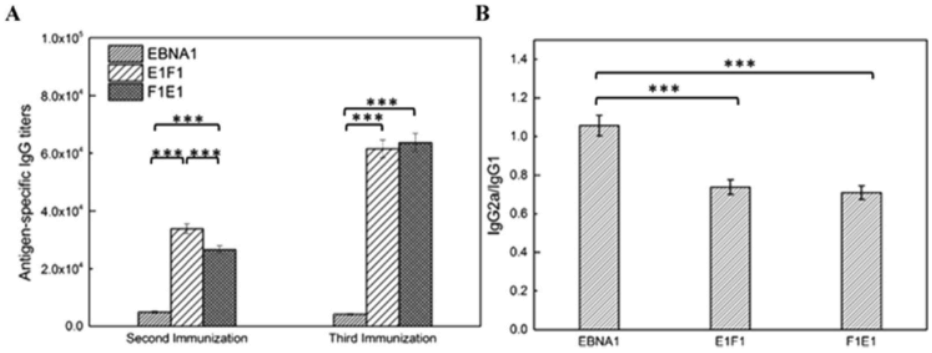


Fig. 4. Production of antigen-specific antibodies in the sera of BALB/c mice. The dosage of groups was 6 μg EBNA1. (A) The antigen-specific IgG titers at day 14 (second immunization) and day 28 (third immunization). (B) The ratio of IgG2a/IgG1 at day 28. Data are expressed as mean SEM (n = 6) (*p < 0.05; **p < 0.01; ***p < 0.001), where n represents the number of mice from each group.

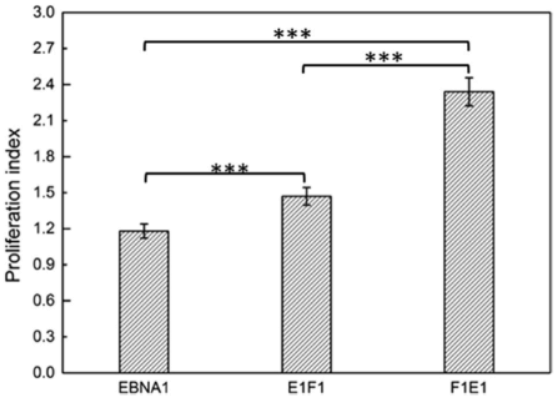


Fig. 5. Proliferative responses of splenocytes responding to antigen simulation ex vivo. Splenocytes were harvested 10 d after the third immunization and restimulated ex vivo with antigen EBNA1 (20 μg/ml). Splenocytes proliferation was measured using CCK-8 kit, and the proliferation index was calculated. (EBNA1: 1.18; E1F1: 1.47; F1E1: 2.34) Data are expressed as mean SEM (n = 6) (*p < 0.05; **p < 0.01; ***p < 0.001), where n represents number of mice from each group.

ker for B cells, which is important to maintain the balance between humoral and antigen-induced response [39]. By measuring the relative number of $CD4^+$ T cells, $CD8^+$ T cells and $CD19^+$ B cells, the strength of humoral immune response and cell-mediated immune response could be compared. As shown in Fig. 6, after activation, cellular levels for $CD4^+$ T cells, $CD8^+$ T cells and $CD19^+$ B cells induced by E1F1 and F1E1 were significantly increased compared to before activation (***p < 0.001). EBNA1 short peptide vaccine showed negligible activation level (Fig. 6A and C). There were approximately 2 times $CD4^+$ T cells (Fig. 6A) and 6 times $CD8^+$ T cells (Fig. 6B) more found in engineered ferritin groups than EBNA1 short peptide, which means activation by EBNA1 was more effective for engineered ferritins with adjuvant groups than short peptide without adjuvant. There were no obvious differences found between E1F1 and F1E1 for $CD4^+$ T cells and $CD19^+$ B cells after activation (Fig. 6A and C). Amount of $CD8^+$ T cells after activation for F1E1 was obviously 11% greater than E1F1 (**p < 0.01) (Fig. 6B). This result further supports that C-terminus insertion induced stronger cell-mediated immune responses than N-terminus insertion.

3.5. Memory lymphocyte T cells differentiation

One of the purposes of vaccination is to acquire immunologic memory by maintaining memory T cells ($CD4^+$ and $CD8^+$) to introduce fast and effective actions on the antigen-specific immune response [40]. Memory T cells are divided into effector T cells (T_{EM}) and central memory T cells (T_{CM}) [41]. The differences between T_{EM} and T_{CM} are: the start-up for T_{EM} is faster compared to T_{CM} but with shorter lifetime. The process that T_{EM} and T_{CM} can be transferred to each other is activated by the stimulator (short EBNA1 peptide). For a strong immune response, cells amount for T_{CM} before the stimulation and T_{EM} after the stimulation, are expected to be significantly high. The expression of CD44 and CD62L population indicates if memory T cells are T_{CM} or T_{EM} [42]. T_{CM} highly express both CD44 and CD62L ($CD44^{hi}$ $CD62L^{hi}$), while T_{EM} are restricted to highly express CD44 only

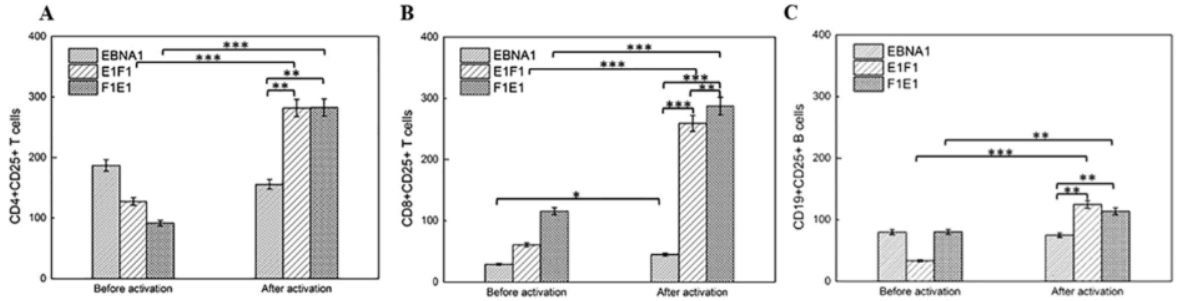


Fig. 6. Frequencies of $CD25^+$ in $CD4^+$, $CD8^+$ T cells and $CD19^+$ B cells were determined by flow cytometry. T cells amount were count for each group before and after activation by short EBNA1 peptide (500 μL, 20 μg/ml). Data are expressed as mean SEM (n = 6) (*p < 0.05; **p < 0.01; ***p < 0.001), where n represents the number of mice from each group. (A) Frequencies of $CD25^+$ in $CD4^+$. (B) Frequencies of $CD25^+$ in $CD8^+$ T cells. (C) Frequencies of $CD25^+$ in $CD19^+$ B cells.

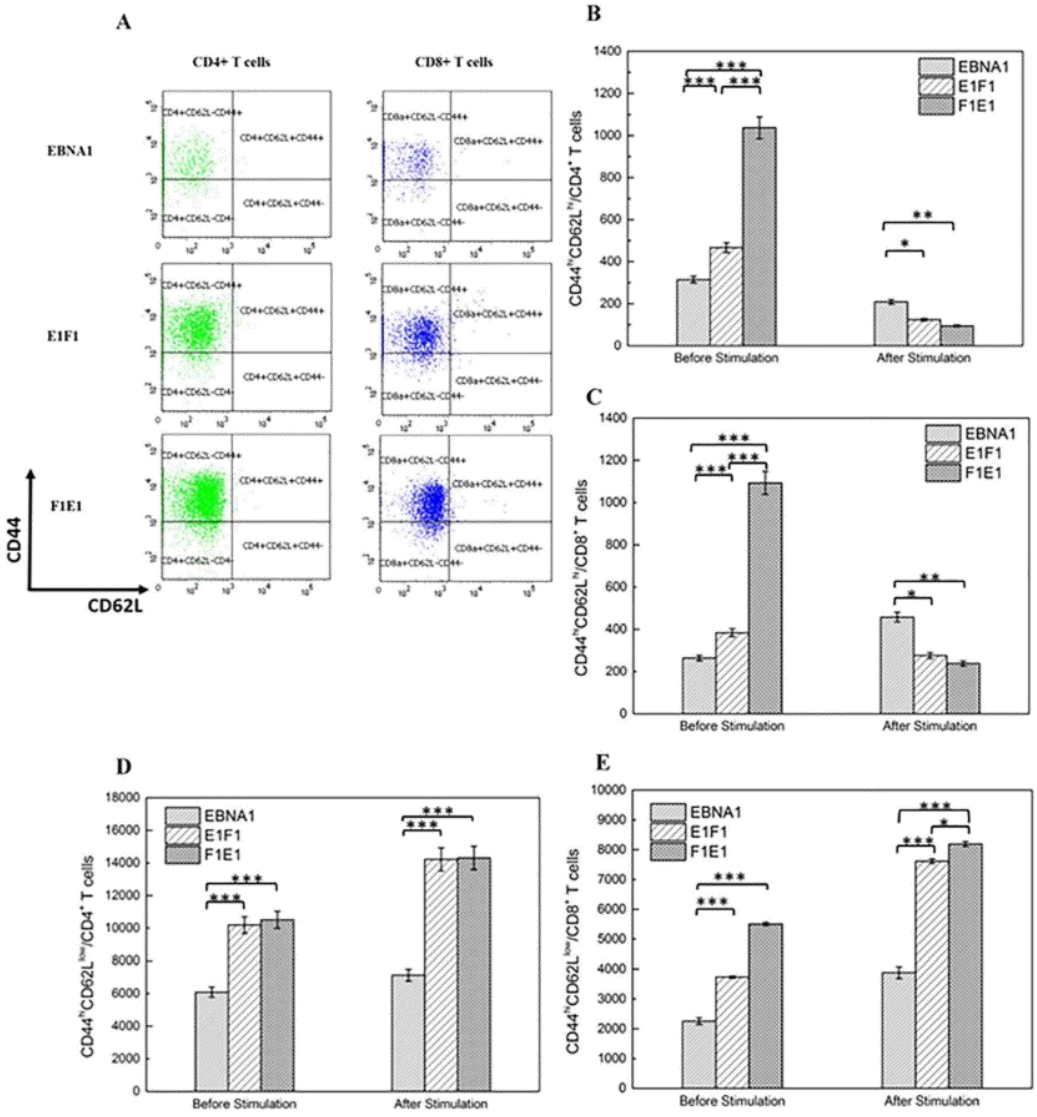


Fig. 7. Frequencies of central ($CD44^{hi}CD62L^{hi}$)/ effector ($CD44^{hi}CD62L^{low}$) memory $CD4^+$ and $CD8^+$ T cells. Mice were immunized three times. Splenocytes were harvested on day 38 after the first immunization and stimulated ex vivo with antigen EBNA1 peptide (500 μL, 20 μg/ml) for 60 h. The frequency of $CD44^{hi}$ $CD62L^{hi}/CD4^+$ T cells, $CD44^{hi}$ $CD62L^{low}/CD4^+$ T cells, $CD44^{hi}$ $CD62L^{hi}/CD8^+$ T cells, and $CD44^{hi}$ $CD62L^{low}/CD8^+$ T cells were measured by flow cytometry. FACS plots in (A) are representative of the mean percentages of six mice in each group. Data are expressed as mean SEM (n = 6) (*p < 0.05; **p < 0.01; ***p < 0.001), where n represents the number of mice from each group. (A) FACS plots. (B) Frequencies of central ($CD44^{hi}CD62L^{hi}$) memory $CD4^+$ T cells. (C) Frequencies of central ($CD44^{hi}CD62L^{hi}$) memory $CD8^+$ T cells. (D) Frequencies of effector ($CD44^{hi}CD62L^{low}$) memory $CD4^+$ T cells. (E) Frequencies of effector ($CD44^{hi}CD62L^{low}$) memory $CD8^+$ T cells.

($CD44^{hi}$ $CD62L^{low}$) [43]. It was observed that before stimulation by EBNA1 epitope, the number of $CD4^+$ and $CD8^+$ T_{CM} for mice immunized with F1E1 was >1000 cells, 2 times greater than immunization with both EBNA1 epitope (~300 cells) and E1F1 (~500 cells) (Fig. 7B and C). After stimulation by EBNA1 epitope, there was a decreasing trend of T_{CM} found for both two groups, because $CD4^+$ and $CD8^+$ T_{CM} were converted to T_{EM}. Overall, the number of T_{EM} for E1F1 and F1E1 was 2 times greater than for EBNA1 epitope (Fig. 7D and E). While E1F1 and F1E1 caused similar $CD4^+$ T_{EM} levels after stimulation (Fig. 7D), F1E1 induced 9% greater number of $CD8^+$ effector memory T cells than E1F1 (*p < 0.05) (Fig. 7E). Hence, C-terminus insertion produces a stronger cell-mediated immune response than N-terminus insertion, because greater number of T_{CM} before stimulation and T_{EM} after the stimulation were found in F1E1 than E1F1.

4. Discussion

This study presents the development of engineered ferritin vaccines against EBV, cooperation with aluminum hydroxide adjuvant. Compared to the EBNA1 short peptide without adjuvant, both N-terminus insertion and C-terminus insertion induced potent humoral and cell-mediated immune responses. The proliferative response was enhanced after applying ferritin as the vaccine carrier. This all confirmed that human ferritin heavy chain is eligible

to be treated as a vaccine carrier. This is possibly caused by the molecular weight difference and addition of adjuvant. Ferritin has uniform architecture with 24 subunits of total molecular weight around 504 kDa. The molecular weight for the short EBNA1 epitope is 1.3 kDa. Hence, the ferritin vaccine platform presenting antigens provides a larger structure than for the EBNA1 epitope structure, explaining the greater induction of humoral and cell-mediated immune responses [44]. This also shows the limitations of peptide vaccine. Peptide vaccine induces weak immune responses and is found to induce fairy weak immune responses even with adjuvant [45,46,47]. Thus, peptide vaccine require carriers to counterbalance the low-molecular nature and low efficiency.

The dominant immune response for the EBNA1 epitope was found to be different to the engineered ferritins groups. The ratio of IgG2a/IgG1 for the short peptide group was slightly greater than 1. For engineered ferritins with adjuvant, the ratio is lowered and the dominant immune response changed from cell-mediated to humoral immune response. The reason may be that, for engineered ferritin groups, aluminum hydroxide adjuvant was added to increase the vaccine efficacy and enhance the immunity [48]. Aluminum adjuvants induce Th2 responses only, which mostly activate humoral immune response, and they are unable to promote cytotoxic T cell-mediated immunity [49,50,51]. EBNA1 epitope without addition of aluminum adjuvants, can elicit Th1-type $CD4^+$ and $CD8^+$ T cells [52,53], which are mostly related to cell-mediated immune response in principle. As a result, cell-mediated immune response is dominant for EBNA1 peptide group, while humoral immune response is dominant for engineered ferritins groups.

In this study, we inserted the antigen at either N-terminus or C-terminus of ferritin to produce E1F1 and F1E1 vaccines, respectively. Structural characterization of the engineered ferritins revealed that for N-terminus insertion EBNA1 epitope was presented on the exterior surface of the ferritin, while C-terminus insertion placed the epitopes into the interior cavity of ferritin. It was found that there was no obvious difference in humoral immune responses for N-terminus insertion and C-terminus insertion. From previous research by Wang *et al.* [8], it was more likely that when epitopes were inserted outside the ferritin cage (N-terminus insertion and the flexible loop region), it was easier to induce greater antibody titers than inserting antigens inside the ferritin cage (C-terminus insertion). Nonetheless, this finding does not really accord with our findings here. We found that C-terminus insertion induces a comparable antibody titer to N-terminus insertion. It may be that many factors result in different behaviors, including different adjuvant types used and different epitopes applied. Hence, our findings suggest that presenting epitopes outside or inside the nanoparticles will not directly determine a stronger or weaker humoral immune response.

Desirable vaccines also achieve antigen-specific cellular response in addition to humoral response. Both N-terminus and C-terminus insertion produced antigen-specific cellular response after stimulation with EBNA1 epitope. Dendritic cells (DCs), which are the most important antigen-presenting cells that initiate and direct adaptive immune responses [54], orderly phagocytosed these engineered ferritin nanoparticles carrying EBNA1 epitope [10]. Following that, the processed EBNA1 epitopes within endosomes were successfully displayed on the surface of DCs with the major histocompatibility complexes (MHCs) MHCI and MHC II, and hence induced EBNA1 epitope specific $CD4^+$ and $CD8^+$ T cells via their T cell receptors (TCRs), shown in Fig. 8. C-terminus insertion showed stronger proliferative responses than N-terminus insertion. Meanwhile, from the evaluation of memory T cells, C-terminus insertion presented a high number of central memory T cells (before stimulation) and effector memory T cells (after stimulation). These results suggest that inserting epitopes at the C-terminus of ferritins may be processed more efficiently by DCs than N-terminus insertions. It is inferred that the positions of antigen and protein conformations result in different antigens-processing pathways by DCs. The possible reason may be the structural differences between E1F1 and F1E1. For N-terminus insertion, the epitopes and 15 units of soft linkers were fully presented out-

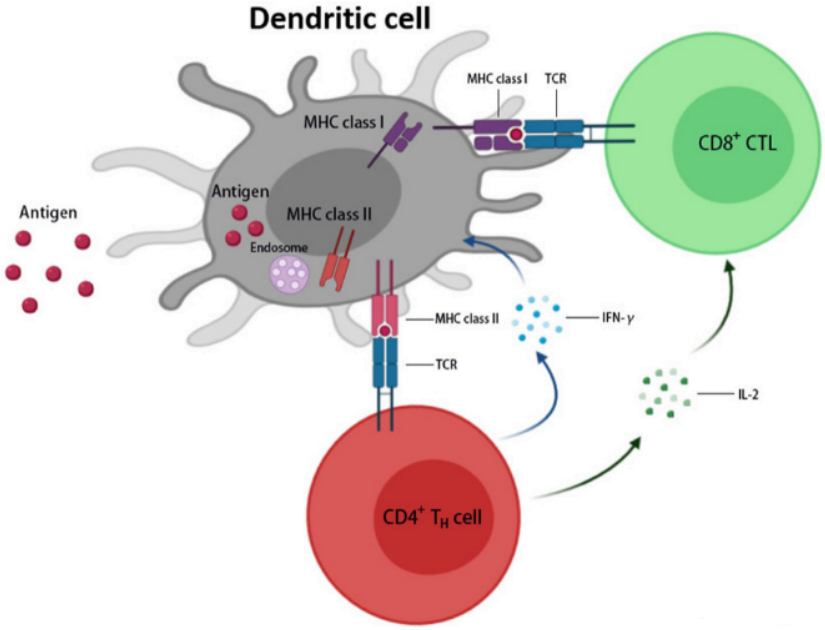


Fig. 8. Dendritic cell processing diagram by Biorender. Antigen presentation is mediated by MHC class I molecules, and the class II molecules, found on the surface of antigen-presenting cells (APCs), such as dendritic cell (DC). MHC class I and class II molecules deliver short peptides to the cell surface in a similar manner. $CD8^+$ T cells recognize peptides on the surface of MHC I molecules while $CD4^+$ T cells recognize the ones on MHC II molecules. T_H cells produce interferon gamma (IFN-γ) that is critical for innate and adaptive immunity against viral. Activated $CD4^+$ T cells and $CD8^+$ T cells are also the major sources of IL-2.

side the ferritin cage. The role of soft linkers ensures the flexibility of epitopes [55] in order to avoid unfavorable interactions between the epitopes and ferritin. On the other hand, soft linkers are not rigid structures and can be easily broken away from the protein nanocage because of changes in the surrounding environment [56]. Antigens may thus be released from ferritin. Compared to N-terminus insertion, epitopes and linkers are protected inside the ferritin cage for C-terminus insertion. On this account, the effective antigen density for C-terminus insertion would be greater than N-terminus insertion due to the protection afforded by the cage. There would be more effective antigen processed by DCs for C-terminus insertion than N-terminus insertion.

In summary, we successfully produced engineered ferritin-based nanoparticles for the prevention of EBV infections. N-terminus and C-terminus locations were selected as the insertion site for the short peptide EBNA1 epitope. N-terminus insertion and C-terminus insertion presented totally different protein structures. N-terminus insertion presented antigen outside the nanoparticle cage, while C-terminus insertion protected antigens inside the cage. This structural difference significantly affects the corresponding immunogenicity, especially in terms of cell-mediated immune response. C-terminus insertion presented stronger proliferative responses and cell-mediated immune response than N-terminus insertion. Our study will be used as a preliminary immunology investigation as it has been proven that, E1F1 and F1E1 are able to induce potent humoral and cell-mediated immune response after stimulating by the short EBNA1 peptide. Our next-step plan will further test associated immune response after stimulating by full-length EBNA1 antigen and also perform virus neutralization assay to further prove the potential of E1F1 and F1E1 in the prevention of EBV virus. In future application, the type of adjuvant and the insertion site for specific antigens in the ferritin will be extremely important to be considered to improve the immune response induced by engineered ferritin vaccines.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the University of Adelaide's DVCR-Bioprocess Facility Fund-2018. We acknowledge Shanxi Education Science "1331 project" special research project (Research and Development of Traditional Chinese Medicine Micro-emulsion and New Biological Preparation), and also thank Adelaide Microscopy Center for assisting to access the transmission electron microscopy. This research was undertaken on the MX1 beamline at the Australian Synchrotron, part of ANSTO.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2021.07.021.

References

[1] López-Sagaseta J, Malito E, Rappuoli R, Bottomley MJ. Self-assembling protein nanoparticles in the design of vaccines. Comput Struct Biotechnol J 2016;14:58–68.
[2] Zhao L, Seth A, Wibowo N, Zhao C-X, Mitter N, Yu C, et al. Nanoparticle vaccines. Vaccine 2014;32(3):327–37.
[3] Diaz D, Care A, Sunna A. Bioengineering strategies for protein-based nanoparticles. Genes (Basel). 2018;9(7):370. https://doi.org/10.3390/genes9070370.
[4] Lee B-R, Ko HK, Ryu JH, Ahn KY, Lee Y-H, Oh SJ, et al. Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. Sci Rep 2016;6(1). https://doi.org/10.1038/srep35182.
[5] Sliepen K, Ozorowski G, Burger JA, van Montfort T, Stunnenberg M, LaBranche C, et al. Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity. Retrovirology. 2015;12(1). https://doi.org/10.1186/s12977-015-0210-4.
[6] Kanekiyo M, Wei C-J, Yassine HM, McTamney PM, Boyington JC, Whittle JRR, et al. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. Nature 2013;499(7456):102–6.
[7] Li CQ, Soistman E, Carter DC. Ferritin nanoparticle technology. A new platform for antigen presentation and vaccine development. Ind Biotechnol 2006;2(2):143–7.
[8] Wang Z, Xu L, Yu H, Lv P, Lei Z, Zeng Y, et al. Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. Biomater Sci 2019;7(5):1794–800.
[9] Zhen Z, Tang W, Todd T, Xie J. Ferritins as nanoplatforms for imaging and drug delivery. Expert Opin Drug Deliv. 2014;11(12):1913–22.
[10] Han J-A, Kang YJ, Shin C, Ra J-S, Shin H-H, Hong SY, et al. Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. Nanomed Nanotechnol Biol Med 2014;10(3):561–9.
[11] Lee J-L, Song H-S, Kim H-J, Park J-H, Chung D-K, Park C-S, et al. Functional expression and production of human H-ferritin in Pichia pastoris. Biotechnol Lett 2003;25:1019–23.
[12] Lee J-H, Seo HS, Song JA, Kwon KC, Lee EJ, Kim HJ, et al. Proteinticle engineering for accurate 3D diagnosis. ACS Nano 2013;7(12):10879–86.
[13] Kwon KC, Ko HK, Lee J, Lee EJ, Kim K, Lee J. Enhanced in vivo tumor detection by active tumor cell targeting using multiple tumor receptor-binding peptides presented on genetically engineered human ferritin nanoparticles. Small. 2016;12(31):4241–53.
[14] Guo J, Xu N, Yao Y, Lin J, Li R, Li J-W. Efficient expression of recombinant human heavy chain ferritin (FTH1) with modified peptides. Protein Expr Purif 2017;131:101–8.
[15] Farrell PJ. Epstein-Barr Virus and Cancer. Annu Rev Pathol 2019;14(1):29–53.
[16] Pattle SB, Farrell PJ. The role of Epstein-Barr virus in cancer. Expert Opin Biol Ther 2006;6(11):1193–205.
[17] Shi Y, Peng S-L, Yang L-F, Chen X, Tao Y-G, Cao Ya. Co-infection of Epstein-Barr virus and human papillomavirus in human tumorigenesis. Chin. J Cancer. 2016;35(1). https://doi.org/10.1186/s40880-016-0079-1.
[18] Mesri E, Feitelson MA, Munger K. Human viral oncogenesis: A cancer hallmarks analysis. Cell Host Microbe 2014;15(3):266–82.
[19] Wilson J, Manet E, Gruffat H, Busson P, Blondel M, Fahraeus R. EBNA1: oncogenic activity, immune evasion and biochemical functions provide targets for novel therapeutic strategies against Epstein-Barr virus- associated cancers. Cancers 2018;10(4):109. https://doi.org/10.3390/cancers10040109.
[20] Destro F, Sforza F, Sicurella M, Marescotti D, Gallerani E, Baldisserotto A, et al. Proteasome inhibitors induce the presentation of an Epstein-Barr virus nuclear antigen 1-derived cytotoxic T lymphocyte epitope in Burkitt's lymphoma cells. Immunology 2011; 133: 105–14.
[21] McPhillips TM, McPhillips SE, Chiu H-J, Cohen AE, Deacon AM, Ellis PJ, et al. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 2002;9(6):401–6.
[22] Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr 2010;66(2):125–32.
[23] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011;67(4):235–42.
[24] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007;40(4):658–74.
[25] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(12):2126–32.
[26] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 2010;66(2):213–21.
[27] DeLano WL. The PyMOL Molecular Graphics System, version 1.5.0.1. New York; 2012.
[28] Butts CA, Swift J, Kang S-g, Di Costanzo L, Christianson DW, Saven JG, et al. Directing Noble Metal ion chemistry within a designed ferritin protein. Biochemistry 2008;47(48):12729–39.
[29] Malyala P, Singh M. Endotoxin limits in formulations for preclinical research. J Pharm Sci 2008;97(6):2041–4.
[30] Hammarström L, Smith CI. IgG subclass changes in response to vaccination. Monogr Allergy 1986;19:241–52.
[31] Lubeck MD, Steplewski Z, Baglia F, Klein MH, Dorrington KJ, Koprowski H. The interaction of murine IgG subclass proteins with human monocyte Fc receptors. J Immunol 1985;135:1299–304.
[32] Williams MA, Bevan MJ. Effector and memory CTL differentiation. Annu Rev Immunol 2007;25(1):171–92.
[33] Poston RN, Morgan RS. Interactions between soluble IgG, complement and cells in lymphocyte and monocyte ADCC. Immunology 1983;50:461–9.
[34] Paul WE, Seder RA. Lymphocyte responses and cytokines. Cell 1994;76(2):241–51.
[35] Datta S, Sarvetnick N. Lymphocyte proliferation in immune-mediated diseases. Trends Immunol 2009;30(9):430–8.

[36] Driesen J, Popov A, Schultze JL. CD25 as an immune regulatory molecule expressed on myeloid dendritic cells. Immunobiology 2008;213(9-10):849–58.
[37] Lederman S, Yellin MJ, Krichevsky A, Belko J, Lee JJ, Chess L. Identification of a novel surface protein on activated CD4+ T cells that induces contact-dependent B cell differentiation (help). J Exp Med 1992;175:1091–101.
[38] Hoyer S, Prommersberger S, Pfeiffer IA, Schuler-Thurner B, Schuler G, Dörrie J, et al. Concurrent interaction of DCs with CD4(+) and CD8(+) T cells improves secondary CTL expansion: It takes three to tango. Eur J Immunol 2014;44 (12):3543–59.
[39] Wang K, Wei G, Liu D. CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. Exp Hematol Oncol 2012;1(1):36. https://doi.org/10.1186/2162-3619-1-36.
[40] Krzych U, Zarling S, Pichugin A. Memory T cells maintain protracted protection against malaria. Immunol Lett 2014;161(2):189–95.
[41] Sallusto F, Geginat J, Lanzavecchia A. Central memory and effector memory T cell subsets: function, generation, and maintenance. Annu Rev Immunol 2004;22:745–63.
[42] Sckisel GD, Mirsoian A, Minnar CM, Crittenden M, Curti B, Chen JQ, et al. Differential phenotypes of memory CD4 and CD8 T cells in the spleen and peripheral tissues following immunostimulatory therapy. J ImmunoTher Cancer 2017;5(1). https://doi.org/10.1186/s40425-017-0235-4.
[43] Kaech SM, Wherry EJ, Ahmed R. Effector and memory T-cell differentiation: implications for vaccine development. Nat Rev Immunol 2002;2(4):251–62.
[44] Pati R, Shevtsov M, Sonawane A. Nanoparticle vaccines against infectious diseases 9:2224-. Front Immunol 2018;9. https://doi.org/10.3389/fimmu.2018.02224.
[45] Skwarczynski M, Toth I. Peptide-based synthetic vaccines. Chem Sci 2016;7 (2):842–54.
[46] Slingluff Jr CL. The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? Cancer J 2011;17:343–50.
[47] Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discovery 2007;6(5):404–14.
[48] HogenEsch H, O'Hagan DT, Fox CB. Optimizing the utilization of aluminum adjuvants in vaccines: you might just get what you want. npj Vaccines 2018;3 (1). https://doi.org/10.1038/s41541-018-0089-x.
[49] Bomford R, Stapleton M, Winsor S, McKnight A, Andronova T. The control of the antibody isotype response to recombinant human immunodeficiency virus gp120 Antigen by Adjuvants. AIDS Res Hum Retroviruses 1992;8 (10):1765–71.
[50] Comoy EE, Capron A, Thyphronitis G. In vivo induction of type 1 and 2 immune responses against protein antigens. Int Immunol 1997;9:523–31.
[51] Jiao X-D, Cheng S, Hu Y-H, Sun Li. Comparative study of the effects of aluminum adjuvants and Freund's incomplete adjuvant on the immune response to an Edwardsiella tarda major antigen. Vaccine 2010;28(7):1832–7.
[52] Bickham K, Münz C, Tsang ML, Larsson M, Fonteneau J-F, Bhardwaj N, et al. EBNA1-specific CD4+ T cells in healthy carriers of Epstein-Barr virus are primarily Th1 in function. J Clin Invest 2001;107(1):121–30.
[53] Münz C. Epstein-barr virus nuclear antigen 1: from immunologically invisible to a promising T cell target. J Exp Med 2004;199:1301–4.
[54] Cohn L, Delamarre L. Dendritic cell-targeted vaccines 5:255-. Front Immunol. 2014;5. https://doi.org/10.3389/fimmu.2014.00255.
[55] Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The role of protein loops and linkers in conformational dynamics and allostery. Chem Rev 2016;116(11):6391–423.
[56] Wriggers W, Chakravarty S, Jennings PA. Control of protein functional dynamics by peptide linkers. Peptide Sci 2005; 80: 736–46.

CHAPTER 7 IMMUNOGENICITY AND VACCINE EFFICACY BOOSTED BY ENGINEERING HUMAN HEAVY CHAIN FERRITIN AND CHIMERIC HEPATITIS B VIRUS CORE NANOPARTICLES

Statement of Authorship

Title of Paper	Immunogenicity and vaccine efficacy boosted by engineering human heavy chain ferritin and chimeric hepatitis B virus core nanoparticles
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Qu, Y.; Zhang, B.; Wang, Y.; Yin, S.; Sun, Y.; Middelberg, A.; Bi, J. Immunogenicity and Vaccine Efficacy Boosted by Engineering Human Heavy Chain Ferritin and Chimeric Hepatitis B Virus Core Nanoparticles. *ACS Applied Bio Materials* 2021, 4, 7147-7156.

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Designed experiment, performed experiments; analysed data; writing manuscript		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	09/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Bingyang Zhang		
Contribution to the Paper	Designed experiment, performed experiment		
Signature		Date	09/09/2021

Name of Co-Author	Yingli Wang		
Contribution to the Paper	Designed experiment, performed experiment, analysed data		
Signature		Date	09/09/2021

Name of Co-Author	Shuang Yin		
Contribution to the Paper	Designed experiment		
Signature		Date	09/09//2021

Name of Co-Author	Yan Sun		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	20/09/2021

Name of Co-Author	Anton Middelberg		
Contribution to the Paper	Experiment design, proof read the manuscript & contribution to ferritin		
Signature		Date	13/10/21

Name of Co-Author	Jingxiu Bi		
Contribution to the Paper	Experiment design, proof read the manuscript		
Signature		Date	12/10/2021

Immunogenicity and vaccine efficacy boosted by engineering human heavy chain ferritin and chimeric hepatitis B virus core nanoparticles

Yiran Qu, Bingyang Zhang, Yingli Wang, Yan Sun, Anton Middelberg, Jingxiu Bi*

Y. Qu, S. Yin, B. Zhang, A/Prof. Dr. J. Bi, Prof. Dr. A. Middelberg

School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide
South Australia 5005 (Australia)

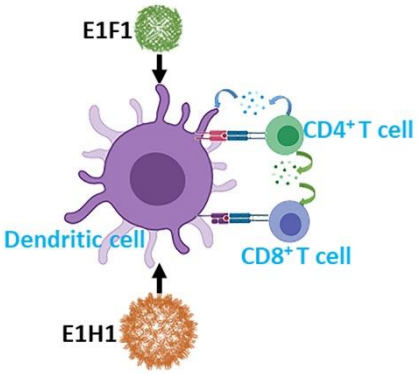
Prof. Dr.Y. Wang
Shanxi University of Traditional Chinese Medicine, Shanxi, China

Prof. Dr. Y. Sun
Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072 (China)

ACS Applied Bio Materials, doi:10.1021/acsabm.1c00738.

Abstract

Human heavy-chain ferritin (HFn) and hepatitis B virus core (HBc) are both nanoparticle proteins presenting well-oriented architecture with constant size and shape, which can be engineered to carry epitopes on the surface of nanoparticle protein cage enabling vaccine design. This study aims to investigate the immunogenicity differences between engineered HFn and chimeric HBc bearing the same epitope. As proof of concept, the model epitope Epstein-Barr nuclear antigen 1 (EBNA1) is inserted at N-terminus of HFn and HBc subunit to produce two vaccine candidates named EBNA1-HFn (E1F1) and EBNA1-HBc (E1H1), respectively. From *in vivo* immunogenicity studies, E1H1 demonstrates the capability to prompt significant humoral and cell-mediated immune response in adjuvant-free formulation. When formulated with aluminum hydroxide adjuvant, E1H1 produces approximately 5× higher titer and 2× stronger proliferative index (PI) than E1F1. These results confirm that the HBc carrier induces a stronger humoral immune response than HFn. On the other hand, from lymphocyte activation experiments, E1F1 induces stronger cell-mediated immune response indicated by 5× more $CD8^+$ T cells and 2× more effector memory T cells in the E1F1 group versus the E1H1 group. Through this study, HFn and HBc are shown to be potentially effective vaccine carrier nanoparticles having subtly different immunological responses.



KEYWORDS: Protein-based vaccine molecule design, hepatitis B virus core, virus like particles, human ferritin heavy chain, Epstein-Barr nuclear antigen 1

7.1 Introduction

The current COVID-19 pandemic has increased urgency for new-engineered reliable vaccines. It has increased the need to develop vaccine platforms meeting criteria including fast manufacture at large scale in response to a pandemic threat and high safety yet will optimal efficacy and efficiency to generate potent and protective immune responses. There is considerable interest in applying recombinant protein nanoparticles as a vaccine delivery system because (i) they can be produced at large scale via cost-effective expression systems such as bacteria [1,2]; (ii) protein nanoparticles are regarded as safe with minimal risks compared to traditional vaccines such as killed or inactivated pathogen vaccines [3]; (iii) nanoparticle proteins are highly repetitive structures that can be developed as efficacious and efficient vaccines to induce potent immune responses [4-6]. There is a wide selection of sizes (5 to 100 nm) [7]. Ferritins and virus like particles (VLPs) are two common examples of nanoparticle proteins.

Ferritins are self-assembling nanoparticle proteins with 24 protein subunits that form spherical cage structures [8]. The inner and outer diameters for a ferritin particle are 8 nm and 12 nm, respectively [9]. Ferritin has recently been explored as a vaccine carrier due to its various advantages including robust thermal and chemical stabilities [10,11], suitable presentation of antigens and presentation of immunogens [10,12-14], low toxicity, high biocompatibility and good biodegradability [11,15]. Recombinant human heavy-chain ferritin (HFn) is one of the well-understood ferritins [16]. The molecular weight of each subunit is about 21 kDa. It was found by Broxmeyer *et al*. that human ferritin heavy chain suppressed immune activity in human *in vivo*[17,18]. They also noticed that ferritin heavy chain notably reduced growth of granulocyte-macrophage, erythroid and multipotential progenitor cells [17-19]. On this account, it is a common

practice that engineered ferritin-based vaccines are often administrated with adjuvants to boost potential immune responses [10,12-14].

Virus-like particles (VLPs) are constructed by viral capsid proteins assembled to a hollow nanocage lacking infectious nucleic acids [4,8]. VLPs are applicable vaccine carriers as they present antigens in a multi-meric and well-organized manner, which induces desirable humoral and cell-mediated immune responses [4,5,20-22]. One of the well-characterized VLPs is the hepatitis B virus core protein (HBc). HBc is made up of 180 equal subunits (T=3) with outer diameters of 27-31 nm [4,23]. There is also a proportion of HBc comprising 240 equal subunits (T=4) with an external diameter of 35 nm [24]. The molecular weight for each subunit is approximately 21 kDa. Because of its strong immunity and extreme flexibility to insert peptide or epitope with up to 300 residues with negligible effects on assembled structure [25-27], HBc is highly promising as a vaccine carrier.

Currently, research directly comparing the immune responses induced by HFn and HBc as vaccine carriers bearing the same epitope has not been reported. Herein, we applied Epstein-Barr nuclear antigen 1 (EBNA1) as the model epitope to develop two vaccine candidates with different molecular designs targeting Epstein-Barr virus (EBV). EBNA1 can be expressed in all EBV-associated tumours to support maintenance and expression of EBV [28]. EBV, belonging to the human herpesviruses family [28], result in serious disease (e.g. Hodgkin's disease, lymphomas and lymphoproliferative disease) [29]. There is no current licensed vaccine available to combat EBV infection [28]. There is some urgency to develop EBV-associated vaccines for controlling EBV infection.

To reveal differences between HFn and HBc as vaccine carriers, short EBNA1 peptide was inserted at the N-terminus of HFn and HBc. The resulting vaccine are candidates named EBNA1-HFn (E1F1) and EBNA1-HBc (E1H1). Before *in-vivo* immunization test, molecular structure was characterized by SDS-PAGE, SEC-MALS and TEM to confirm the purity of E1H1 (> 95 %) and E1F1 (> 95 %), hydrodynamic radius and molecular weight (E1H1: 17.4 nm and 4740 kDa; E1F1: 7.3 nm and 562 kDa). After immunization in mice with (E1H1 and E1F1) and without (E1H1) adjuvant, immune response (humoral and cell-mediated), proliferative index between E1H1 and E1F1 were compared. *In-vivo* experiment demonstrates HBc VLP carrier can induce potent immune response in adjuvant-free formulation. Boosted by adjuvant, E1H1 induces greater humoral immune and proliferative response than E1F1. However, E1F1 induces greater cell-mediated immune response than E1H1. This paper sheds new light on chimeric protein as vaccine carriers to boost immunogenicity and vaccine efficacy, which can be produced at large scale in a short period via cost-effective process to against pandemic infection.

7.2 Experimental Section

7.2.1 Nanoparticle proteins expression

The detailed information about plasmid construction for EBNA1-ferritin (E1F1) can be found in our previous research [11]. For EBNA1-HBc (E1H1), plasmid was built using pET 30a vector. The restriction sites were *Nde I* and *BamH I* [11]. Information about protein sequence is shown in Table S7.1, Supporting Information. *E.coli* BL21 (DE3) (Invitrogen, USA) was used to transform plasmids. Epitope short EBNA1 peptide (HPVGEADYFEY) (taken from EBNA1 aa 407 to aa 417), was inserted at N-terminus of either HFn or HBc. Single colony expressing each protein was incubated in lysogeny broth (LB) medium with antibiotic kanamycin (1 mM). The growing process was at 37 °C for 14-16 h in a shaking incubator at 180 rpm. In every 500 mL LB-kanamycin (100 μg mL^{-1}) medium, 10 ml of cultured medium from 50 mL was added to further grow bacteria (37 °C, 200 rpm, 4 h, OD_{600} = 0.8), then 1 mM IPTG was added to express proteins (37 °C, 200 rpm, 4 h). Bacterial solution was centrifuged to collect bacteria pellet. The sonication followed the procedure [11] with lysis buffer (E1F1: 20 mM phosphate buffer (PB), 2 mM EDTA, pH 7.0; E1H1: 20 mM Tris-HCl, 3 mM EDTA, 1 mM Phenylmethanesulfonyl fluoride (PMSF), 0.1 % Triton X100, pH 8.0). Soluble expressed protein was collected by centrifuging bacterial lysate at 13,751 xg for 30 min. SDS-PAGE (12 %) was conducted to check soluble expression level for E1F1 and E1H1.

7.2.2 Purification of recombinant E1F1 and recombinant E1H1

For purification of E1F1, supernatant of cell lysate containing soluble expressed protein was diluted to protein concentration to 4 mg mL^{-1} with Mili-Q water [11]. The pH of diluted supernatant was altered to 5.0 utilizing 1 M NaCl, NaAc-HAc (1M, pH 5.0). Heating (60 °C, 10 min) was applied to precipitate protein impurities, following by centrifugation (19,802 xg,

4°C, 10 min). Resultant supernatant was mixed with an ammonium sulfate (AS) solution (3M), and a phosphate buffer (1M) to final AS concentration 1 M at pH 6.5. Hydrophobic interaction chromatography (HIC) was applied to separate nucleic acid. A HiTrap Octyl FF column (GE Healthcare Life Sciences) was used for E1F1 purification by binding process and AKTA Pure (GE Healthcare Life Sciences) was adopted. The column was firstly equilibrated by 0.1 M PB with 1 M AS, pH 6.5. After discarding the flow through fraction, target protein E1H1 was directly eluted by 100 % elution buffer (20 mM PB, pH 6.5). The collected fractions from the HIC column were then injected into a HiTrap Desalting column (GE Healthcare Life Sciences) to buffer exchange into 20 mM PB, pH 7.0 for further use.

In terms of E1H1 purification, different purification steps were applied. Briefly, supernatant containing soluble E1H1 of cell lysate was mixed with 1 M ammonium sulfate and equilibrated for 30 min at 25 °C, pH 7.4. The mixture was centrifuged (19,802 xg, 4 °C, 10 min) to harvest E1H1. Pellets of E1H1 were collected and dissolved in 20 mM Tris-HCl, 4 M Urea, pH 9.0. Dissolved E1H1 was then dialyzed against 20 mM PB, pH 7.0 for further use. Purity of E1F1 and E1H1 was analysed using 12% reducing SDS-PAGE.

7.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Expression level of soluble protein and purity after purification were determined by SDS-PAGE. The procedure followed our previous research [11]. Briefly, diluted samples (1 mg mL^{-1}) mixing with loading buffer [11] were disassembled by heating (100 °C, 10 min). Lanes were separated by electrophoresis (Bio-Rad, USA). Stain the gel for 1 h and wash the background using washing buffer [11].

7.2.4 Transmission electron microscopy (TEM)

TEM was utilized to characterize conformations of purified E1F1 and E1H1. Purified E1F1 and E1H1 were diluted to 0.1 mg mL^{-1}. Protein samples (10 μL) were added on grids (carbon-coated copper) and settled for 5 min. Grids were then cleaned by RO water and negatively stained for 2 min [11]. Grids were analysed with TEM (Philips CM100, 100 kV)(Field Electron and Ion Company, USA).

7.2.5 Size exclusion chromatography multi-angle light scattering (SEC-MALS)

SEC-MALS was demonstrated to detect hydrodynamic radius and molecular weights of purified proteins. For E1F1 characterization, Superose 6 Increase 10/300 GL (GE Healthcare, USA) was connected to High Performance Liquid Chromatography (HPLC) (Shimazdu, Japan) coupled with a multi-angle laser light scattering (MALS) Wyatt DAWN® HELEOS II and Optilab T-rEx (Wyatt Technology, Santa Barbara, CA USA) [11]. The equilibration buffer was 20 mM PB, pH 7.0. For E1H1 characterization, TSKgel G4000SWXL (TOSOH Bioscience, Japan) was attached to HPLC-MALS. The equilibration buffer was 20 mM PB, pH 7.4. The loading volume was 50 µl and minimum protein concentration was: E1F1=5 mg mL^{-1} and E1H1=1 mg mL^{-1}, respectively. The signal intensity at 280 nm versus retention time was plotted, and protein hydrodynamic radius and molecular weight were analysed by ASTRA software (v. 6.1).

7.2.6 Adjuvant adsorption percentage determination

Purified E1F1 and E1H1 were diluted with Mili-Q water to 1mg mL^{-1}. Aluminum hydroxide adjuvant (750 µL, 0.8 mg mL^{-1}, SERVA, Sweden) was mixed with 750 µL diluted protein samples (E1F1 and E1H1), and were stored at room temperature (25 °C) for 24 hours. After incubation, protein-adjuvant mixtures were centrifuged at 2100g for 20 min. The supernatants were collected to measure protein concentration by Bradford method (Bio-Rad, USA). The adjuvant adsorption (%) was calculated using following equation (7.1).

$$\text{Adjuvant adsorption (\%)} = (M_1-M_2)/M_1 \times 100\% \quad (7.1)$$

M_1: mass of protein before adding adjuvant

M_2: mass of protein in the supernatant after adding adjuvant and centrifugation

7.2.7 Particle size measurement

Mastersizer 2000 (Malvern, USA) was used to measure the particle size for aluminum hydroxide adjuvant (10 mg mL^{-1}), E1F1 with adjuvant (10 mg mL^{-1}), and E1H1 with adjuvant (5 mg mL^{-1}). The measurement options were set as: *Refractive index: 1.66; Adsorption: 0.01*. For each sample, 5 mL was added to make detectable particle size measurements. Each sample was measured for three times. All measurements were analysed by Mastersizer software (Malvern, USA).

7.2.8 Animal immunization

Animal immunization experiments were planned and undertaken by authors in China and were approved by the Medical Ethics Committee of ShanXi University of Traditional Chinese Medicine (Approval Number 2019LL137).

Sample groups (n=10/treatment) of 6-8-week-old female BALB/c mice were immunized via intraperitoneal injection applied into the peritoneum [30] (Group 1: 100 μg purified E1F1 (1 mg mL^{-1}) in PBS with 200 μL adjuvant (0.8 mg mL^{-1}); Group 2: 100 μg purified E1H1 (1 mg mL^{-1}) in PBS; Group 3: 100 μg purified E1H1 (1 mg mL^{-1}) in PBS with 200 μL adjuvant (0.8 mg mL^{-1}). All three samples were injected with equivalent 6 μg of EBNA1 epitope. The detailed calculation was shown in supporting information S3. Another two control groups (n=10/treatment) were injected with 200 μL PBS (Blank) and 100 μg of Ovalbumin (OVA) [30], respectively. Serum was collected 10 days after the first (at day 24) and the second boost (at day 38) and stored at -80 °C before further measurement.

7.2.9 Enzyme-linked immunosorbent assay (ELISA) and antigen-specific IgG isotype

The detailed procedure followed our published research [30]. Short EBNA1 peptide (HPVGEADYFEY) was adsorbed for 14 h at 4 °C in 96-well plates [30]. After being blocked with serum at 37 °C for 2 h, the plate was added with mouse sera of serial dilutions collected before (Day 24 and Day 38). The plate was further incubated at 37 °C for 2vh. After washing with PBS for three times, 200 μL goat anti-mouse IgG horseradish peroxidase-conjugated antibody of 1:5000 dilution was added to the plate and incubated at 37 °C for 2 h [30]. Colour was developed by 3,3′,5,5′-Tetramethylbenzidine (TMB) Single-Component Substrate solution for 10 min. Sulphuric acid (1 M) was utilized to stop this reaction. Microplate reader (Perlong, China) (450 nm) was used to examine plates.

Steps prior to plates washing were similar to ELISA [30]. Antibodies goat anti-mouse IgG1 and IgG2a (1:1000 dilution, 100 μL) were loaded to each plate [30]. 200 μL of rabbit anti-goat IgG horseradish peroxidase-conjugated antibody (1:5000 dilution) was loaded to the plate. The following steps were same as ELISA to develop colour, stop reaction and plates analysis.

7.2.10 T-cell proliferation assay

Single-cell suspensions (4 x 10^6 cells mL^{-1}) were cultured in RPMI 1640 (Gibco, Germany) with antigen (EBNA1: 20 μg mL^{-1}), and incubated at 37 °C for 2 days. CCK-8 (10 μl) (Dojindo, Japan) was added per well (37 °C, 4 h). Microplate reader (Perlong, China) was used to analyse results. The proliferation indexes (PI) was calculated by equations 7.2[30] and 7.3[30]:

$$\text{Proliferation rate} = \frac{\text{A(sample)} - \text{A}_0\text{(control)}}{\text{A}_0\text{(control)}} \quad (7.2)$$

A (sample): A450 nm of the sample group with CCK-8 and stimulator (short EBNA1 peptide)

A_0 (control): A450 nm of the control group with CCK-8 only

$$\text{Proliferation Index} = \frac{\text{Proliferation rate (sample)}}{\text{Proliferation rate (PBS)}} \quad (7.3)$$

7.2.11 Lymphocyte activation and memory T-cells of splenocytes

The procedure to culture splenocytes (10^6 cells) from immunized mice followed referenced procedure [31]. Splenocytes were stained with fluorescent-labelled anti-mouse antibodies: CD4, CD8a (PerCP-Cy 5.5) monoclonal antibody, CD19 (eFluor 780), CD25, CD44 (PE-Cyanine 7) and CD62L (L-Selectin) monoclonal antibodies (Thermo fisher, USA) at 4 °C [30]. Number of cells for CD25/CD4 T, CD25/CD8 T, CD25/CD19 B, activated lymphocytes ($CD69^+$), central memory T-cells ($CD44^{hi}CD62L^{hi}$), and effector memory T-cells ($CD44^{hi}CD62L^{low}$) was measured by a FACSCanto II flow cytometry [30,31]. After flow cytometry was calibrated using standard BD beads, gates were circled to locate $CD4^+$ T, $CD8^+$ T and $CD19^+$ B cells first. CD25 was circled in $CD4^+$ T, $CD8^+$ T and $CD19^+$ B cells to determine lymphocyte activation. Cross gates of CD44 and CD62L in $CD4^+$ T, $CD8^+$ T and $CD19^+$ B cells were drawn to detect memory T cells. FACSCanto II software was used to demonstrate data analysis.

7.2.12 Statistics

Data were shown in Mean ± Standard deviation (SD).[30] Statistics was analysed by GraphPad Prism (2021). P-value was calculated using One-way ANOVA method (GraphPad Prism 2021), where *$p<0.05$, **$p<0.01$ and ***$p<0.001$.

7.3 Results and discussion

7.3.1 Characterization of E1F1 and E1H1

Nanoparticle protein characterization is important to confirm that key quality criteria including good purity and quaternary structure are satisfied prior to biological experiments. As shown by Figure 7.1A, both E1F1 and E1H1 achieved purity above 95 %. The molecular weight of E1F1 monomer by SDS- PAGE (Figure 7.1A) is around 24 kDa (lane 3), slightly larger than E1H1 (20 kDa) (lane 6). Based on the molecular weight prediction server ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss), the theoretical molecular weight for E1F1 monomer and E1H1 monomer were 23.3 kDa and 22.4 kDa, respectively, consistent with experiment. Figure 7.1B portrays the assembled structures for HFn and HBc carrier. Both assembled HFn and HBc show particulate structure; HBc comprises more subunits than HFn and therefore has a larger structure. Monomer structures for E1F1 and E1H1 are also displayed to portray how antigens were inserted. Model epitope EBNA1 was inserted at the N-terminus of HFn and HBc monomers (Figure 7.1B). Hence, for both HFn and HBc, epitopes were exposed on the surface of the protein nanoparticles. Compared to HBc monomer, an extra 15-residues $(GGSGG)_3$ of soft linker was inserted for E1F1 between the epitope and ferritin monomer to confer flexibility. As shown in Figures 7.1C and 7.1D, the molecular weight and hydrodynamic radius measured by SEC-MALS for assembled E1F1 nanoparticles were 562 kDa and 7.3 nm, comparing with 4740 kDa and 17.4 nm for assembled E1H1 (T=3). E1H1 is 8 times heavier and 2 times larger than E1F1. Theoretical assembled E1H1 (T=3) and E1F1 molecular weights are 4032 kDa and 559.2 kDa, respectively. The difference between measured data and theoretical prediction is likely due to error in the MALS measurement. The reported particle diameter for HBc and HFn are 35 nm and 12 nm, respectively. The measured sizes show reasonable correlation with these (E1H1 diameter: 33.8 nm; E1F1: 14.6 nm). Lastly, TEM images confirmed that both E1F1 and E1H1 were successfully assembled and confirmed

the size difference, which is consistent with the hydrodynamic radius result (Figures 7.1E and 7.1F). Overall, characterization confirmed that the protein nanoparticles exhibited high quality, were properly formed, and were virus-like particles rather than merely virus-sized aggregates, giving confidence that biological results would be meaningful.

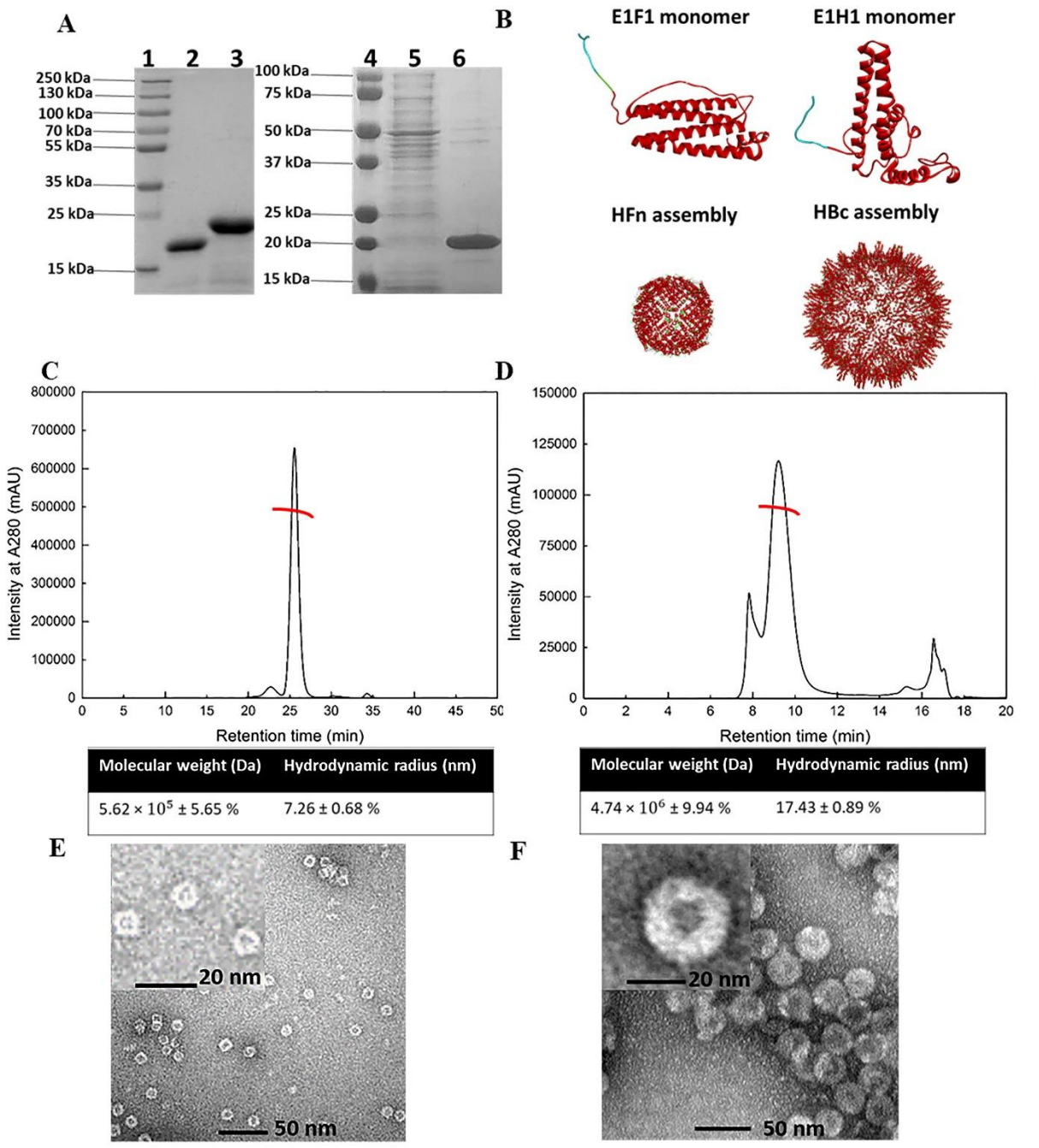


Figure 7.1 Characterization of E1F1 (ferritin carrier) and E1H1 (HBc carrier). A) Coomassie blue stained sodium dodecyl sulfate (SDS): lane 1, protein marker 1; lane 2, HFn marker (~21 kDa); lane 3, E1F1 (~24 kDa); lane 4, protein marker 2; lane 5, impurities (N/A); lane 6: E1H1 (~20 kDa). B) Mimic diagrams by Discovery Studio 2020. Red denotes carrier, blue denotes EBNA1 short peptide epitope, and green denotes soft linkers $(GGSGG)_3$ (i) E1F1 monomer; (ii) E1H1 monomer; (iii) assembled HFn; (iv) assembled HBc. C) Molar weight and particle size of assembled E1F1 measured by SEC-MALS; D) Molar weight and particle size of assembled E1H1 measured by SEC-MALS; E) Assembled E1F1 TEM image; F) Assembled E1H1 TEM image.

7.3.2 Antigen-specific antibodies titers evaluation

Sera IgG responses to short EBNA1 peptide were determined to compare humoral immune response induced by E1F1 with that caused by E1H1. It was clearly observed from Figure 7.2A that, after the second boost (Day 38), both E1H1 with and without adjuvant produced 5 times higher titers than E1F1 with adjuvant (***$p<0.001$). In addition, the minor difference between E1H1 with and without adjuvant (*$p<0.5$) proved that non-adjuvanted E1H1 can induce a potent humoral immune response. This result reveals the advantage of HBc VLP as a potential vaccine carrier able to achieve a stronger humoral immune response than that for a ferritin vaccine carrier.

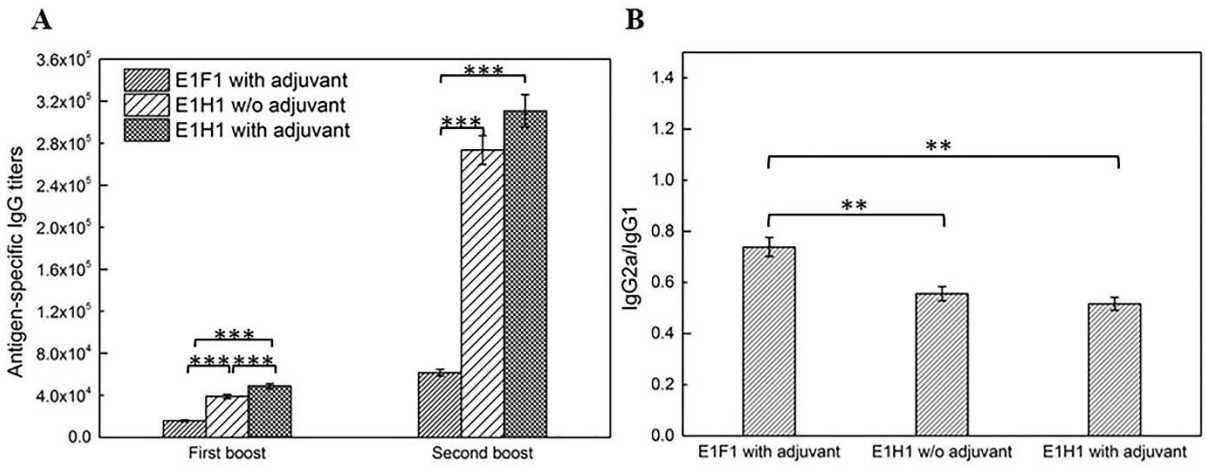


Figure 7.2 EBNA1-specific antibodies in sera of BALB/c mice. A) IgG titers 10 days after 1st boost (Day 24) and 2nd boost (Day 38). B) IgG2a/IgG1 at Day 38. Data are shown in mean ± SEM (n=3) (*$p<0.05$; **$p<0.01$; ***$p<0.001$).

To identify whether immunization response was dominant in humoral or cell-mediated, IgG2a and IgG1 were compared [32]. IgG2a/IgG1 is an indicator, which defines that: Th1-favored immunization pathway if IgG2a/IgG1 < 1; and/or Th2-biased response if IgG2a/IgG1 > 1 [33]. The Th1 response involves secretion of cytokines, for example, IL-12, TNF-α and IFN-γ. These cytokines initiate cell-mediated response for growing cytotoxic T lymphocyte (CTL) cells to

combat viral infections [30,34]. The Th2 response regulates cytokines (IL-4, IL-5 and IL-6). These cytokines are important to initiate humoral immune responses [30,35]. From Figure 7.2B, IgG2a/IgG1 ratios for all three groups (E1F1 with adjuvant, E1H1 with or without adjuvant), were lower than 1, indicating a humoral immune response bias for all three groups.

7.3.3 Comparisons of splenocytes proliferative responses

Proliferative responses induced by E1F1 and E1H1 were determined by splenocyte proliferation assay. The level of splenocyte proliferative responses by using proliferative index (PI) can reflect the strength of immune response induced by E1H1 or E1F1 [36]. The individual PI for E1F1 with adjuvant and for E1H1 with or without adjuvant were 2.0, 3.9 and 3.1 (Figure 7.3). It was also found that E1H1 with or without adjuvant induced a higher PI than E1F1 with adjuvant (**p<0.01). PI of E1H1 without adjuvant is slightly lower than E1H1 with adjuvant.

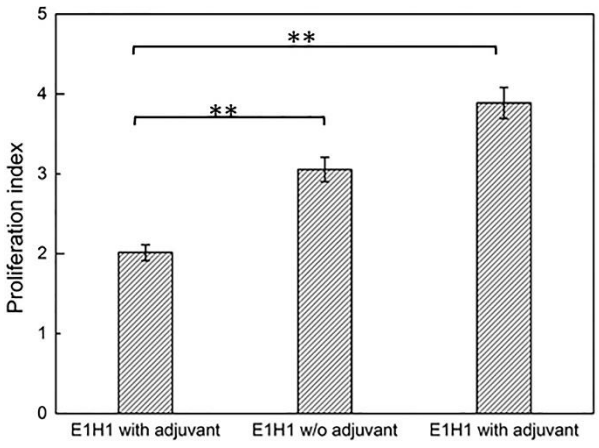


Figure 7.3 Proliferation index for E1H1 with adjuvant, E1H1 w/o adjuvant and E1H1 with adjuvant measured by CCK-8 kit. Splenocytes were cultured at Day 38 (10 days after 2nd boost) and restimulated *ex vivo* with EBNA1 antigen. Data are shown in mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001).

7.3.4 Lymphocyte activation

To compare the strength of humoral with cell-mediated immune responses induced by E1F1 and E1H1, CD25 activation level on $CD19^+$B cells, $CD4^+$T cells, and $CD8^+$T cells was found [37]. CD25 is a premature marker [37], to examine B and T cell activation level after stimulation by short EBNA1 [30]. CD19, as a B cell biomarker, is significant to regulate humoral responses [38]. The primary role for $CD4^+$T cells, T helper cells, is to drive on adaptive immune responses [39]. Figure 7.4A showed that $CD4^+$T cells identify peptides displayed on major histocompatibility complex (MHC) class II molecules, from antigen presenting cells (APCs) [40]. Dendritic cells (DCs) are examples of APCs, which activate and control adaptive immune responses [41]. Compared with $CD4^+$T cells, $CD8^+$T cells mainly communicate peptides presented by MHC I class molecules [40]. After recognition and activation by EBNA1, $CD8^+$T cells secret cytokines such as TNF-α and IFN-γ correlated to the cell-mediated immune response [30]. Overall, the number of $CD19^+$B cells reflects the strength of the humoral immune response, while $CD8^+$T cell number indicates the strength of the cell-mediated immune response.

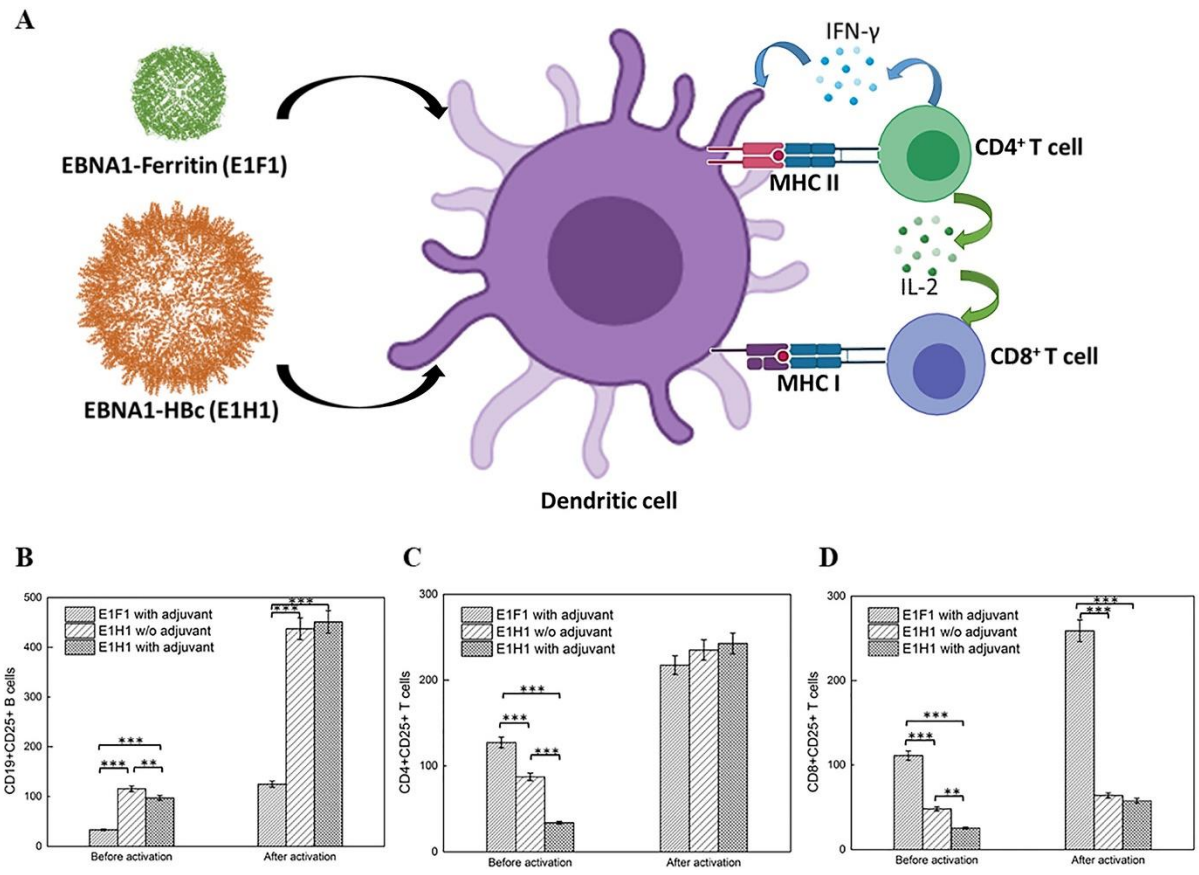


Figure 7.4 Activation of dendritic cell by EBNA1 antigens determined by flow cytometry. A) Mimic dendritic cell processing diagram by Biorender©. B) Frequency of CD25+ in CD19+ B cells. C) CD25+ in CD4+ T cells. D) CD25+ in CD8+ T cells. Data are shown in mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001).

In accordance with IgG titer results, cellular level of $CD19^{+}$B cells for E1H1 with adjuvant were the highest (Figure 7.4B), approximately 4 times higher than E1F1 with adjuvant. There was no significant difference between E1H1 with and without adjuvant regarding the number of $CD19^{+}$B cells after activation. This result suggests E1H1 can induce a potent humoral immune response by itself without adjuvant boost. Number of $CD4^{+}$T cells in three groups was similar (Figure 7.4C). $CD8^{+}$T cell number after activation for E1F1 was roughly 5 times higher than the other two groups (Figure 7.4D). This result suggests E1F1 with adjuvant produces the strongest cell-mediated immune response.

7.3.5 Memory lymphocyte T cells differentiation

To test if E1H1 and E1F1 have the ability to confer long-term protection, memory lymphocyte T cells were further determined [42]. Memory T cells can be classified as effector T cells (T_{EM}) and central memory T cells (T_{CM}) [43]. T_{EM} cells work faster than T_{CM} cells, nevertheless, the effect period is shorter. Marked by CD44 and CD62L, T_{EM} cell is represented as $CD44^{hi}$ $CD62L^{low}$, whereas T_{CM} cell is represented as $CD44^{hi}$ $CD62L^{hi}$.[44] Figure 7.5A showed the flow cytometry cross gates for CD44 and CD62L. The detailed gating information is shown in Figure S7.1 (Supporting Information). Before the stimulation by the short peptide EBNA1, the number of $CD4^+$ and $CD8^+$ T_{CM} cells in E1F1 group was 2 times higher than E1H1 with or without adjuvant groups (HBc carrier) (Figure 7.5B and 7.5C). After the stimulation by short EBNA1 peptide, the number of T_{EM} cells increased as $CD4^+$ and $CD8^+$ T_{CM} cells were converted to T_{EM} cells. There were approximately 8 % greater as many $CD4^+$ T_{EM} cells found in the E1F1 compared with E1H1 (***$p < 0.001$) (Figures 7.5D). It was found that approximately 5 % and 3 % greater as many $CD8^+$ T_{EM} cells in E1F1 group than E1H1 w/o adjuvant and E1H1 with adjuvant, respectively (***$p < 0.001$) (Figure 7.5E). The percentage of $CD4^+$ and $CD8^+$ T_{CM} cells were significantly low (<1 %) compared with $CD4^+$ and $CD8^+$ T_{EM} cells (>10 %). These results were considered as valuable by comparing these with the research using similar immunology study methods while humoral immune response was dominant (T_{CM} cells: ~1 %; T_{EM} cells: ~10 %) [31]. Based on the above results, E1F1 induces a stronger cell-mediated immune response than E1H1 with or without adjuvant.

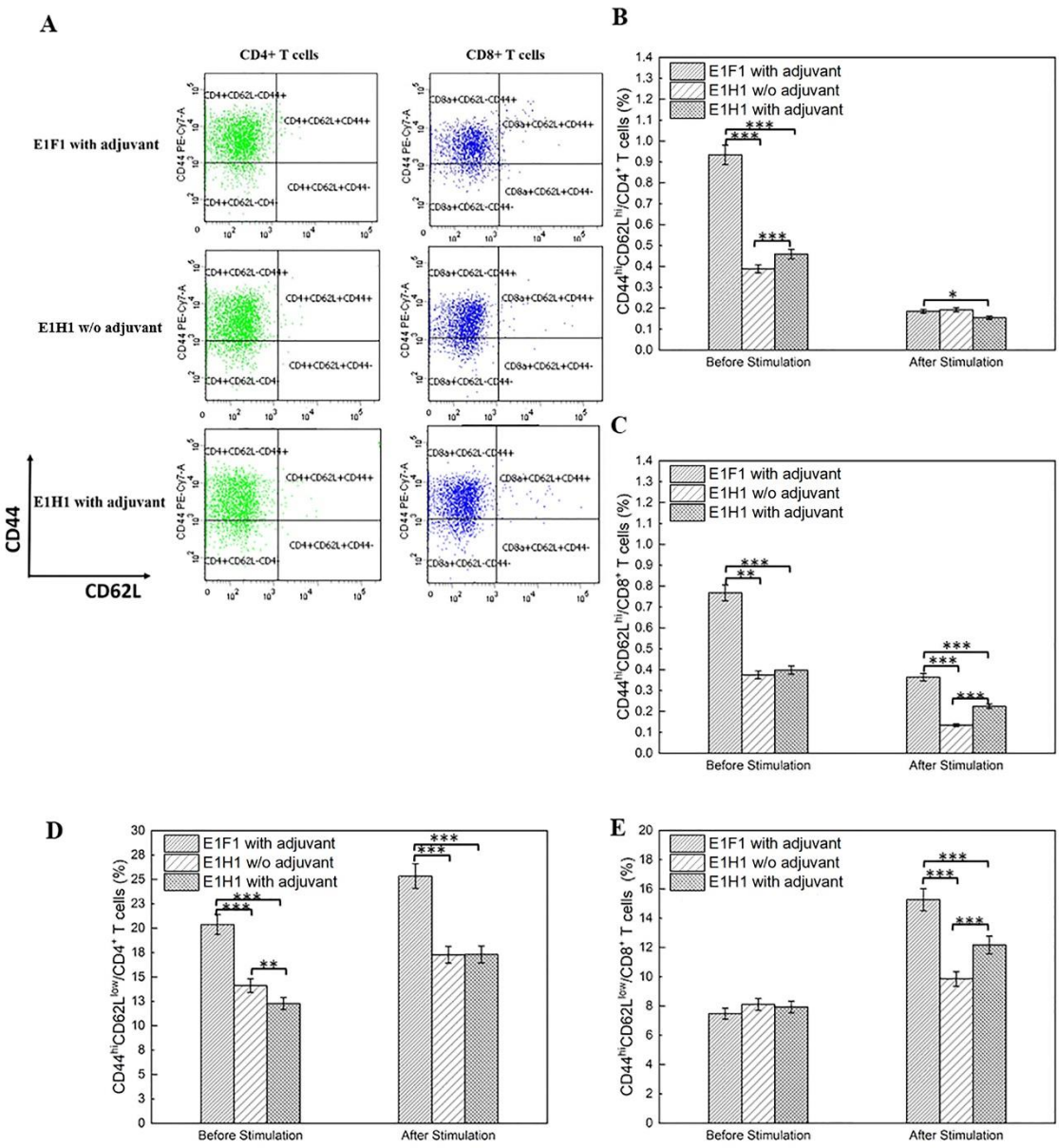


Figure 7.5 Frequencies of TCM (CD44hiCD62Lhi) and TEM (CD44hiCD62Llow) cells. A) FACS plots. FACS plots are representative of the mean percentage of six mice in each group. B) CD4+ TCM cells (%). C) CD8+ TCM cells (%). D) CD4+ TEM cells. E) CD8+ TEM cells (%). Data are shown in mean ± SEM (n=6) (*p<0.05; **p<0.01; ***p<0.001).

7.6.6 Adjuvant adsorption fraction determination and epitope density calculations

To confirm whether E1F1 and E1H1 can be adsorbed by aluminum hydroxide adjuvant, adjuvant adsorption fraction was determined. E1F1 was completely adsorbed by adjuvant (Table 7.1), whereas average adsorption percentage for E1H1 was 54 %. The particle size for E1F1 and E1H1 after being adsorbed by aluminum hydroxide adjuvant was confirmed by Mastersizer (Figure S7.2, Supporting Information). The particle size distribution for E1F1 and E1H1 with adjuvant was similar (Table S7.2, Supporting Information), and there was only a minor shift compared to the fresh aluminum hydroxide adjuvant size distribution.

Table 7.1 Adjuvant adsorption percentage for E1F1 and E1H1.

Sample name	**Adjuvant adsorption (%)**			
	1	2	3	Average
EBNA1-Ferritin	100	100	100	100 ± 0 %
EBNA1-HBc	52	54	56	54 ± 1.63 %

To understand differences in terms of epitope density, further calculations were performed. It is assumed that effect of epitope density on the immunity responses is based on the difference in antigen organization on the single protein particle.

Each single E1H1 protein particle carried 240 (T=4) or 180 epitopes (T=3), while each single E1F1 protein particle carried 24 epitopes. The epitope density ratio of E1H1 and E1F1 was 10 (T=4) or 7.5 (T=3) (S7.4, Supporting Information). That demonstrates the epitope density of E1H1 was always higher than E1F1

7.4. Discussion

For all three groups, namely, E1F1 with adjuvant, E1H1 with or without adjuvant, the dominant immune response was humoral. The EBNA1 short peptide (HPVGEADYFEY) itself belongs to $CD4^+$ T cells epitope and primarily triggers Th1 responses [45], which are responsible for activating cell-mediated immune response. However, these were not consistent with our results. This immune response shift has been found in several previous studies [46-48], mostly because of effects of the adjuvant and vaccine carrier. Aluminum hydroxide has been proven to be one of the adjuvants that mainly induces a Th2 response to primarily activate humoral immune response [49-52]. Therefore, the dominant immune response for E1F1 with adjuvant depends on aluminum hydroxide adjuvant. The HBC VLP itself has been known to induce a predominantly IgG1 antibody response (Th2) [53]. Thus, E1H1 with or without adjuvant tends to be humoral immune response dominant.

Moreover, with the same antigen applied (EBNA1), HFn (E1F1) and HBc (E1H1) showed different immunogenicity. IgG titers demonstrated that E1H1 produced greater humoral immune responses than E1F1. The proliferation assay indicated that E1H1 enhanced the proliferation index compared to E1F1. There are three possible reasons. In the first place, HBc-VLP has self-immunogenicity in the human body while HFn does not [54]. HFn-based vaccines are expected to be co-administrated with adjuvant to boost the immune response [10,12,14]. By contrast, it has been confirmed by the IgG titer results that adjuvant-free E1H1 induced a potent humoral immune response. This highlights the advantage of HBc-VLP as vaccine carrier. The adjuvant-free formulation is significantly favored in vaccination, because aluminum adjuvant was found to have risks in serious immunological disorders [55]. Secondly, HBc (5040 kDa) is obviously a larger molecule than HFn molecule (504 kDa), and also has larger particle size of 35 nm compared with 12 nm. Research by Chang *et al*. suggests that protein nanoparticle vaccine size affects significantly the immune responses, especially in terms of dendritic cell

(DC) processing [56]. The optimal nanoparticle size range to induce strong immune responses is from 40 nm to 3 μm [57,58]. Therefore, E1H1 (HBc based vaccine) is expected to induce stronger immune responses. Last but not least, the epitope density difference between E1F1 and E1H1 results in different immunogenicity. Studies have demonstrated that epitope density is closely associated with humoral immune responses specific to epitopes [59-61]. Findings showed that humoral immune response was enhanced with increasing epitope density [60]. Study by Liu *et al.* demonstrated that fusion proteins (S-transferase) carrying higher M2e influenza epitopes ($M2e_8$) induced polyclonal antibodies with greater an average affinity constant, and showed higher average avidity, compared with the one of low epitope density ($M2e_1$, $M2e_4$) [62]. It can be concluded here that single protein particle with high epitope density induces stronger immunogenicity compared with the one with low epitope density. Proven by experimental data *in vivo*, E1H1 produces stronger humoral immune response than E1F1, which is consistent with the calculated data, E1H1 shows a higher epitope density than E1F1.

Cell-mediated immune responses for E1F1 and E1H1 were also compared in our study. Interestingly, from both lymphocyte activation and memory lymphocyte T cells differentiation results, E1F1 induced an enhanced cell-mediated immune response than E1H1. This underlines the advantage of human ferritin in T cell immunology. It was noted by Lee *et al*. that, compared with other nanoparticle proteins (*E. coli* DNA binding protein, Thermoplasma acidophilum proteasome and HBV capsid) [63], a ferritin vaccine carrier fast drift to lymph nodes (LNs) with short incubation time (< 1 min), and the accumulation of human heavy-chain ferritin in the LNs lasted for an adequately long time (6 days) [9,63]. This is presumably caused by that human ferritin can interact with T lymphocytes [9]. It has been found that, human ferritin has a strong binding affinity to mucin domain-2 (TIM-2) and T cell immunoglobulin to control cellular immunity

[9,64,65]. The prolonged accumulation of human ferritin ensures that antigens can be efficiently exposed to immune cells in LNs [9], which may in turn enhance the immunogenicity of a vaccine [66]. This may explain why E1F1 (HFn carrier) showed a stronger cell-immune response than E1H1 (HBc vaccine carrier).

7.5. Conclusion

This study proves the success of developing HFn and HBc as vaccine carriers inserting EBNA1 short peptide, to potentially combat EBV infections. Both nanoparticles prompt significant humoral and cell-mediated immune responses. Importantly, EBNA1-HBc (E1H1) can induce immune responses without adjuvant. HBc vaccine carrier (E1H1) also has a higher epitope density than ferritin carrier (E1F1). This explained well why E1H1 produced more than 5 times the antibody titer and 2 times stronger proliferative responses than E1F1. On the other hand, E1F1 utilized the benefit of HFn on strong binding to T cell immunoglobulin. The prolonged accumulation of HFn in lymph nodes likely ensured that sufficient epitopes can be processed to immune cells. This in turn induced a stronger cell-mediated immune response than E1H1. Overall, this study has shown a successful application of HFn and HBc as vaccine carriers. This initial proof-of-concept study can potentially lead to further vaccine development against urgent viral challenges, such as EBV and new emerging viruses.

7.6 References

1 Chuan, Y. P. *et al.* High-level expression of soluble viral structural protein in *Escherichia coli*. *Journal of Biotechnology* **134**, 64-71, doi:https://doi.org/10.1016/j.jbiotec.2007.12.004 (2008).

2 Liew, M. W. *et al.* Microbial production of virus-like particle vaccine protein at gram-per-litre levels. *Journal of Biotechnology* **150**, 224-231, doi:10.1016/j.jbiotec.2010.08.010 (2010).

3 Butkovich, N. *et al.* Advancements in protein nanoparticle vaccine platforms to combat infectious disease. *WIREs Nanomedicine and Nanobiotechnology* **n/a**, e1681, doi:https://doi.org/10.1002/wnan.1681 (2020).

4 Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances* **41**, 107547, doi:https://doi.org/10.1016/j.biotechadv.2020.107547 (2020).

5 Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **10**, 787-796, doi:10.1038/nri2868 (2010).

6 Mohsen, M. O. *et al.* Major findings and recent advances in virus–like particle (VLP)-based vaccines. *Seminars in Immunology* **34**, 123-132, doi:https://doi.org/10.1016/j.smim.2017.08.014 (2017).

7 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine: Nanotechnology, Biology and Medicine* **2**, 137-149, doi:https://doi.org/10.1016/j.nano.2006.07.009 (2006).

8 Doll, T. A. P. F. *et al.*Nanoscale assemblies and their biomedical applications. *Journal of The Royal Society Interface* **10**, 20120740, doi:10.1098/rsif.2012.0740 (2013).

9 Lee, B.-R. *et al.* Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. *Scientific reports* **6**, 35182-35182, doi:10.1038/srep35182 (2016).

10 López-Sagaseta, J. *et al.* Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

11 Qu, Y. *et al.* Stability of engineered ferritin nanovaccines investigated by combined molecular simulation and experiments. *The Journal of Physical Chemistry B* **125**, 3830-3842, doi:10.1021/acs.jpcb.1c00276 (2021).

12 Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102, doi:10.1038/nature12202 https://www.nature.com/articles/nature12202#supplementary-information (2013).

13 Li, C. Q. *et al.* Ferritin nanoparticle technology: A new platform for antigen presentation and vaccine development. *Industrial Biotechnology* **2**, 143-147, doi:10.1089/ind.2006.2.143 (2006).

14 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

15 Zhen, Z. *et al.* Ferritins as nanoplatforms for imaging and drug delivery. *Expert opinion on drug delivery* **11**, 1913-1922, doi:10.1517/17425247.2014.941354 (2014).

16 Hempstead, P. D. *et al.* Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution11Edited by R. Huber. *Journal of Molecular Biology* **268**, 424-448, doi:https://doi.org/10.1006/jmbi.1997.0970 (1997).

17 Broxmeyer, H. E. *et al.* Suppressive effects *in vivo* of purified recombinant human H-subunit (acidic) ferritin on murine myelopoiesis. *Blood* **73**, 74-79 (1989).

18 Wang, W. *et al.* Serum ferritin: Past, present and future. *Biochimica et Biophysica Acta* **1800**, 760-769, doi:10.1016/j.bbagen.2010.03.011 (2010).

19 Broxmeyer, H. E. *et al.* The influence of purified recombinant human heavy-subunit and light-subunit ferritins on colony formation *in vitro* by granulocyte-macrophage and erythroid progenitor cells. *Blood* **68**, 1257-1263 (1986).

20 Gamvrellis, A. *et al.* Vaccines that facilitate antigen entry into dendritic cells. *Immunology & Cell Biology* **82**, 506-516, doi:10.1111/j.0818-9641.2004.01271.x (2004).

21 Slütter, B. & Jiskoot, W. Sizing the optimal dimensions of a vaccine delivery system: a particulate matter. *Expert Opinion on Drug Delivery* **13**, 167-170, doi:10.1517/17425247.2016.1121989 (2016).

22 Zhao, Q. *et al.* Virus-like particle-based human vaccines: quality assessment based on structural and functional properties. *Trends in Biotechnology* **31**, 654-663, doi:https://doi.org/10.1016/j.tibtech.2013.09.002 (2013).

23 Golmohammadi, R. *et al.* The crystal structure of bacteriophage Qβ at 3.5 å resolution. *Structure* **4**, 543-554, doi:https://doi.org/10.1016/S0969-2126(96)00060-3 (1996).

24 Wynne, S. A. *et al.*The crystal structure of the human hepatitis B virus capsid. *Molecular Cell* **3**, 771-780, doi:https://doi.org/10.1016/S1097-2765(01)80009-5

(1999).

25 Baltabekova, A. Z. *et al.* SplitCore technology allows efficient production of virus-like particles presenting a receptor-contacting epitope of human IgE. *Molecular Biotechnology* **57**, 746-755, doi:10.1007/s12033-015-9867-0 (2015).

26 Walker, A. *et al.* SplitCore: An exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. *Scientific Reports* **1**, 5, doi:10.1038/srep00005 (2011).

27 Li, Z. *et al.* Strong hydrophobicity enables efficient purification of HBc VLPs displaying various antigen epitopes through hydrophobic interaction chromatography. *Biochemical Engineering Journal* **140**, 157-167, doi:https://doi.org/10.1016/j.bej.2018.09.020 (2018).

28 Farrell, P. J. Epstein–Barr virus and cancer. *Annual Review of Pathology: Mechanisms of Disease* **14**, 29-53, doi:10.1146/annurev-pathmechdis-012418-013023 (2019).

29 Pattle, S. B. & Farrell, P. J. The role of Epstein–Barr virus in cancer. *Expert Opinion on Biological Therapy* **6**, 1193-1205, doi:10.1517/14712598.6.11.1193 (2006).

30 Qu, Y. *et al.* Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* **39**, 4830-4841, doi:https://doi.org/10.1016/j.vaccine.2021.07.021 (2021).

31 Guo, F. *et al.* Prompt and robust humoral immunity elicited by a conjugated chimeric malaria antigen with a truncated flagellin. *Bioconjugate Chemistry* **29**, 761-770, doi:10.1021/acs.bioconjchem.7b00320 (2018).

32 Hammarström, L. & Smith, C. I. IgG subclass changes in response to vaccination. *Monographs in Allergy* **19**, 241-252 (1986).

33 Lubeck, M. D. *et al.* The interaction of murine IgG subclass proteins with human monocyte Fc receptors. *The Journal of Immunology* **135**, 1299-1304 (1985).

34 Williams, M. A. & Bevan, M. J. Effector and memory CTL differentiation. *Annual Review Immunology* **25**, 171-192, doi:10.1146/annurev.immunol.25.022106.141548 (2007).

35 Poston, R. N. & Morgan, R. S. Interactions between soluble IgG, complement and cells in lymphocyte and monocyte ADCC. *Immunology* **50**, 461-469 (1983).

36 Datta, S. & Sarvetnick, N. Lymphocyte proliferation in immune-mediated diseases. *Trends in Immunology* **30**, 430-438, doi:https://doi.org/10.1016/j.it.2009.06.002 (2009).

37 Driesen, J. *et al.* CD25 as an immune regulatory molecule expressed on myeloid dendritic cells. *Immunobiology* **213**, 849-858,

doi:https://doi.org/10.1016/j.imbio.2008.07.026 (2008).

38 Wang, K. *et al.*. CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. *Experimental Hematology & Oncology* **1**, 36-36, doi:10.1186/2162-3619-1-36 (2012).

39 Lederman, S. *et al.* Identification of a novel surface protein on activated CD4+ T cells that induces contact-dependent B cell differentiation (help). *Journal of Experimental Medicine* **175**, 1091-1101, doi:10.1084/jem.175.4.1091 (1992).

40 Laidlaw, B. J. *et al.* The multifaceted role of CD4+ T cells in CD8+ T cell memory. *Nature Reviews Immunology* **16**, 102-111, doi:10.1038/nri.2015.10 (2016).

41 Cohn, L. & Delamarre, L. Dendritic cell-targeted vaccines. *Frontiers in immunology* **5**, 255-255, doi:10.3389/fimmu.2014.00255 (2014).

42 Krzych, U. *et al.* Memory T cells maintain protracted protection against malaria. *Immunology Letters* **161**, 189-195, doi:10.1016/j.imlet.2014.03.011 (2014).

43 Sallusto, F. *et al.* Central memory and effector memory T cell subsets: Function, generation, and maintenance. *Annual Review of Immunology* **22**, 745-763, doi:10.1146/annurev.immunol.22.012703.104702 (2004).

44 Kaech, S. M. *et al.* Effector and memory T-cell differentiation: implications for vaccine development. *Nature Reviews Immunology* **2**, 251-262, doi:10.1038/nri778 (2002).

45 Bickham, K. *et al.* EBNA1-specific CD4+ T cells in healthy carriers of Epstein-Barr virus are primarily Th1 in function. *Journal of Clinical Investigation* **107**, 121-130, doi:10.1172/JCI10209 (2001).

46 Korsholm, K. S. *et al.* T-helper 1 and T-helper 2 adjuvants induce distinct differences in the magnitude, quality and kinetics of the early inflammatory response at the site of injection. *Immunology* **129**, 75-86, doi:10.1111/j.1365-2567.2009.03164.x (2010).

47 Ciabattini, A. *et al.* Modulation of primary immune response by different vaccine adjuvants. *Frontiers in Immunology* **7**, doi:10.3389/fimmu.2016.00427 (2016).

48 Awate, S. *et al.* Mechanisms of action of adjuvants. *Frontiers in immunology* **4**, 114-114, doi:10.3389/fimmu.2013.00114 (2013).

49 Bomford, R. *et al.* The control of the antibody isotype response to recombinant human immunodeficiency virus gp120 antigen by adjuvants. *AIDS Research and Human Retroviruses* **8**, 1765-1771, doi:10.1089/aid.1992.8.1765 (1992).

50 Comoy, E. E. *et al. In vivo* induction of type 1 and 2 immune responses against protein antigens. *International Immunology* **9**, 523-531, doi:10.1093/intimm/9.4.523 (1997).

51 Jiao, X. D. *et al.* Comparative study of the effects of aluminum adjuvants and Freund's

incomplete adjuvant on the immune response to an Edwardsiella tarda major antigen. *Vaccine* **28**, 1832-1837, doi:10.1016/j.vaccine.2009.11.083 (2010).

52 Hogenesch, H. Mechanism of immunopotentiation and safety of aluminum adjuvants. *Frontiers in immunology* **3**, 406-406, doi:10.3389/fimmu.2012.00406 (2013).

53 Sominskaya, I. *et al.* Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. *Clinical and Vaccine Immunology* **17**, 1027, doi:10.1128/CVI.00468-09 (2010).

54 Tan, A. *et al.* Immune response in hepatitis B virus infection. *Cold Spring Harbor Perspectives in Medicine* **5**, a021428-a021428, doi:10.1101/cshperspect.a021428 (2015).

55 Tomljenovic, L. & Shaw, C. A. Aluminum vaccine adjuvants: are they safe? *Current Medicinal Chemistry* **18**, 2630-2637, doi:10.2174/092986711795933740 (2011).

56 Chang, T. Z. *et al.* Effects of ovalbumin protein nanoparticle vaccine size and coating on dendritic cell processing. *Biomaterials science* **5**, 223-233, doi:10.1039/c6bm00500d (2017).

57 Fifis, T. *et al.* Size-dependent immunogenicity: therapeutic and protective properties of nano-vaccines against tumors. *The Journal of Immunology* **173**, 3148-3154, doi:10.4049/jimmunol.173.5.3148 (2004).

58 Tran, K. K. & Shen, H. The role of phagosomal pH on the size-dependent efficiency of cross-presentation by dendritic cells. *Biomaterials* **30**, 1356-1362, doi:10.1016/j.biomaterials.2008.11.034 (2009).

59 Jegerlehner, A. *et al.* Regulation of IgG antibody responses by epitope density and CD21-mediated costimulation. *European Journal of Immunology* **32**, 3305-3314, doi:10.1002/1521-4141(200211)32:11<3305::Aid-immu3305>3.0.Co;2-j (2002).

60 Liu, W. *et al.* High epitope density in a single recombinant protein molecule of the extracellular domain of influenza A virus M2 protein significantly enhances protective immunity. *Vaccine* **23**, 366-371, doi:10.1016/j.vaccine.2004.05.028 (2004).

61 Liu, W. & Chen, Y. H. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. *European Journal of Immunology* **35**, 505-514, doi:10.1002/eji.200425749 (2005).

62 Liu, W. & Chen, Y.-H. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine

development and a preliminary investigation about B cell discrimination of monomeric proteins. *European Journal of Immunology* **35**, 505-514, doi:https://doi.org/10.1002/eji.200425749 (2005).

63 Lee, B.-R. *et al.* Engineered human ferritin nanoparticles for direct delivery of tumor antigens to lymph node and cancer immunotherapy. *Scientific Reports* **6**, 35182, doi:10.1038/srep35182 (2016).

64 de Souza, A. J. & Kane, L. P. Immune regulation by the TIM gene family. *Immunologic Research* **36**, 147-155, doi:10.1385/IR:36:1:147 (2006).

65 Fargion, S. *et al.* Specific binding sites for H-ferritin on human lymphocytes: modulation during cellular proliferation and potential implication in cell growth control. *Blood* **78**, 1056-1061 (1991).

66 Bachmann, M. F. *et al.* Long-lived memory CD8+ T cells are programmed by prolonged antigen exposure and low levels of cellular activation. *European Journal of Immunology* **36**, 842-854, doi:10.1002/eji.200535730 (2006).

7.7 Supporting Information

7.7.1 Supporting figures

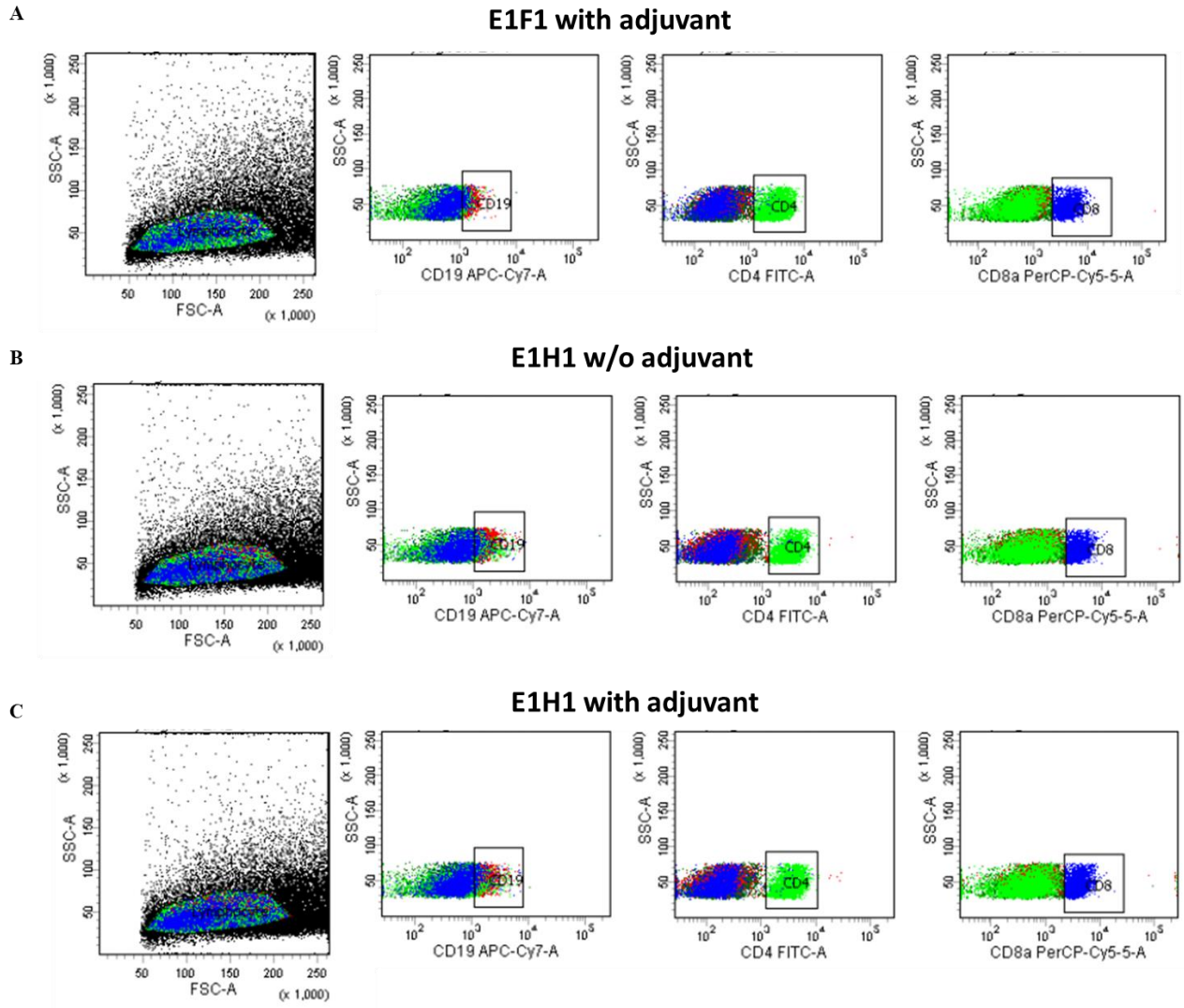


Figure S7.1 Flow cytometry gates showing CD19, CD4 and CD8. CD19 is shown in red colour; CD 4 is shown in green; and CD8 is shown in blue. A) E1F1 with adjuvant. B) E1H1 without adjuvant. C) E1H1 with adjuvant.

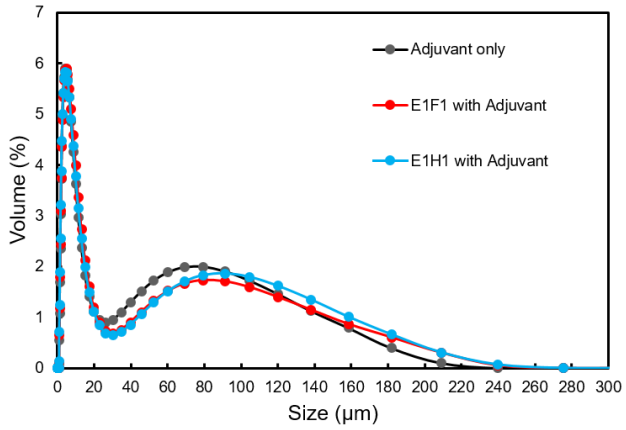


Figure S7.2 Particle size for aluminum hydroxide adjuvant, E1F1 with adjuvant and E1H1 with adjuvant by Mastersizer.

7.7.2 Supporting tables

Table S7.1 Amino acid sequence of E1F1 and E1H1.

Name	Amino acid Sequence
EBNA1-Ferritin **(Ferritin with inserted EBNA1 epitope at the N-terminus of ferritin)**	**HPVGEADYFEY**GGSGGGGSGGGGSGGTTASTSQVRQNYHQDSEAAINRQINLELYASYV YLSMSYYFDRDDVALKNFAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDI KKPDCDDWESGLNAMECALHLEKNVNQSLLELHKLATDKNDPHLCDFIETHYLNEQVK AIKELGDHVTNLRKMGAPESGLAEYLFDKHTLGDSDNES
EBNA1-HBc **(HBc with inserted EBNA1 epitope at the N-terminus of HBc)**	**HPVGEADYFEY**MDIDPYKEFGASVELLSFLPSDFFPSIRDLLDTASALYREALESPEHCSPH HTALRQAILCWGELMNLATWVGSNLEDPASRELVVSYVNVNMGLKIRQLLWFHISCLTFG RETVLEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVVRRRGRSPRRRTPSPRRRRSQSPRRR RSQSRESQC

Table S7.2 Particle size by Mastersizer.

Name	d (0.1) µm	d (0.5) µm	d (0.9) µm
Aluminium adjuvant	1.95 (± 0.03)	5.66 (± 0.10)	66.44 (± 0.62)
E1F1 with adjuvant	1.96 (± 0.02)	5.61 (± 0.21)	80.20 (± 5.25)
E1H1 with adjuvant	1.99 (± 0.09)	5.62 (± 0.13)	81.76 (± 8.59)

*d (0.1), d (0.5) and d (0.9) represents the diameter of a sphere at which 10 %, 50 % or 90 % of the particles in the samples are smaller on a number basis.

7.7.3 Calculation of injected antigen dose

EBNA1 short antigen: Mw = 1326 Da
E1F1 carries 24 antigens in each particle, whereas E1H1 carries 240 (T=4) or 180 (T=3) antigens in each particle.

E1F1
m = 100 µg; Mw = 559.2 kDa; $n = \frac{m}{M_w} = 1.79 \times 10^{-10}$ mol
Moles of antigen = $1.79 \times 10^{-10} \times 24 = 4.29 \times 10^{-9}$ mol
Mass of antigen = $4.29 \times 10^{-9} \times 1326 \approx 6$ µg

E1H1

m = 100 µg;

T=4
Mw = 5376 kDa; $n = \frac{m}{M_w} = 1.86 \times 10^{-11}$ mol
Moles of antigen = $1.86 \times 10^{-11} \times 240 = 4.46 \times 10^{-9}$ mol
Mass of antigen = $4.46 \times 10^{-9} \times 1326 \approx 6$ µg

T=3
Mw = 4032 kDa; $n = \frac{m}{M_w} = 2.48 \times 10^{-11}$ mol
Moles of antigen = $2.48 \times 10^{-11} \times 180 = 4.46 \times 10^{-9}$ mol
Mass of antigen = $4.46 \times 10^{-9} \times 1326 \approx 6$ µg

7.7.4 Calculation of epitope density ratio (number basis)

E1H1 (T=4)

$$\frac{\text{Epitope density}_{\text{E1H1}}}{\text{Epitope density}_{\text{E1F1}}} = \frac{240}{24} = 10$$

E1H1 (T=3)

$$\frac{\text{Epitope density}_{\text{E1H1}}}{\text{Epitope density}_{\text{E1F1}}} = \frac{180}{24} = 7.5$$

Epitope density is expressed in number of epitopes per single protein particle.

Publication included in Chapter 7

Qu, Y. *et al.* Immunogenicity and vaccine efficacy boosted by engineering human heavy chain ferritin and chimeric hepatitis B virus core nanoparticles. *ACS Applied Bio Materials* doi: 10.1021/acsabm.1c00738

ACS APPLIED BIO MATERIALS

www.acsabm.org Article

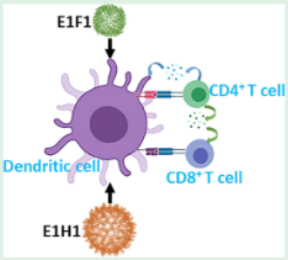
Immunogenicity and Vaccine Efficacy Boosted by Engineering Human Heavy Chain Ferritin and Chimeric Hepatitis B Virus Core Nanoparticles

Yiran Qu,[‖] Bingyang Zhang,[‖] Yingli Wang, Shuang Yin, Yan Sun, Anton Middelberg, and Jingxiu Bi*

Cite This: *ACS Appl. Bio Mater.* 2021, 4, 7147−7156 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Human heavy-chain ferritin (HFn) and hepatitis B virus core (HBc) are both nanoparticle proteins presenting a well-oriented architecture with constant size and shape, which can be engineered to carry epitopes on the surface of the nanoparticle protein cage, enabling vaccine design. This study aims to investigate the immunogenicity differences between engineered HFn and chimeric HBc bearing the same epitope. As a proof of concept, the model epitope Epstein−Barr nuclear antigen 1 (EBNA1) is inserted at the N-terminus of the HFn and HBc subunit to produce two vaccine candidates named EBNA1−HFn (E1F1) and EBNA1−HBc (E1H1), respectively. From *in vivo* immunogenicity studies, E1H1 demonstrates the capability to prompt significant humoral and cell-mediated immune responses in adjuvant-free formulation. When formulated with the aluminum hydroxide adjuvant, E1H1 produces approximately 5× higher titer and 2× stronger proliferation index (PI) than E1F1. These results confirm that the HBc carrier induces a stronger humoral immune response than HFn. On the other hand, from lymphocyte activation experiments, E1F1 induces a stronger cell-mediated immune response indicated by 5× more $CD8^+$T cells and 2× more effector memory T cells in the E1F1 group versus the E1H1 group. Through this study, HFn and HBc are shown to be potentially effective vaccine carrier nanoparticles having subtly different immunological responses.



KEYWORDS: *protein-based vaccine molecule design, hepatitis B virus core, virus like particles, human ferritin heavy chain, Epstein−Barr nuclear antigen 1*

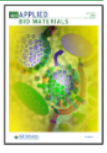
1. INTRODUCTION

The current COVID-19 pandemic has increased urgency for new-engineered reliable vaccines. It has increased the need to develop vaccine platforms meeting criteria, including fast manufacture at a large scale in response to a pandemic threat and high safety yet with optimal efficacy and efficiency to generate potent and protective immune responses. There is considerable interest in applying recombinant protein nanoparticles as a vaccine delivery system because (i) they can be produced at large scale *via* cost-effective expression systems such as bacteria;[1,2] (ii) protein nanoparticles are regarded as safe with minimal risks compared to traditional vaccines such as killed or inactivated pathogen vaccines;[3] and (iii) nanoparticle proteins are highly repetitive structures that can be developed as efficacious and efficient vaccines to induce potent immune responses.[4−6] There is a wide selection of sizes (5−100 nm).[7] Ferritins and virus-like particles (VLPs) are two common examples of nanoparticle proteins.

Ferritins are self-assembling nanoparticle proteins with 24 protein subunits that form spherical cage structures.[8] The inner and outer diameters for a ferritin particle are 8 and 12 nm, respectively.[9] Ferritin has recently been explored as a vaccine carrier due to its various advantages including robust thermal and chemical stabilities,[10,11] suitable presentation of antigens and presentation of immunogens,[10,12−14] low toxicity, high biocompatibility and good biodegradability.[11,15] Recombinant human heavy-chain ferritin (HFn) is one of the well-understood ferritins.[16] The molecular weight of each subunit is about 21 kDa. It was found by Broxmeyer *et al.* that human ferritin heavy chain suppressed immune activity in human *in vivo*.[17,18] They also noticed that ferritin heavy chain notably reduced growth of granulocyte-macrophage, erythroid and multipotential progenitor cells.[17−19] On this account, it is a common practice that engineered ferritin-based vaccines are often administrated with adjuvants to boost potential immune responses.[10,12−14]

VLPs are constructed by viral capsid proteins assembled to a hollow nanocage lacking infectious nucleic acids.[4,8] VLPs are applicable vaccine carriers as they present antigens in a multimeric and well-organized manner, which induces desirable humoral and cell-mediated immune responses.[4,5,20−22] One of the well-characterized VLPs is the hepatitis B virus core

Received: June 28, 2021
Published: September 8, 2021



protein (HBc). HBc is made up of 180 equal subunits ($T = 3$) with outer diameters of 27−31 nm.[4,23] There is also a proportion of HBc comprising 240 equal subunits ($T = 4$) with an external diameter of 35 nm.[24] The molecular weight for each subunit is approximately 21 kDa. Because of its strong immunity and extreme flexibility to insert peptide or epitope with up to 300 residues with negligible effects on assembled structure,[25−27] HBc is highly promising as a vaccine carrier.

Currently, research directly comparing the immune responses induced by HFn and HBc as vaccine carriers bearing the same epitope has not been reported. Herein, we applied Epstein−Barr nuclear antigen 1 (EBNA1) as the model epitope to develop two vaccine candidates with different molecular designs targeting Epstein−Barr virus (EBV). EBNA1 can be expressed in all EBV-associated tumors to support maintenance and expression of EBV.[28] EBV, belonging to the human herpesviruses family,[28] result in serious disease (*e.g.* Hodgkin's disease, lymphomas and lymphoproliferative disease).[29] There is no current licensed vaccine available to combat EBV infection.[28] There is some urgency to develop EBV-associated vaccines for controlling EBV infection.

To reveal differences between HFn and HBc as vaccine carriers, short EBNA1 peptide was inserted at the N-terminus of HFn and HBc. The resulting vaccine are candidates named EBNA1−HFn (E1F1) and EBNA1−HBc (E1H1). Before *in vivo* immunization test, molecular structure was characterized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), size exclusion chromatography multi-angle light scattering (SEC-MALS) and transmission electron microscopy (TEM) to confirm the purity of E1H1 (>95%) and E1F1 (>95%), hydrodynamic radius and molecular weight (E1H1: 17.4 nm and 4740 kDa; E1F1: 7.3 nm and 562 kDa). After immunization in mice with (E1H1 and E1F1) and without (E1H1) adjuvant, immune response (humoral and cell-mediated), proliferation index between E1H1 and E1F1 were compared. *In vivo* experiment demonstrates HBc VLP carrier can induce potent immune response in adjuvant-free formulation. Boosted by adjuvant, E1H1 induces greater humoral immune and proliferative response than E1F1. However, E1F1 induces greater cell-mediated immune response than E1H1. This paper sheds new light on chimeric protein as vaccine carriers to boost immunogenicity and vaccine efficacy, which can be produced at large scale in a short period *via* cost-effective process to against pandemic infection.

2. EXPERIMENTAL SECTION

2.1. Nanoparticle Protein Expression. The detailed information about plasmid construction for EBNA1−ferritin (E1F1) can be found in our previous research.[11] For EBNA1−HBc (E1H1), plasmid was built using the pET 30a vector. The restriction sites were *Nde I* and *BamH I*.[11] Information about protein sequence is shown in Table S1, Supporting Information. *Escherichia coli* BL21 (DE3) (Invitrogen, USA) was used to transform plasmids. Epitope short EBNA1 peptide (HPVGEADYFEY) (taken from EBNA1 aa 407 to aa 417) was inserted at the N-terminus of either HFn or HBc. Single colony expressing each protein was incubated in the lysogeny broth (LB) medium with antibiotic kanamycin (1 mM). The growing process was at 37 °C for 14−16 h in a shaking incubator at 180 rpm. In every 500 mL of the LB−kanamycin (100 μg/mL) medium, 10 mL of the cultured medium from 50 mL was added to further grow bacteria (37 °C, 200 rpm, 4 h, OD_{600} = 0.8), then 1 mM IPTG was added to express proteins (37 °C, 200 rpm, 4 h). Bacterial solution was centrifuged to collect the bacteria pellet. The sonication followed the procedure[11] with lysis buffer (E1F1: 20 mM phosphate buffer (PB), 2 mM EDTA, pH 7.0; E1H1: 20 mM Tris-HCl, 3 mM EDTA, 1 mM phenylmethanesulfonyl fluoride, 0.1% Triton X-100, pH 8.0). Soluble-expressed protein was collected by centrifuging bacterial lysate at 13,751*g* for 30 min. SDS-PAGE (12%) was conducted to check the soluble expression level for E1F1 and E1H1.

2.2. Purification of Recombinant E1F1 and Recombinant E1H1. For the purification of E1F1, the supernatant of the cell lysate containing soluble expressed protein was diluted to protein concentration to 4 mg/mL with Mili-Q water.[11] The pH of the diluted supernatant was altered to 5.0 utilizing 1 M NaCl, NaAc−HAc (1 M, pH 5.0). Heating (60 °C, 10 min) was applied to precipitate protein impurities, followed by centrifugation (19,802*g*, 4 °C, 10 min) to remove protein impurities. The resultant supernatant was mixed with an ammonium sulfate (AS) solution (3 M) and a phosphate buffer (1 M) to final AS concentration 1 M at pH 6.5. Hydrophobic interaction chromatography (HIC) was applied to separate nucleic acid. A HiTrap Octyl FF column (GE Healthcare Life Sciences) was used for E1F1 purification by the binding process and AKTA Pure (GE Healthcare Life Sciences) was adopted. The column was first equilibrated by 0.1 M PB with 1 M AS, pH 6.5. After discarding the flow through the fraction, target protein E1H1 was directly eluted by 100% elution buffer (20 mM PB, pH 6.5). The collected fractions from the HIC column were then injected into a HiTrap Desalting column (GE Healthcare Life Sciences) to buffer exchange into 20 mM PB, pH 7.0 for further use.

In terms of E1H1 purification, different purification steps were applied. Briefly, the supernatant containing soluble E1H1 of the cell lysate was mixed with 1 M ammonium sulfate and equilibrated for 30 min at 25 °C, pH 7.4. The mixture was centrifuged (19,802*g*, 4 °C, 10 min) to harvest E1H1. Pellets of E1H1 were collected and dissolved in 20 mM Tris-HCl, 4 M urea, pH 9.0. Dissolved E1H1 was then dialyzed against 20 mM PB, pH 7.0 for further use. Purity of E1F1 and E1H1 was analyzed using 12% reducing SDS-PAGE.

2.3. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis. Expression level of the soluble protein and purity after purification were determined by SDS-PAGE. The procedure followed our previous research.[11] Briefly, the diluted samples (1 mg/mL) mixed with loading buffer[11] were disassembled by heating (100 °C, 10 min). Lanes were separated by electrophoresis (Bio-Rad, USA). The gel for 1 h was stained and the background was washed using washing buffer.[11]

2.4. Transmission Electron Microscopy. TEM was utilized to characterize conformations of purified E1F1 and E1H1. Purified E1F1 and E1H1 were diluted to 0.1 mg/mL. Protein samples (10 μL) were added on grids (carbon-coated copper) and settled for 5 min. Grids were then cleaned by RO water and negatively stained for 2 min[11] Grids were analyzed with TEM (Philips CM100, 100 kV) (Field Electron and Ion Company, USA).

2.5. Size Exclusion Chromatography Multiangle Light Scattering. SEC-MALS was demonstrated to detect hydrodynamic radius and molecular weights of purified proteins. For E1F1 characterization, Superose 6 Increase 10/300 GL (GE Healthcare, USA) was connected to high-performance liquid chromatography (HPLC) (Shimadzu, Japan) coupled with a multiangle laser light scattering (MALS) Wyatt DAWN HELEOS II and Optilab T-rEx (Wyatt Technology, Santa Barbara, CA USA).[11] The equilibration buffer was 20 mM PB, pH 7.0. For E1H1 characterization, TSKgel G4000SWXL (TOSOH Bioscience, Japan) was attached to HPLC-MALS. The equilibration buffer was 20 mM PB, pH 7.4. The loading volume was 50 μL and minimum protein concentration was as follows: E1F1 = 5 mg/mL and E1H1 = 1 mg/mL, respectively. The signal intensity at 280 nm versus retention time was plotted, and protein hydrodynamic radius and molecular weight were analyzed by ASTRA software (v. 6.1).

2.6. Adjuvant Adsorption Percentage Determination. Purified E1F1 and E1H1 were diluted with Mili-Q water to 1 mg/mL. An aluminum hydroxide adjuvant (750 μL, 0.8 mg/mL, SERVA, Sweden) was mixed with 750 μL of the diluted protein samples (E1F1 and E1H1) and stored at room temperature (25 °C) for 24 h. After incubation, protein-adjuvant mixtures were centrifuged at 2100*g* for 20 min. The supernatants were collected to measure the protein

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

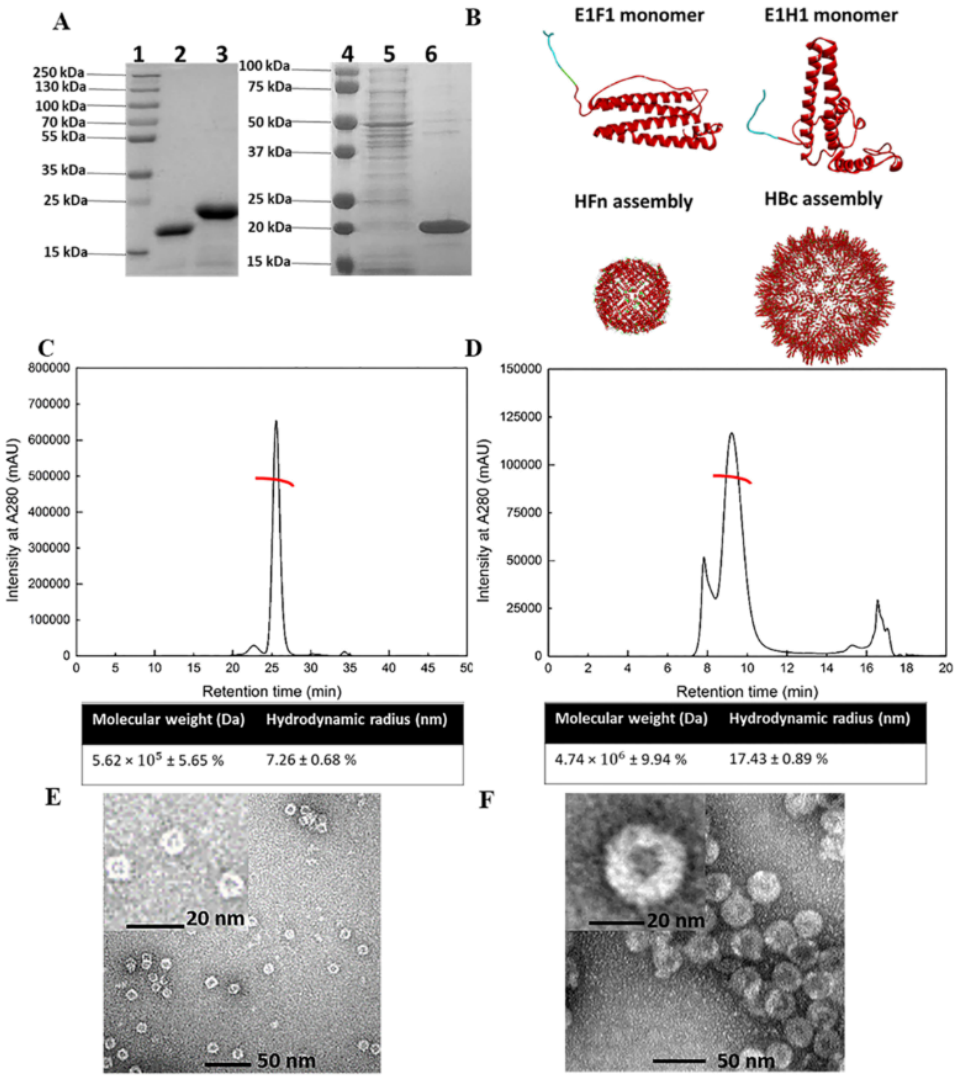


Figure 1. Characterization of E1F1 (ferritin carrier) and E1H1 (HBc carrier). (A) Coomassie blue stained sodium dodecyl sulfate (SDS): lane 1, protein marker 1; lane 2, HFn marker (~21 kDa); lane 3, E1F1 (~24 kDa); lane 4, protein marker 2; lane 5, impurities (N/A); and lane 6: E1H1 (~20 kDa). (B) Mimic diagrams by Discovery Studio 2020. Red denotes carrier, blue denotes EBNA1 short peptide epitope, and green denotes soft linkers $(GGSGG)_3$ (i) E1F1 monomer; (ii) E1H1 monomer; (iii) assembled HFn; and (iv) assembled HBc. (C) Molar weight and particle size of assembled E1F1 measured by SEC-MALS; (D) molar weight and particle size of assembled E1H1 measured by SEC-MALS; (E) assembled E1F1 TEM image; and (F) assembled E1H1 TEM image.

concentration by the Bradford method (Bio-Rad, USA). The adjuvant adsorption (%) was calculated using the following equation (eq 1).

$$\text{Adjuvant adsorption (\%)} = (M_1 - M_2)/M_1 \times 100\% \quad (1)$$

M_1: mass of protein before adding adjuvant. M_2: mass of protein in the supernatant after adding the adjuvant and centrifugation.

2.7. Particle Size Measurement. A Mastersizer 2000 (Malvern, USA) was used to measure the particle size for the aluminum hydroxide adjuvant (10 mg/mL), E1F1 with an adjuvant (10 mg/mL), and E1H1 with an adjuvant (5 mg/mL). The measurement options were set as follows: refractive index: 1.66 and adsorption: 0.01. For each sample, 5 mL was added to make detectable particle size measurements. Each sample was measured three times. All the measurements were analyzed by Mastersizer software (Malvern, USA).

2.8. Animal Immunization. Animal immunization experiments were planned and undertaken by authors in China and were approved by the Medical Ethics Committee of ShanXi University of Traditional Chinese Medicine (Approval Number 2019LL137).

Sample groups (n = 10/treatment) of 6–8 week-old female BALB/c mice were immunized *via* intraperitoneal injection applied into the peritoneum[30] group 1: 100 μg of purified E1F1 (1 mg/mL) in PBS with 200 μL of adjuvant (0.8 mg/mL); group 2: 100 μg of purified E1H1 (1 mg/mL) in PBS; group 3: 100 μg of purified E1H1 (1 mg/mL) in PBS with 200 μL of the adjuvant (0.8 mg/mL). All the three samples were injected with equivalent 6 μg of the EBNA1 epitope. The detailed calculation was shown in Supporting Information, S3. Another two control groups (n = 10/treatment) were injected with 200 μL of PBS (blank) and 100 μg of ovalbumin,[30] respectively. The serum was collected 10 days after the first (at day 24) and the second boost (at day 38) and stored at −80 °C before further measurement.

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

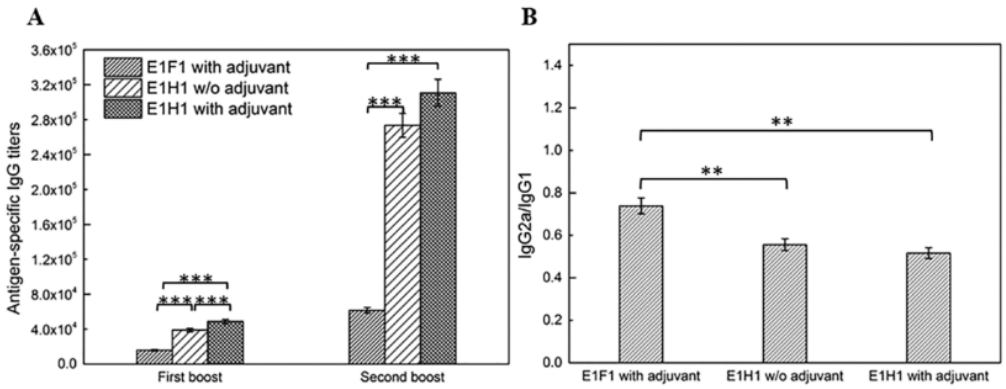


Figure 2. EBNA1-specific antibodies in sera of BALB/c mice. (A) IgG titers 10 days after first boost (day 24) and second boost (day 38). (B) IgG2a/IgG1 at day 38. Data are shown in mean SEM (n = 3) (*p < 0.05; **p < 0.01; ***p < 0.001).

2.9. Enzyme-Linked Immunosorbent Assay and Antigen-Specific IgG Isotype. The detailed procedure followed our published research.[30] Short EBNA1 peptide (HPVGEADYFEY) was adsorbed for 14 h at 4 °C in 96-well plates.[30] After being blocked with the serum at 37 °C for 2 h, the plate was added with mouse sera of serial dilutions collected before (day 24 and day 38). The plate was further incubated at 37 °C for 2 h. After washing with PBS three times, 200 μL of goat anti-mouse IgG horseradish peroxidase-conjugated antibody of 1:5000 dilution was added to the plate and incubated at 37 °C for 2 h.[30] Color was developed by 3,3′,5,5′-tetramethylbenzidine single-component substrate solution for 10 min. Sulfuric acid (1 M) was utilized to stop this reaction. A microplate reader (Perlong, China) (450 nm) was used to examine plates.

Steps prior to plate washing were similar to the enzyme-linked immunosorbent assay (ELISA).[30] Antibodies goat anti-mouse IgG1 and IgG2a (1:1000 dilution, 100 μL) were loaded to each plate.[30] 200 μL of rabbit anti-goat IgG horseradish peroxidase-conjugated antibody (1:5000 dilution) was loaded to the plate. The following steps were the same as ELISA to develop color, stop reaction, and plate analysis.

2.10. T-Cell Proliferation Assay. Single-cell suspensions (4×10^6 cells/mL) were cultured in RPMI 1640 (Gibco, Germany) with an antigen (EBNA1: 20 μg/mL) and incubated at 37 °C for 2 days. CCK-8 (10 μL) (Dojindo, Japan) was added per well (37 °C, 4 h). A microplate reader (Perlong, China) was used to analyze results. The proliferation indexes (PI) was calculated by eqs 2 and 3[30]

$$\text{proliferation rate} = \frac{A(\text{sample}) - A_0(\text{control})}{A_0(\text{control})} \quad (2)$$

A (sample): A_{450} nm of the sample group with CCK-8 and stimulator (short EBNA1 peptide). A_0 (control): A_{450} nm of the control group with CCK-8 only

$$\text{proliferation index} = \frac{\text{proliferation rate (sample)}}{\text{proliferation rate (PBS)}} \quad (3)$$

2.11. Lymphocyte Activation and Memory T-Cells of Splenocytes. The procedure to culture splenocytes (10^6 cells) from immunized mice followed the referenced procedure.[31] Splenocytes were stained with fluorescent-labeled anti-mouse antibodies: CD4, CD8a (PerCP-Cy 5.5) monoclonal antibody, CD19 (eFluor 780), CD25, CD44 (PE−Cyanine 7), and CD62L (L-selectin) monoclonal antibodies (Thermo Fisher, USA) at 4 °C.[30] The number of cells for CD25/CD4 T, CD25/CD8 T, CD25/CD19 B, activated lymphocytes ($CD69^+$), central memory T-cells ($CD44^{hi}CD62L^{hi}$), and effector memory T-cells ($CD44^{hi}CD62L^{low}$) was measured by a FACSCanto II flow cytometry.[30,31] After flow cytometry was calibrated using standard BD beads, gates were circled to locate $CD4^+$T, $CD8^+$T, and $CD19^+$B cells first. CD25 was circled in $CD4^+$T, $CD8^+$T, and $CD19^+$B cells to determine lymphocyte activation. Cross gates of CD44 and CD62L in $CD4^+$T, $CD8^+$T, and $CD19^+$B cells were drawn to detect memory T cells. FACSCanto II software was used to demonstrate data analysis.

2.12. Statistics. Data were shown in mean standard deviation.[30] Statistics was analyzed by GraphPad Prism (2021). The P-value was calculated using the one-way ANOVA method (GraphPad Prism 2021), where *p < 0.05, **p < 0.01, and ***p < 0.001.

3. RESULTS

3.1. Characterization of E1F1 and E1H1. Nanoparticle protein characterization is important to confirm that key quality criteria including good purity and quaternary structure are satisfied prior to biological experiments. As shown by Figure 1A, both E1F1 and E1H1 achieved purity above 95%. The molecular weight of the E1F1 monomer by SDS-PAGE (Figure 1A) is around 24 kDa (lane 3), slightly larger than E1H1 (20 kDa) (lane 6). Based on the molecular weight prediction server ExPaSy Bioinformatics Resource Portal (SIB Swiss Institute of Bioinformatics, Swiss), the theoretical molecular weights for E1F1 monomer and E1H1 monomer were 23.3 and 22.4 kDa, respectively, consistent with the experiment. Figure 1B portrays the assembled structures for HFn and HBc carriers. Both assembled HFn and HBc show a particulate structure; HBc comprises more subunits than HFn and thus has a larger structure. Monomer structures for E1F1 and E1H1 are also displayed to portray how antigens were inserted. Model epitope EBNA1 was inserted at the N-terminus of HFn and HBc monomers (Figure 1B). Hence, for both HFn and HBc, epitopes were exposed on the surface of the protein nanoparticles. Compared to the HBc monomer, an extra 15-residues $(GGSGG)_3$ of the soft linker was inserted for E1F1 between the epitope and ferritin monomer to confer flexibility. As shown in Figure 1C,D, the molecular weight and hydrodynamic radius measured by SEC-MALS for assembled E1F1 nanoparticles were 562 kDa and 7.3 nm, comparing with 4740 kDa and 17.4 nm for assembled E1H1 (T = 3). E1H1 is eight times heavier and two times larger than E1F1. Theoretical assembled E1H1 (T = 3) and E1F1 molecular weights are 4032 and 559.2 kDa, respectively. The difference between the measured data and theoretical prediction is likely due to the error in the MALS measurement. The reported particle diameter for HBc and HFn are 35 and 12 nm, respectively. The measured sizes show a reasonable correlation with these (E1H1 diameter: 33.8 nm; E1F1: 14.6 nm). Lastly, TEM images confirmed that both E1F1 and E1H1 were

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

successfully assembled and confirmed the size difference, which is consistent with the hydrodynamic radius result (Figure 1E,F). Overall, characterization confirmed that the protein nanoparticles exhibited high quality, were properly formed, and were VLPs rather than merely virus-sized aggregates, giving confidence that biological results would be meaningful.

3.2. Antigen-Specific Antibody Titer Evaluation. Sera IgG responses to the short EBNA1 peptide were determined to compare the humoral immune response induced by E1F1 with that caused by E1H1. It was clearly observed from Figure 2A that, after the second boost (day 38), both E1H1 with and without an adjuvant produced five times higher titers than E1F1 with an adjuvant (***$p < 0.001$). In addition, the minor difference between E1H1 with and without the adjuvant (*$p < 0.5$) proved that non-adjuvanted E1H1 can induce a potent humoral immune response. This result reveals the advantage of HBc VLP as a potential vaccine carrier able to achieve a stronger humoral immune response than that for a ferritin vaccine carrier.

To identify whether the humoral or cell-mediated immunization response was dominant, IgG2a and IgG1 were compared.[32] IgG2a/IgG1 is an indicator, which defines that: Th1-favored immunization pathway if IgG2a/IgG1 < 1 and/or Th2-biased response if IgG2a/IgG1 > 1.[33] The Th1 response involves the secretion of cytokines, for example, IL-12, TNF-α, and IFN-γ. These cytokines initiate a cell-mediated response for growing cytotoxic T lymphocyte cells to combat viral infections.[30,34] The Th2 response regulates cytokines (IL-4, IL-5, and IL-6). These cytokines are important to initiate humoral immune responses.[30,35] From Figure 2B, IgG2a/IgG1 ratios for all the three groups (E1F1 with an adjuvant, E1H1 with, or without an adjuvant) were lower than 1, indicating a humoral immune response bias for all three groups.

3.3. Comparisons of Splenocyte Proliferative Responses. Proliferative responses induced by E1F1 and E1H1 were determined by the splenocyte proliferation assay. The level of splenocyte proliferative responses by using proliferation index (PI) can reflect the strength of the immune response induced by E1H1 or E1F1.[36] The individual PI for E1F1 with the adjuvant and for E1H1 with or without adjuvant were 2.0, 3.9, and 3.1 (Figure 3). It was also found that E1H1 with or without adjuvant induced a higher PI than E1F1 with the adjuvant (**$p < 0.01$). PI of E1H1 without the adjuvant is slightly lower than E1H1 with the adjuvant.

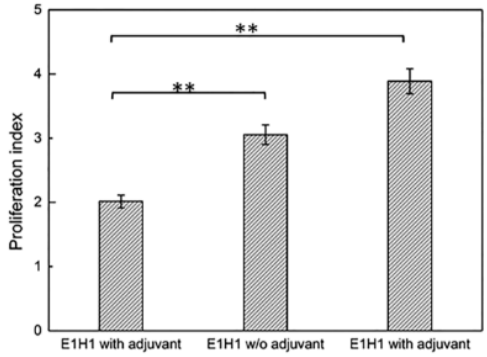


Figure 3. PI for E1H1 with the adjuvant, E1H1 w/o the adjuvant, and E1H1 with the adjuvant measured by a CCK-8 kit. Splenocytes were cultured at day 38 (10 days after second boost) and restimulated *ex vivo* with the EBNA1 antigen. Data are shown in mean SEM ($n = 6$) (*$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$).

3.4. Lymphocyte Activation. To compare the strength of humoral with cell-mediated immune responses induced by E1F1 and E1H1, the CD25 activation levels on $CD19^+$B cells, $CD4^+$T cells, and $CD8^+$T cells were found.[37] CD25 is a premature marker[37] to examine B- and T-cell activation level after stimulation by short EBNA1.[30] CD19, as a B-cell biomarker, is significant to regulate humoral responses.[38] The primary role for $CD4^+$T cells, T helper cells, is to drive on adaptive immune responses.[39] Figure 4A shows that $CD4^+$T cells identify peptides displayed on major histocompatibility complex (MHC) class II molecules, from antigen presenting cells (APCs).[40] Dendritic cells (DCs) are examples of APCs, which activate and control adaptive immune responses.[41] Compared with $CD4^+$T cells, $CD8^+$T cells mainly communicate peptides presented by MHC I class molecules.[40] After recognition and activation by EBNA1, $CD8^+$T cells secret cytokines such as TNF-α and IFN-γ correlated to the cell-mediated immune response.[30] Overall, the number of $CD19^+$B cells reflects the strength of the humoral immune response, while the $CD8^+$T cell number indicates the strength of the cell-mediated immune response.

In accordance with IgG titer results, the cellular level of $CD19^+$B cells for E1H1 with the adjuvant was the highest (Figure 4B), approximately 4 times higher than E1F1 with the adjuvant. There was no significant difference between E1H1 with and without adjuvants regarding the number of $CD19^+$B cells after activation. This result suggests E1H1 can induce a potent humoral immune response by itself without an adjuvant boost. The number of $CD4^+$T cells in three groups was similar (Figure 4C). $CD8^+$T cell number after activation for E1F1 was roughly five times higher than the other two groups (Figure 4D). This result suggests E1F1 with an adjuvant produces the strongest cell-mediated immune response.

3.5. Memory Lymphocyte T-Cell Differentiation. To test if E1H1 and E1F1 have the ability to confer long-term protection, memory lymphocyte T cells were further determined.[42] Memory T cells can be classified as the effector T cells (T_{EM}) and central memory T cells (T_{CM}).[43] T_{EM} cells work faster than T_{CM} cells, nevertheless, the effect period is shorter. Marked by CD44 and CD62L, the T_{EM} cell is represented as $CD44^{hi}$ $CD62L^{low}$, whereas the T_{CM} cell is represented as $CD44^{hi}$ $CD62L^{hi}$.[44] Figure 5A shows the flow cytometry cross gates for CD44 and CD62L. The detailed gating information is shown in Figure S1 (Supporting Information). Before the stimulation by the short peptide EBNA1, the number of $CD4^+$ and $CD8^+$ T_{CM} cells in the E1F1 group was two times higher than E1H1 with or without adjuvant groups (HBc carrier) (Figure 5B,C). After the stimulation by the short EBNA1 peptide, the number of T_{EM} cells increased as $CD4^+$ and $CD8^+$ T_{CM} cells were converted to T_{EM} cells. There were approximately 8% greater $CD4^+$ T_{EM} cells found in the E1F1 compared with E1H1 (***$p < 0.001$) (Figure 5D). It was found that approximately 5 and 3% greater $CD8^+$ T_{EM} cells in the E1F1 group than the E1H1 w/o adjuvant and E1H1 with the adjuvant, respectively (***$p < 0.001$) (Figure 5E). The percentage of $CD4^+$ and $CD8^+$ T_{CM} cells were significantly low (<1%) compared with $CD4^+$ and $CD8^+$ T_{EM} cells (>10%). These results were considered as valuable by comparing these with the research using a similar immunology study methods while the humoral immune

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

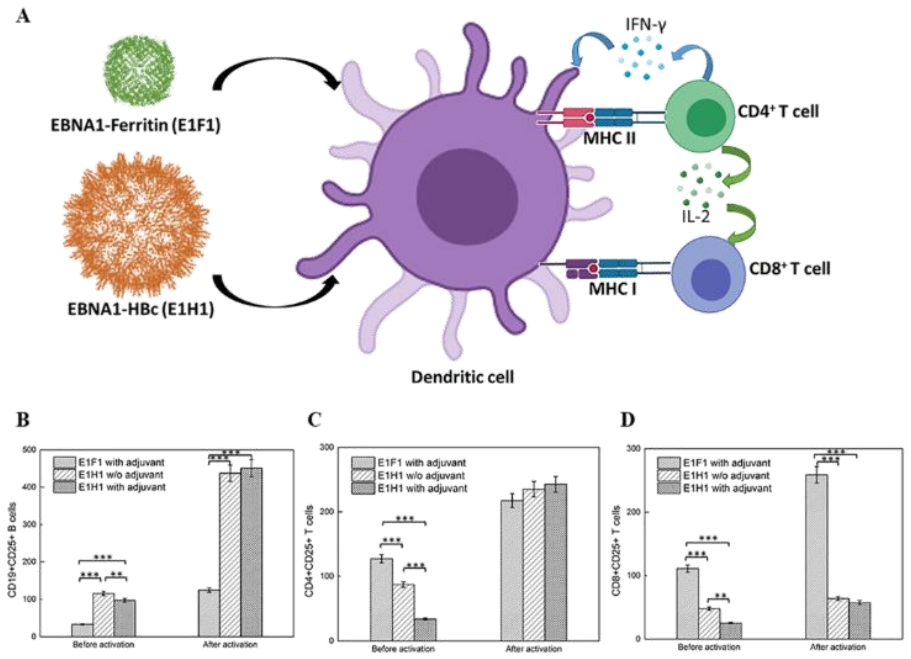


Figure 4. Activation of DC by EBNA1 antigens determined by flow cytometry. (A) Mimic DC processing diagram by Biorender. (B) Frequency of $CD25^+$ in $CD19^+$B cells. (C) $CD25^+$ in CD^{4+}T cells. (D) $CD25^+$ in $CD8^+$T cells. Data are shown in mean SEM ($n = 6$) (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

response was dominant (T_{CM} cells: ~1%; T_{EM} cells: ~10%).[31] Based on the above results, E1F1 induces a stronger cell-mediated immune response than E1H1 with or without adjuvants.

3.6. Adjuvant Adsorption Fraction Determination and Epitope Density Calculations. To check if E1F1 and E1H1 can be adsorbed by the aluminum hydroxide adjuvant, the adjuvant adsorption fraction was determined. E1F1 was completely adsorbed by the adjuvant (Table 1), whereas the average adsorption percentage for E1H1 was 54%. The particle size for E1F1 and E1H1 after being adsorbed by the aluminum hydroxide adjuvant was confirmed by Mastersizer (Figure S2, Supporting Information). The particle size distribution for E1F1 and E1H1 with the adjuvant was similar (Table S2, Supporting Information), and there was only a minor shift compared to the fresh aluminum hydroxide adjuvant size distribution.

To understand the differences in terms of epitope density, further calculations were performed. It is assumed that the effect of epitope density on the immunity responses is based on the difference in the antigen organization on the single protein particle.

Each single E1H1 protein particle carried 240 ($T = 4$) or 180 epitopes ($T = 3$), while each single E1F1 protein particle carrying 24 epitopes. The epitope density ratio of E1H1 and E1F1 was 10 ($T = 4$) or 7.5 ($T = 3$) (S4, Supporting Information). This demonstrates that the epitope density of E1H1 was always higher than E1F1.

4. DISCUSSION

For all three groups, namely, E1F1 with an adjuvant, E1H1 with, or without adjuvant, the dominant immune response was humoral. The EBNA1 short peptide (HPVGEADYFEY) itself belongs to $CD4^+$ T-cell epitope and primarily triggers Th1 responses,[45] which are responsible for activating the cell-mediated immune response. However, these were not consistent with our results. This immune response shift has been found in several previous studies,[46−48] mostly because of the effects of the adjuvant and vaccine carrier. Aluminum hydroxide has been proven to be one of the adjuvants that mainly induces a Th2 response to primarily activate the humoral immune response.[49−52] Therefore, the dominant immune response for E1F1 with the adjuvant depends on the aluminum hydroxide adjuvant. The HBC VLP itself has been known to induce a predominantly IgG1 antibody response (Th2).[53] Thus, E1H1 with or without adjuvant tends to be humoral immune response dominant.

Moreover, with the same antigen applied (EBNA1), HFn (E1F1), and HBc (E1H1) showed different immunogenicity. IgG titers demonstrated that E1H1 produced greater humoral immune responses than E1F1. The proliferation assay indicated that E1H1 enhanced the PI compared to E1F1. There are three possible reasons. In the first place, HBc-VLP has self-immunogenicity in the human body while HFn does not.[54] HFn-based vaccines are expected to be co-administrated with the adjuvant to boost the immune response.[10,12,14] By contrast, it has been confirmed by the IgG titer results that the adjuvant-free E1H1 induced a potent humoral immune response. This highlights the advantage of HBc-VLP as a vaccine carrier. The adjuvant-free formulation is significantly favored in vaccination because the aluminum adjuvant was found to have risks in serious immunological disorders.[55] Second, HBc (5040 kDa) is obviously a larger molecule than the HFn molecule (504 kDa) and also has larger particle size of 35 nm compared with 12 nm. Research by Chang *et al.* suggests that the protein nanoparticle vaccine size affects

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

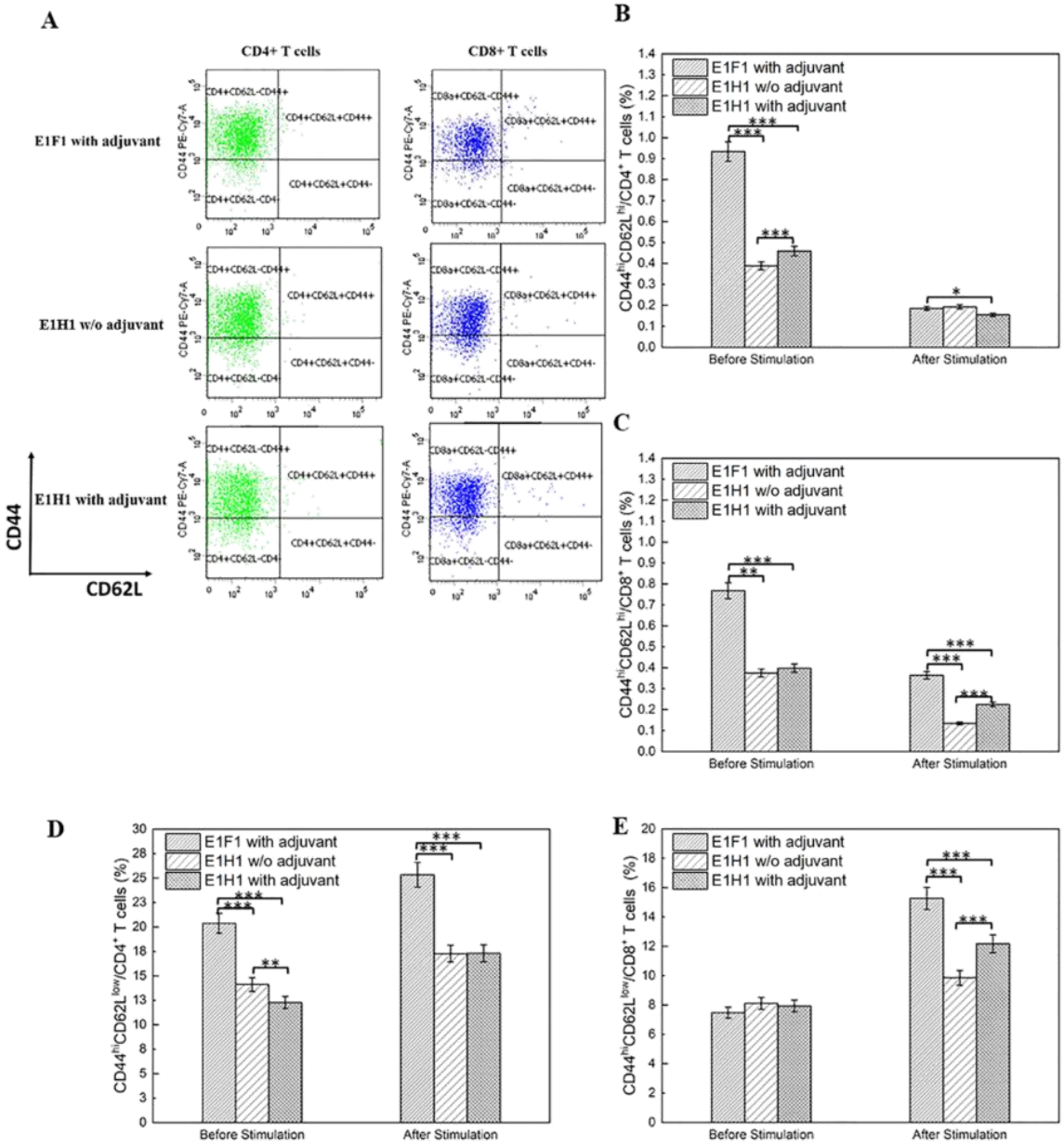


Figure 5. Frequencies of T_{CM} ($CD44^{hi}CD62L^{hi}$) and T_{EM} ($CD44^{hi}CD62L^{low}$) cells. (A) FACS plots. FACS plots are the representative of the mean percentage of six mice in each group. (B) $CD4^+$ T_{CM} cells (%). (C) $CD8^+$ T_{CM} cells (%). (D) $CD4^+$ T_{EM} cells. (E) $CD8^+$ T_{EM} cells (%). Data are shown in mean SEM ($n = 6$) (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

Table 1. Adjuvant Adsorption Percentage for E1F1 and E1H1

	adjuvant adsorption (%)				
sample name	1	2	3	average (%)	
EBNA1−ferritin	100	100	100	100	0
EBNA1−HBc	52	54	56	54	1.63

significantly the immune responses, especially in terms of DC processing.[56] The optimal nanoparticle size range to induce strong immune responses is from 40 nm to 3 μm.[57,58] Therefore, E1H1 (HBc based vaccine) is expected to induce stronger immune responses. Last but not least, the epitope density difference between E1F1 and E1H1 results in different immunogenicity. Studies have demonstrated that epitope density is closely associated with humoral immune responses specific to epitopes.[59−61] Findings showed that the humoral immune response was enhanced with increasing epitope density.[60] The study by Liu *et al.* demonstrated that fusion proteins (*S*-transferase) carrying higher M2e influenza epitopes ($M2e_8$) induced polyclonal antibodies with a greater average affinity constant and showed higher average avidity, compared

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

with the one of low epitope density ($M2e_1$ and $M2e_4$).[62] It can be concluded here that a single protein particle with high epitope density induces stronger immunogenicity compared with the one with low epitope density. Proven by experimental data *in vivo*, E1H1 produces a stronger humoral immune response than E1F1, which is consistent with the calculated data, E1H1 shows a higher epitope density than E1F1.

Cell-mediated immune responses for E1F1 and E1H1 were also compared in our study. Interestingly, from both lymphocyte activation and memory lymphocyte T-cell differentiation results, E1F1 induced an enhanced cell-mediated immune response than E1H1. This underlines the advantage of human ferritin in T cell immunology. It was noted by Lee *et al.* that, compared with other nanoparticle proteins (*E. coli* DNA binding protein, *Thermoplasma acidophilum* proteasome and HBV capsid),[63] a ferritin vaccine carrier fast drifts to lymph nodes (LNs) with a short incubation time (<1 min) and the accumulation of human heavy-chain ferritin in the LNs lasted for an adequately long time (6 days).[9,63] This is presumably caused due to the fact that human ferritin can interact with T lymphocytes.[9] It has been found that human ferritin has a strong binding affinity to mucin domain-2 (TIM-2) and T-cell immunoglobulin to control cellular immunity.[9,64,65] The prolonged accumulation of human ferritin ensures that antigens can be efficiently exposed to immune cells in LNs,[9] which may in turn enhance the immunogenicity of a vaccine.[66] This may explain why E1F1 (HFn carrier) showed a stronger cell-immune response than E1H1 (HBc vaccine carrier).

5. CONCLUSIONS

This study proves the success of developing HFn and HBc as vaccine carriers by inserting the EBNA1 short peptide to potentially combat EBV infections. Both nanoparticles prompt significant humoral and cell-mediated immune responses. Importantly, EBNA1−HBc (E1H1) can induce immune responses without the adjuvant. The HBc vaccine carrier (E1H1) also has a higher epitope density than the ferritin carrier (E1F1). This explained well why E1H1 produced more than five times the antibody titer and two times stronger proliferative responses than E1F1. On the other hand, E1F1 utilized the benefit of HFn on strong binding to the T-cell immunoglobulin. The prolonged accumulation of HFn in LNs likely ensured that sufficient epitopes can be processed to the immune cells. This in turn induced a stronger cell-mediated immune response than E1H1. Overall, this study has shown a successful application of HFn and HBc as vaccine carriers. This initial proof-of-concept study can potentially lead to further vaccine development against urgent viral challenges, such as EBV and new emerging viruses.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsabm.1c00738.

Flow cytometry gates showing CD19, CD4, and CD8; particle size for the aluminum hydroxide adjuvant, E1F1 with an adjuvant and E1H1 with an adjuvant by Mastersizer; amino acid sequence of E1F1 and E1H1; particle size by Mastersizer; calculation of injected antigen dose; and calculation of epitope density ratio (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jingxiu Bi − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0001-7056-8572; Email: Jingxiu.bi@adelaide.edu.au

Authors

Yiran Qu − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia;* orcid.org/0000-0002-3536-925X

Bingyang Zhang − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Yingli Wang − *Shanxi University of Traditional Chinese Medicine, Taiyuan, Shanxi 030024, China*

Shuang Yin − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Yan Sun − *Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;* orcid.org/0000-0001-5256-9571

Anton Middelberg − *School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsabm.1c00738

Author Contributions

‖Y.Q and B.Z. have the same contribution to this paper

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully appreciate Shanxi Education Science "1331 project" special research project (Research and Development of Traditional Chinese Medicine Micro-emulsion and New Biological Preparation).

■ REFERENCES

(1) Chuan, Y. P.; Lua, L. H. L.; Middelberg, A. P. J. High-level expression of soluble viral structural protein in Escherichia coli. *J. Biotechnol.* **2008**, *134*, 64−71.

(2) Liew, M. W. O.; Rajendran, A.; Middelberg, A. P. J. Microbial production of virus-like particle vaccine protein at gram-per-litre levels. *J. Biotechnol.* **2010**, *150*, 224−231.

(3) Butkovich, N.; Li, E.; Ramirez, A.; Burkhardt, A. M.; Wang, S. W. Advancements in protein nanoparticle vaccine platforms to combat infectious disease. *WIREs Nanomed. Nanobiotechnol.* **2021**, *13*, No. e1681.

(4) Demchuk, A. M.; Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology AdvanceBiotechnol. Adv.* **2020**, *41*, 107547.

(5) Bachmann, M. F.; Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat. Rev. Immunol.* **2010**, *10*, 787−796.

(6) Mohsen, M. O.; Zha, L.; Cabral-Miranda, G.; Bachmann, M. F. Major findings and recent advances in virus−like particle (VLP)-based vaccines. *Semin. Immunol.* **2017**, *34*, 123−132.

(7) Lee, L.; Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomed. Nanotechnol. Biol. Med.* **2006**, *2*, 137−149.
(8) Doll, T. A. P. F.; Raman, S.; Dey, R.; Burkhard, P. Nanoscale assemblies and their biomedical applications. *J. R. Soc. Interface* **2013**, *10*, 20120740.
(9) Lee, B.-R.; Ko, H. K.; Ryu, J. H.; Ahn, K. Y.; Lee, Y.-H.; Oh, S. J.; Na, J. H.; Kim, T. W.; Byun, Y.; Kwon, I. C.; Kim, K.; Lee, J. Engineered Human Ferritin Nanoparticles for Direct Delivery of Tumor Antigens to Lymph Node and Cancer Immunotherapy. *Sci. Rep.* **2016**, *6*, 35182.
(10) López-Sagaseta, J.; Malito, E.; Rappuoli, R.; Bottomley, M. J. Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 58−68.
(11) Qu, Y.; Wang, L.; Yin, S.; Zhang, B.; Jiao, Y.; Sun, Y.; Middelberg, A.; Bi, J. Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments. *J. Phys. Chem. B* **2021**, *125*, 3830−3842.
(12) Kanekiyo, M.; Wei, C.-J.; Yassine, H. M.; McTamney, P. M.; Boyington, J. C.; Whittle, J. R. R.; Rao, S. S.; Kong, W.-P.; Wang, L.; Nabel, G. J. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **2013**, *499*, 102.
(13) Li, C. Q.; Soistman, E.; Carter, D. C. Ferritin nanoparticle technology. A new platform for antigen presentation and vaccine development. *Ind. Biotechnol.* **2006**, *2*, 143−147.
(14) Wang, Z.; Xu, L.; Yu, H.; Lv, P.; Lei, Z.; Zeng, Y.; Liu, G.; Cheng, T. Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomater. Sci.* **2019**, *7*, 1794−1800.
(15) Zhen, Z.; Tang, W.; Todd, T.; Xie, J. Ferritins as nanoplatforms for imaging and drug delivery. *Expet Opin. Drug Deliv.* **2014**, *11*, 1913−1922.
(16) Hempstead, P. D.; Yewdall, S. J.; Fernie, A. R.; Lawson, D. M.; Artymiuk, P. J.; Rice, D. W.; Ford, G. C.; Harrison, P. M. Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution11Edited by R. Huber. *J. Mol. Biol.* **1997**, *268*, 424−448.
(17) Broxmeyer, H.; Williams, D.; Geissler, K.; Hangoc, G.; Cooper, S.; Bicknell, D.; Levi, S.; Arosio, P. Suppressive effects in vivo of purified recombinant human H-subunit (acidic) ferritin on murine myelopoiesis. *Blood* **1989**, *73*, 74−79.
(18) Wang, W.; Knovich, M. A.; Coffman, L. G.; Torti, F. M.; Torti, S. V. Serum ferritin: Past, present and future. *Biochim. Biophys. Acta* **2010**, *1800*, 760−769.
(19) Broxmeyer, H.; Lu, L.; Bicknell, D.; Williams, D.; Cooper, S.; Levi, S.; Salfeld, J.; Arosio, P. The influence of purified recombinant human heavy-subunit and light-subunit ferritins on colony formation in vitro by granulocyte-macrophage and erythroid progenitor cells. *Blood* **1986**, *68*, 1257−1263.
(20) Gamvrellis, A.; Leong, D.; Hanley, J. C.; Xiang, S. D.; Mottram, P.; Plebanski, M. Vaccines that facilitate antigen entry into dendritic cells. *Immunol. Cell Biol.* **2004**, *82*, 506−516.
(21) Slütter, B.; Jiskoot, W. Sizing the optimal dimensions of a vaccine delivery system: a particulate matter. *Expet Opin. Drug Deliv.* **2016**, *13*, 167−170.
(22) Zhao, Q.; Li, S.; Yu, H.; Xia, N.; Modis, Y. Virus-like particle-based human vaccines: quality assessment based on structural and functional properties. *Trends Biotechnol.* **2013**, *31*, 654−663.
(23) Golmohammadi, R.; Fridborg, K.; Bundule, M.; Valegård, K.; Liljas, L. The crystal structure of bacteriophage Qβ at 3.5 å resolution. *Structure* **1996**, *4*, 543−554.
(24) Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W. The Crystal Structure of the Human Hepatitis B Virus Capsid. *Mol. Cell* **1999**, *3*, 771−780.
(25) Baltabekova, A. Z.; Shagyrova, Z. S.; Kamzina, A. S.; Voykov, M.; Zhiyenbay, Y.; Ramanculov, E. M.; Shustov, A. V. SplitCore Technology Allows Efficient Production of Virus-Like Particles Presenting a Receptor-Contacting Epitope of Human IgE. *Mol. Biotechnol.* **2015**, *57*, 746−755.
(26) Walker, A.; Skamel, C.; Nassal, M. SplitCore: An exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. *Sci. Rep.* **2011**, *1*, 5.
(27) Li, Z.; Wei, J.; Yang, Y.; Ma, X.; Hou, B.; An, W.; Hua, Z.; Zhang, J.; Li, Y.; Ma, G.; Zhang, S.; Su, Z. Strong hydrophobicity enables efficient purification of HBc VLPs displaying various antigen epitopes through hydrophobic interaction chromatography. *Biochem. Eng. J.* **2018**, *140*, 157−167.
(28) Farrell, P. J. Epstein−Barr Virus and Cancer. *Annu. Rev. Pathol.: Mech. Dis.* **2019**, *14*, 29−53.
(29) Pattle, S. B.; Farrell, P. J. The role of Epstein−Barr virus in cancer. *Expet Opin. Biol. Ther.* **2006**, *6*, 1193−1205.
(30) Qu, Y.; Zhang, B.; Wang, Y.; Yin, S.; Pederick, J. L.; Bruning, J. B.; Sun, Y.; Middelberg, A.; Bi, J. Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* **2021**, *39*, 4830−4841.
(31) Guo, F.; Liu, Y.; Zhang, C.; Wang, Q.; Wang, L.; Gao, Y.; Bi, J.; Wang, H.; Su, Z. Prompt and Robust Humoral Immunity Elicited by a Conjugated Chimeric Malaria Antigen with a Truncated Flagellin. *Bioconjugate Chem.* **2018**, *29*, 761−770.
(32) Hammarström, L.; Smith, C. I. IgG subclass changes in response to vaccination. *Monogr. Allergy* **1986**, *19*, 241−252.
(33) Lubeck, M. D.; Steplewski, Z.; Baglia, F.; Klein, M. H.; Dorrington, K. J.; Koprowski, H. The interaction of murine IgG subclass proteins with human monocyte Fc receptors. *J. Immunol.* **1985**, *135*, 1299.
(34) Williams, M. A.; Bevan, M. J. Effector and memory CTL differentiation. *Annu. Rev. Immunol.* **2007**, *25*, 171−192.
(35) Poston, R. N.; Morgan, R. S. Interactions between soluble IgG, complement and cells in lymphocyte and monocyte ADCC. *Immunology* **1983**, *50*, 461.
(36) Datta, S.; Sarvetnick, N. Lymphocyte proliferation in immune-mediated diseases. *Trends Immunol.* **2009**, *30*, 430−438.
(37) Driesen, J.; Popov, A.; Schultze, J. L. CD25 as an immune regulatory molecule expressed on myeloid dendritic cells. *Immunobiology* **2008**, *213*, 849−858.
(38) Wang, K.; Wei, G.; Liu, D. CD19: a biomarker for B cell development, lymphoma diagnosis and therapy. *Exp. Hematol. Oncol.* **2012**, *1*, 36.
(39) Lederman, S.; Yellin, M. J.; Krichevsky, A.; Belko, J.; Lee, J. J.; Chess, L. Identification of a novel surface protein on activated CD4+ T cells that induces contact-dependent B cell differentiation (help). *J. Exp. Med.* **1992**, *175*, 1091−1101.
(40) Laidlaw, B. J.; Craft, J. E.; Kaech, S. M. The multifaceted role of CD4+ T cells in CD8+ T cell memory. *Nat. Rev. Immunol.* **2016**, *16*, 102−111.
(41) Cohn, L.; Delamarre, L. l. Dendritic cell-targeted vaccines. *Front. Immunol.* **2014**, *5*, 255.
(42) Krzych, U.; Zarling, S.; Pichugin, A. Memory T cells maintain protracted protection against malaria. *Immunol. Lett.* **2014**, *161*, 189−195.
(43) Sallusto, F.; Geginat, J.; Lanzavecchia, A. Central Memory and Effector Memory T Cell Subsets: Function, Generation, and Maintenance. *Annu. Rev. Immunol.* **2004**, *22*, 745−763.
(44) Kaech, S. M.; Wherry, E. J.; Ahmed, R. Effector and memory T-cell differentiation: implications for vaccine development. *Nat. Rev. Immunol.* **2002**, *2*, 251−262.
(45) Bickham, K.; Münz, C.; Tsang, M. L.; Larsson, M.; Fonteneau, J.-F.; Bhardwaj, N.; Steinman, R. EBNA1-specific CD4+ T cells in healthy carriers of Epstein-Barr virus are primarily Th1 in function. *J. Clin. Invest.* **2001**, *107*, 121−130.
(46) Korsholm, K. S.; Petersen, R. V.; Agger, E. M.; Andersen, P. T-helper 1 and T-helper 2 adjuvants induce distinct differences in the magnitude, quality and kinetics of the early inflammatory response at the site of injection. *Immunology* **2010**, *129*, 75−86.
(47) Ciabattini, A.; Pettini, E.; Fiorino, F.; Pastore, G.; Andersen, P.; Pozzi, G.; Medaglini, D. Modulation of Primary Immune Response by Different Vaccine Adjuvants. *Front. Immunol.* **2016**, *7*, 427.

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

(48) Awate, S.; Babiuk, L. A.; Mutwiri, G. Mechanisms of action of adjuvants. *Front. Immunol.* **2013**, *4*, 114.
(49) Bomford, R.; Stapleton, M.; Winsor, S.; McKnight, A.; Andronova, T. The Control of the Antibody Isotype Response to Recombinant Human Immunodeficiency Virus gp120 Antigen by Adjuvants. *AIDS Res. Hum. Retrovir.* **1992**, *8*, 1765−1771.
(50) Comoy, E.; Capron, A.; Thyphronitis, G. In vivo induction of type 1 and 2 immune responses against protein antigens. *Int. Immunol.* **1997**, *9*, 523−531.
(51) Jiao, X.-d.; Cheng, S.; Hu, Y.-h.; Sun, L. Comparative study of the effects of aluminum adjuvants and Freund's incomplete adjuvant on the immune response to an Edwardsiella tarda major antigen. *Vaccine* **2010**, *28*, 1832−1837.
(52) Hogenesch, H. Mechanism of immunopotentiation and safety of aluminum adjuvants. *Front. Immunol.* **2013**, *3*, 406.
(53) Sominskaya, I.; Skrastina, D.; Dislers, A.; Vasiljev, D.; Mihailova, M.; Ose, V.; Dreilina, D.; Pumpens, P. Construction and Immunological Evaluation of Multivalent Hepatitis B Virus (HBV) Core Virus-Like Particles Carrying HBV and HCV Epitopes. *Clin. Vaccine Immunol.* **2010**, *17*, 1027.
(54) Tan, A.; Koh, S.; Bertoletti, A. Immune Response in Hepatitis B Virus Infection. *Cold Spring Harbor Perspect. Med.* **2015**, *5*, a021428.
(55) Tomljenovic, L.; Shaw, C. A. Aluminum vaccine adjuvants: are they safe? *Curr. Med. Chem.* **2011**, *18*, 2630−2637.
(56) Chang, T. Z.; Stadmiller, S. S.; Staskevicius, E.; Champion, J. A. Effects of ovalbumin protein nanoparticle vaccine size and coating on dendritic cell processing. *Biomater. Sci.* **2017**, *5*, 223−233.
(57) Fifis, T.; Gamvrellis, A.; Crimeen-Irwin, B.; Pietersz, G. A.; Li, J.; Mottram, P. L.; McKenzie, I. F. C.; Plebanski, M. Size-dependent immunogenicity: therapeutic and protective properties of nano-vaccines against tumors. *J. Immunol.* **2004**, *173*, 3148−3154.
(58) Tran, K. K.; Shen, H. The role of phagosomal pH on the size-dependent efficiency of cross-presentation by dendritic cells. *Biomaterials* **2009**, *30*, 1356−1362.
(59) Jegerlehner, A.; Storni, T.; Lipowsky, G.; Schmid, M.; Pumpens, P.; Bachmann, M. F. Regulation of IgG antibody responses by epitope density and CD21-mediated costimulation. *Eur. J. Immunol.* **2002**, *32*, 3305−3314.
(60) Liu, W.; Peng, Z.; Liu, Z.; Lu, Y.; Ding, J.; Chen, Y.-H. High epitope density in a single recombinant protein molecule of the extracellular domain of influenza A virus M2 protein significantly enhances protective immunity. *Vaccine* **2004**, *23*, 366−371.
(61) Liu, W.; Chen, Y.-H. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. *Eur. J. Immunol.* **2005**, *35*, 505−514.
(62) Liu, W.; Chen, Y.-H. High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. *Eur. J. Immunol.* **2005**, *35*, 505−514.
(63) Lee, B.-R.; Ko, H. K.; Ryu, J. H.; Ahn, K. Y.; Lee, Y.-H.; Oh, S. J.; Na, J. H.; Kim, T. W.; Byun, Y.; Kwon, I. C.; Kim, K.; Lee, J. Engineered Human Ferritin Nanoparticles for Direct Delivery of Tumor Antigens to Lymph Node and Cancer Immunotherapy. *Sci. Rep.* **2016**, *6*, 35182.
(64) de Souza, A. J.; Kane, L. P. Immune regulation by the TIM gene family. *Immunol. Res.* **2006**, *36*, 147−156.
(65) Fargion, S.; Fracanzani, A.; Brando, B.; Arosio, P.; Levi, S.; Fiorelli, G. Specific binding sites for H-ferritin on human lymphocytes: modulation during cellular proliferation and potential implication in cell growth control. *Blood* **1991**, *78*, 1056−1061.
(66) Bachmann, M. F.; Beerli, R. R.; Agnellini, P.; Wolint, P.; Schwarz, K.; Oxenius, A. Long-lived memory CD8+ T cells are programmed by prolonged antigen exposure and low levels of cellular activation. *Eur. J. Immunol.* **2006**, *36*, 842−854.

https://doi.org/10.1021/acsabm.1c00738
ACS Appl. Bio Mater. 2021, 4, 7147−7156

CHAPTER 8 CONCLUSIONS AND FUTURE RESEARCH

8.1 Conclusions

The current COVID-19 pandemic has increased urgency to develop reliable vaccines. Traditional vaccines, for example, live-attenuated vaccine strains or inactivated killed vaccine pathogens, have disadvantages of imposing high risks of reversion into disease-causing viruses, or inducing weak immune responses. Therefore, modern vaccines using nanoparticles are developed. EBC vaccine, as one of examples, is practically promising because it has: 1) a precise control over the structure and functionality; and 2) mimics viral structures to boost antigen presentation and strong immunogenicity. EBC vaccine is structured with three (3) main components, NPC, linker and epitope.

In this Thesis, HFn was the main interested NPC to be developed as an EBC vaccine. Based on findings from **Chapter 2**, it was found that HFn exhibits a number of advantages including a: 1) highly organized unique protein nanocage in order to convenient display of epitopes; 2) strong protein stability against thermal and chemical denaturants; 3) highly safe self-assembled human proteins with low toxicity, good biocompatibility and biodegradability; and 4) ability to induce potent cell-mediated immune response by strong binding to T cell immunoglobulin and TIM-2. This Thesis follows the process: molecular design → protein purification and characterization → vaccine immunology study.

Key conclusions from this thesis work include:

1. With applied same epitope (EBNA1), both MDS and experimental data confirmed that engineered ferritin with N-terminus insertion (E1F1) was more hydrophobic and less stable against thermal, pH and chemical denaturants than that from C-terminus insertion (F1E1). The less hydrophobic SASA provided F1E1 less interactions between non-

polar residues with water molecules. Therefore, a more stable protein structure was observed (**Chapter 3**).

2. Flexible linker length molecular design can be used to significantly boost protein stability. N-terminus insertion with long linker (15 residues) ($E_1L_{15}F_1$) exhibited less protein surface hydrophobicity because of steric shielding, and greater stability against thermal and chemical denaturants (GdnHCl and urea) than that with short linker (3 residues) ($E_1L_3F_1$). C-terminus insertion with long linker (15 residues) ($F_1L_{15}E_1$) had similar hydrophobicity to short linker (3 residues) ($F_1L_3E_1$), and greater thermal stability together with GdnHCl-denaturant stability. However, stability against urea denaturant for both two groups was similar (**Chapter 4**).

3. C-terminus of HFn has shown significant potential for maintaining protein stability, therefore, variant study of helix E in C-terminus was performed. The hot spots in helix E were predicted by MDS as aa 167 and aa171. Five (5) variants (C1, C2, C3, C4 and C5) were constructed by altering electrostatic or hydrophobic interfaces. It was found that variants C1, C2 and C5 can be folded properly with good conformational stability. However, variant C3 exhibited significant proportion of aggregates, and variant C4 was not folded properly. All of five variants had decreased thermal stability confirmed by both MDS and experimental data (**Chapter 5**).

4. Immunology study confirmed that the insertion site is affected significantly on vaccine immunogenicity, especially in terms of cell-mediated immune response. C-terminus insertion (F1E1) presented stronger proliferative responses and cell-mediated immune response than N-terminus insertion (E1F1) (**Chapter 6**).

5. Immunology study demonstrated that vaccine immunogenicity was significantly affected by particle size and molecular weight of NPCs. E1F1 utilised the benefit of HFn on strong binding to T cell immunoglobulin to induce stronger cell-mediated immune response than E1H1. However, E1H1 (HBc vaccine carrier) induced stronger humoral immune response and proliferative response than E1F1 (HFn vaccine carrier) because of higher epitope density (**Chapter 7**).

8.2 Future research directions

This Thesis work has underscored the importance of vaccine rational design by considering epitope-insertion sites, linker length and hot spots replacements in order to produce highly stable and efficacious HFn-based vaccines. The preliminary immunology study has revealed the great potential of HFn as vaccine carrier. Future research directions will necessarily include:

1. Flexible loop region is another promising insertion site when applying HFn as vaccine carrier. Corresponding immunology study is expected to get direct comparisons of immunogenicity amongst these 3 insertion sites (N-terminus, C-terminus and flexible loop region).

2. Flexible linker composition is the other interesting aspect to affect protein stability. By altering amino acid compositions, effect of flexible linker with different composition on protein stability can be investigated.

3. A focus on using MDS to predict replacement of which amino acids with hot spots to boost protein stability. This will build on the finding that this Thesis has shown that replacement of hot spots with positive-charged, non-charged, more hydrophobic or more hydrophilic residues, all significantly affect protein folding and stability.

4. DSC analysis indicated multiple denaturing temperature values for some HFn derivatives. Future analysis based on SEC-MALS can be coupled with DSC, CD and fluorescence spectrometry to generate in-depth understandings of intermediate stages of HFn-based nanoparticles during thermal denaturation.

5. Immunity experiments including stimulating by the full length of EBNA1 antigen

together with virus neutralisation assay to confirm the efficacy of vaccines.

6. A more detailed cell immunity examination for HFn based vaccine to confirm that HFn possibly regulates cell immunity more efficiently than HBc.

APPENDIX REVIEW PAPER

Statement of Authorship

Title of Paper	Using molecular dynamics simulation (MDS) to predict performance of Epitope-based chimeric vaccine: a critical review
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Qu, Y.; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Using molecular dynamics simulation (MDS) to predict performance of Epitope-based chimeric vaccine: a critical review. *Journal of Chemical Information and Modeling.*

Principal Author

Name of Principal Author (Candidate)	Yiran Qu		
Contribution to the Paper	Critical review analysis; writing manuscript		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth Davey		
Contribution to the Paper	Proof read the manuscript		
Signature		Date	11/10/2021

Name of Co-Author	Yan Sun
Contribution to the Paper	Proof read the manuscript

Signature		Date	20/09/2021

Name of Co-Author	Anton Middelberg
Contribution to the Paper	Proof read the manuscript
Signature	Date 13/10/21

Name of Co-Author	Jingxiu Bi
Contribution to the Paper	Proof read the manuscript
Signature	Date 12/10/2021

Using molecular dynamics simulation (MDS) to predict performance of Epitope-based chimeric vaccine: a critical review

*Yiran Qu, Kenneth Davey, Yan Sun, Anton Middelberg, and Jingxiu Bi**

Y. Qu, A/Prof. Dr. K. Davey, A/Prof. Dr. J. Bi

School of Chemical Engineering and Advanced Materials, The University of Adelaide, SA 5005, Australia

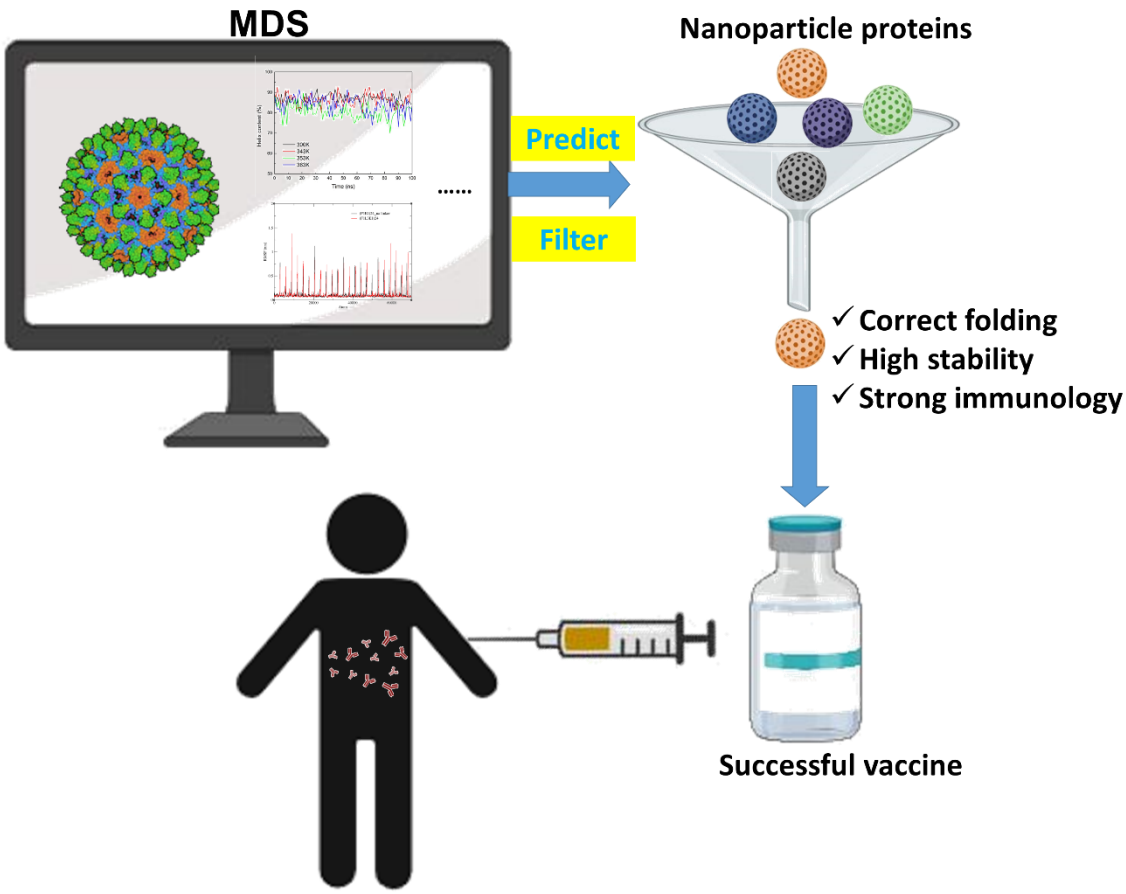
Prof. Dr. A. Middelberg

Division of Research and Innovation, The University of Adelaide, SA 5005, Australia

Prof. Dr. Y. Sun

Department of Biochemical Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China

TOC



KEYWORDS

Molecular dynamics simulation (MDS), nanoparticle protein, vaccine, epitope immunogenicity, protein structure prediction, protein stability.

Abstract

The COVID-19 pandemic has underscored the importance of reliable vaccines. Protein-based vaccines are attractive because of efficacy and safety. Epitope-based chimeric vaccines have attracted significant research interest because it exhibits: 1) a wide range of size from 5 to 100 nm; 2) cost-effective production, and; 3) highly repetitive structures to induce immune response(s). The practical challenge however is to carefully design innovative epitope-based chimeric vaccine, rather than rely on time-consuming and costly experiments. Molecular dynamics simulation(s) (MDS) is an emerging and important computational tool with application to vaccine development. Despite this, the use of MDS to aid chimeric vaccine nanoparticle design has not been reviewed in detail. Here we present a timely review to provide a critical assessment of use of MDS in design of nanoparticle protein-based vaccines. We focus sequentially on: 1) epitope immunogenicity prediction; 2) predicting fusion protein structures, and; 3) evaluating protein stability. We show that MDS can be used judiciously to design a highly stable and strongly immunogenic vaccine that can be readily examined later experimental validation(s). We conclude that MDS can significantly boost confidence in vaccine manufacture. Findings will be significant to researchers in simulation methods to guide practical vaccine development.

A1 Introduction

Timely and reliable vaccine development is important. This is underscored by the recent COVID-19 pandemic. The practical challenge is to design and develop innovative protein-based vaccines to replace traditional ones, such as live attenuated vaccines and inactivated pathogens, to increase immunogenicity and safety [1,2]. The overarching aims with innovative epitope-based chimeric (Ebc) vaccine include, to induce strong immune responses, be safe and to remain stable.

EBC vaccines consists of two (2) parts, 1) nanoparticle protein cage (NPC), and; 2) inserted epitope; these are shown schematically in Figure A1. Ebc are advantageous because they exhibit, 1) a wide range of sizes from 5 to 100 nm [3], 2) cost-effective production,[4] and; 3) highly repetitive structures to induce immune response(s) [4-6].

There are however significant practical challenges to development of EBC vaccines [1,7-9]. Development of vaccines includes but is not limited to: 1) molecular design/selection of epitopes; 2) design of energy-minimized fusion protein structures; 3) protein expression by host cell lines; 4) protein purification, and; 5) *in vitro* and *in vivo* testing.

Significant time-consuming and costly experiments are needed to develop desirable vaccine candidates. This is because there are many unknown factors, including epitope immunogenicity and protein stability. The result is often a low rate of successful vaccine production [10].

Molecular dynamics simulation (MDS) is an emerging design tool that appears to reduce the need for experimental testing and has potential to increase the likely success rate in vaccine development.

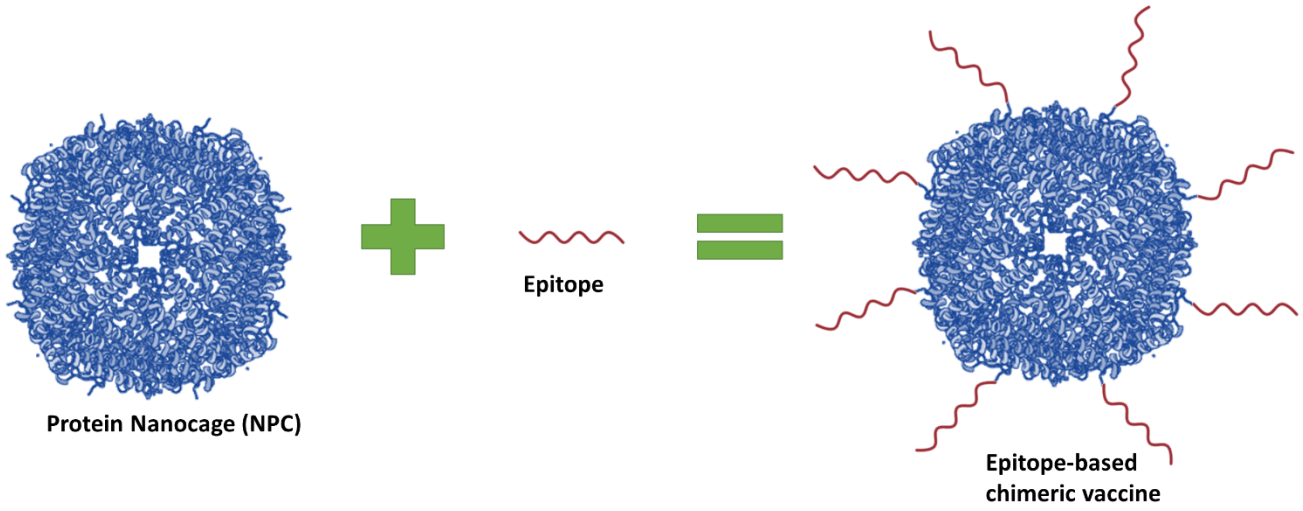


Figure A1 Schematic of epitope-based chimeric vaccine consisting of protein nanocage and epitope in BioRender©.

MDS is used to determine physical movement of three dimensional (3D) particles in a given dynamic environment and an applied force field. Since parallel developments in high-performance computing (HPC), large protein molecules can now be investigated *in silico* using MDS. Recent applications of MDS in protein field focus on prediction of peptides and protein structures [11-13], molecular docking and drug design [14-16], and prediction of protein stability [17,18].

A schematic showing how MDS is used to design and predict performance of proposed nanoparticle protein-based vaccines is given as Figure A2. It is seen from the figure there are 3 serial steps, namely, prediction of: M1) epitope immunogenicity; M2) protein structure, and; M3) protein stability, via *in silico* analyses.

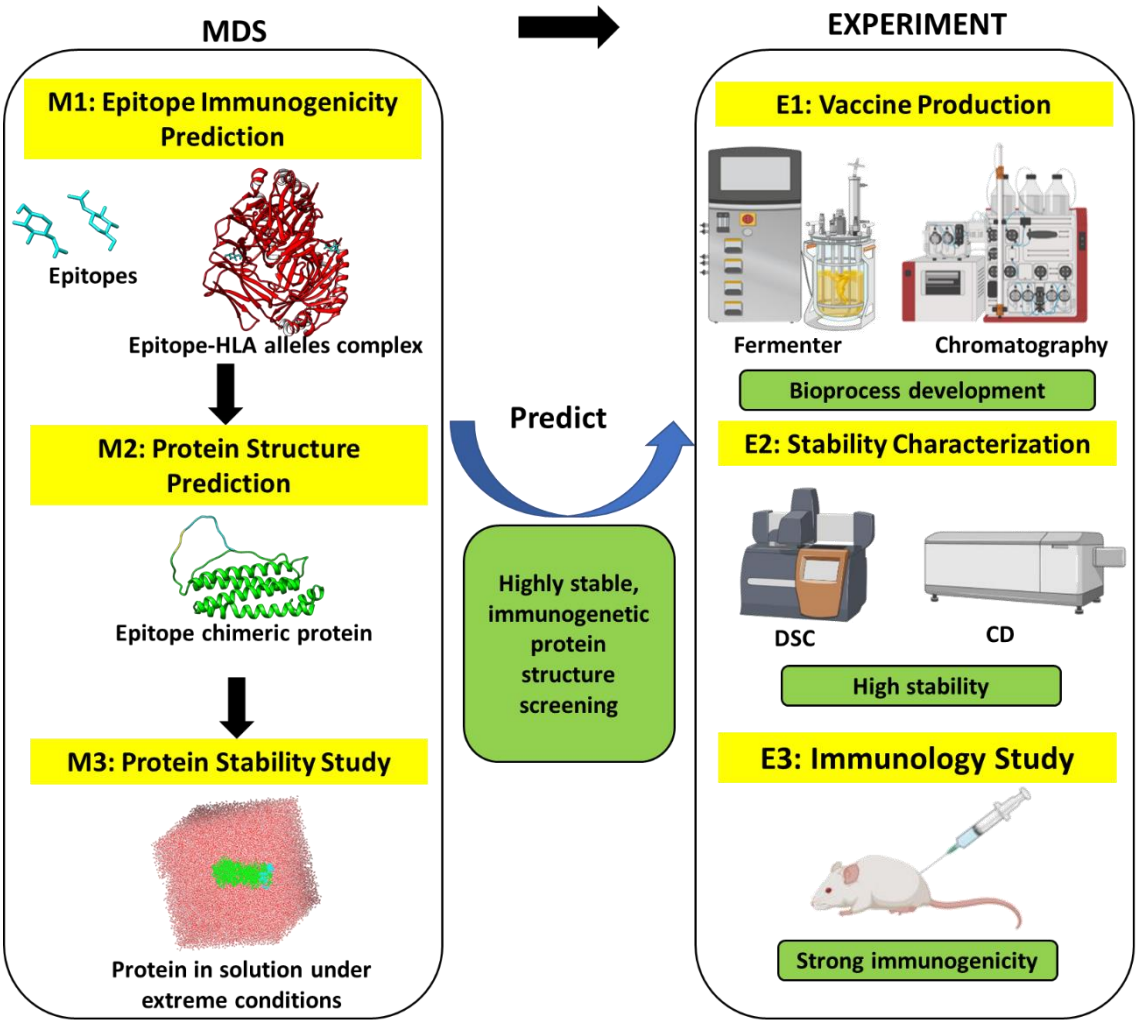


Figure A2 Schematic of Molecular Dynamics Simulation (MDS) for monitoring performance of nanoparticle protein-based vaccines in which DSC = Differential Scanning Calorimetry and CD = Circular Dichroism. Protein structures were created in UCSF Chimera. Cartoons of equipment were prepared in BioRender©.

Prediction of epitope immunogenicity by MDS (Figure A2, M1) filters the suitable epitope which is, most, 1) specific to a target disease, and; 2) efficacious. Epitopes are evaluated *in silico* by calculating the binding energy between epitopes and receptors. The lowest binding energy indicates the strongest binding with receptors. MDS is used to filter the epitope-receptor complex with greatest stability. The selected epitope is the most likely to induce significant

immune response [19]. Step M1 significantly improves success rates in later experimental immunological studies (Figure A2, E3) [19]. Step M2 is achieved via construction of energy-minimized molecular-structure for nanoparticle proteins (NPs) inserted with epitopes. If the crystal structure of NPCs has been registered in the Protein Data Bank (PDB) [20], M2 can be readily applied following assessing of structural stability data. However, where NPCs cannot be found in PDB, and limited molecular structural information is available, MDS becomes a critical tool to simulate protein folding and to predict energy-minimized protein structure with insertion of the epitopes [21]. Step M2 ensures protein structural resonance to support an optimal vaccine production (Figure A2, E1).

Step 3, M3 stability studies in MDS, have been found to be highly consistent with experimental data. This consistency between MDS prediction and experiment is advantageous to design the most stable vaccine with reduced time and cost. For example, protein thermal stability *in silico* agreed well with experimentally determined protein melting temperature using differential scanning calorimetry (DSC) [22,23]. MDS can provide protein secondary structure *in silico* information.[11,24] This is advantageous as it can replace experiments using circular dichroism (CD) [25, 26]. Protein stability studies via MDS achieves the same purpose, and it gives the most stable structures for experiment (E2) with both optimized time and cost.

It is clear that there is therefore practical benefit in using MDS to predict performance of nanoparticle protein-based vaccine. Despite this however it has not been reviewed in detail.

Here for the first time we present a timely review to provide a critical assessment of uses of MDS for designing nanoparticle protein-based vaccine. We focus sequentially on: 1) epitope immunogenicity prediction, M1; 2) predicting fusion protein structures, M2, and; 3) evaluating

protein stability, M3

We use case studies to show that when used judiciously, MDS can be relied on in the design of highly stable and strongly immunogenic vaccine, which can be readily identified for later experimental validation(s). We conclude that MDS can significantly boost confidence in vaccine manufacture. We discuss future application and challenges arising from MDS. We expect findings to be significant to researchers in simulation methods to guide practical vaccine development.

A2 Prediction of epitope immunogenicity

It is the most important in vaccination to achieve high immunogenicity by inducing potent immune responses. Therefore, selection, optimization and validation of suitable epitopes are important to ensure efficacy and safety.

A flowchart for the method to epitope design and immunogenicity prediction via computation is presented as Figure A3 [27-31]. This shows serial steps including: 1) viral strain selection, 2) protein sequence preparation, 3) epitope prediction, 4) vaccine structure construction, 5) allergenicity and toxicity, 6) population convergence analysis, 7) molecular docking analysis, and; 8) MDS.

Prior to molecular docking analyses, a number of databases are available to predict and analyze epitopes including IEDB [32], SYEPEITHI [33] and ProPred1 [34]. These databases are all designed to screen and predict epitopes of interest for a target disease [27-31]. The immunogenicity for specific epitopes is then predicted by considering interactions between epitopes and viral strains.

Together, Molecular docking analysis and MDS, are of particular interest in this Review. This is because Molecular docking and MDS quantify interaction between epitope and alleles i.e. Molecular docking is the step prior to MDS to prepare a 'ligand-receptor' complex [35]. Molecular docking is a technique used to predict how the ligand interacts with the receptor. This is done by calculating the binding energy between the ligand and receptor [36]. Epitopes, or antigens, are identified as 'ligands', whilst different alleles and immune receptors are treated as 'receptors', Figure A4. The docking analysis scores these complexes based on binding affinities. Binding affinities are computed from Equation A1 [37], which gives numerically the energy difference between bound and unbound ligands and proteins. The complex with the most negative value for binding affinity is transferred to MDS to validate the binding interaction [28], and to determine stability [27].

$$\Delta G_{bind} = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{R-R} - V_{unbound}^{R-R}) + \left(V_{bound}^{R-L} - V_{unbound}^{R-L} + \Delta S_{conf}\right)$$

Equation A1

Where L is ligand; P is receptor; conf is conformational entropy.

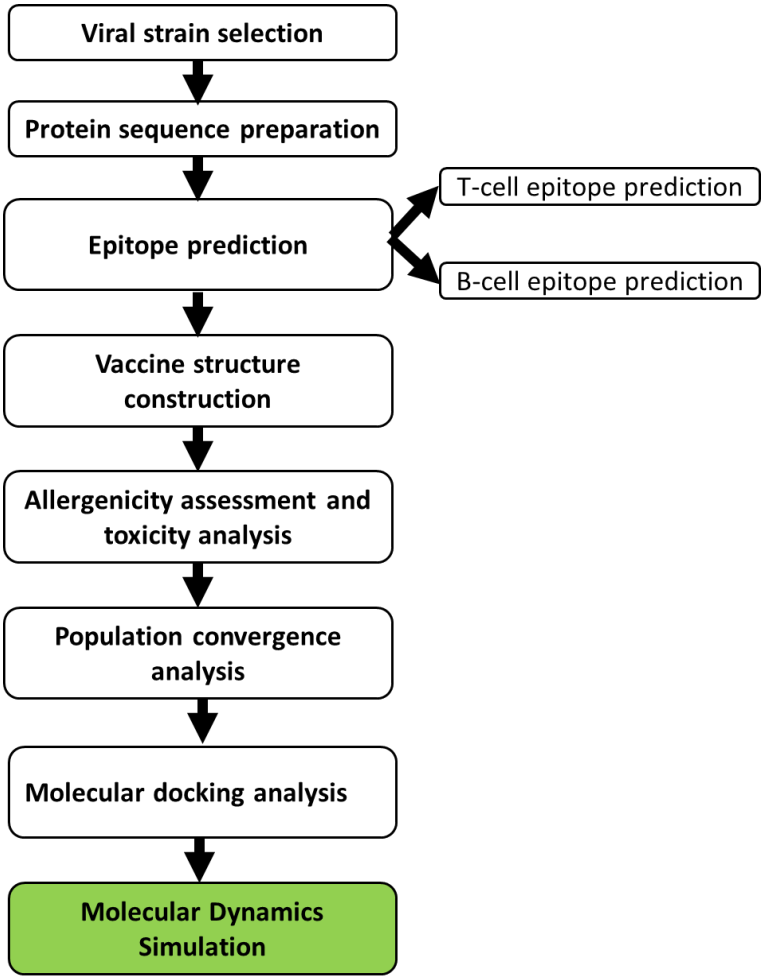


Figure A3 Flowchart for prediction of epitope immunogenicity.

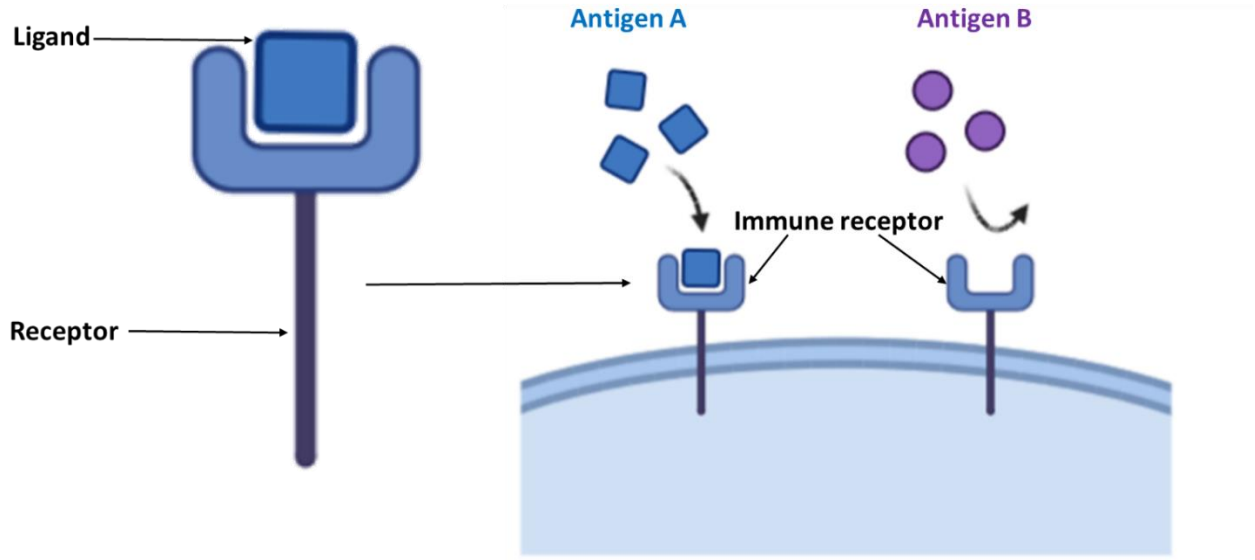


Figure A4 Schematic of antigen-immune receptors interaction in BioRender©. The antigen is considered as 'ligand' and immune receptor as 'receptor'.

A2.1 Application of MDS in epitope immunogenicity prediction

Hasan *et al*. conducted research to develop a vaccine against Marburg virus [27]. PEP-FOLD3 is an online server to predict peptide structures based on amino acid sequences via a *de novo* approach [38]. By defining the specific patch of a protein, all possible peptide-protein complexes are predicted.[39] Through this tool, they obtained 12 T-cell epitopes, 6 from envelope glycoprotein and 6 from matrix protein VP40. All selected epitopes were confirmed by performing docking analysis with Human leukocyte antigen (HLA) molecules, named, respectively, HLA-A*11:01 and HLA-DRB1*04:01.

It was found that the epitope with sequence VQEDDLAAGLSWIPF from the envelope glycoprotein was bound by HLA-DRB1*04:01 (-7.8 kcal mol^{-1}), and the epitope with sequence VPAWLPLGIMSNFEY from matrix protein VP40 demonstrated binding energy of -7.0 kcal

mol^{-1}) [27]. VP1-epitope 'APIDFDPVP' was found to interact with HLA-A*11:01 most strongly in terms of the least free binding energy (-9.5 kcal/mol) [27].

These authors focused on MDS to investigate stability and mobility of proteins with selected epitopes via normal mode analysis (NMA) which includes calculation of deformability, eigenvalues, B-factors and covariance [27]. Deformability determines the degree to which applying a force can make a given molecule change its shape. Eigenvalue is related to the motion stiffness, which is proportional to energy input to deform the structure. The less the eigenvalue, the more readily likely the deformation [40].

The eigenvalue for the complex V1-TLR3 with highly immunogenic epitopes, such as VP1, was $1.02e^{-04}$. However, these authors did not state whether this value was sufficiently low so as to result in deformation of the structure. They rather claimed that deformation of the structure was unlikely to occur, based on location of hinges in the chain (Figure SA1, Supporting Information). They compared the essential dynamics of V1-TLR3 from MDS with its normal mode of protein and found that the deformability value was < 1 for all residues. This is considered not significant.

This work by Hasan *et al*. focused on using molecular docking analysis to determine immunogenicity of listed epitopes based on binding affinity. Importantly, however, there was a lack of MDS to validate complex stability.

Docking analysis is actually weak in analyzing interactions between epitopes and immune receptors. MDS will be more accurate in predicting binding interactions if a dynamic system

is considered at the atomic level [41-46]. MDS docking analysis followed by MDS determine both binding interaction and complex stability [28,29,31,47].

Gupta *et al*. applied docking analysis to develop vaccines against carcinoembryonic antigen-related cell adhesion (CEACAM). These is a highly conserved region in cancers [28]. Seven (7) predicted epitopes were docked to HLA allele. The complexes formed were confirmed using ligand interaction study in MDS.

These authors compared the stability of the epitope-HLA complexes to that of original HLA complex structures. The free binding energy was computed to justify immunogenicity of varied epitopes. Findings showed the designed vaccine was highly likely to be efficacious. This was because: 1) epitopes bound strongly to the major histocompatibility complex (MHC) molecules were screened, and; 2) stability of complex with selected epitope over time has been validated.

A2.2 Analysis

Lessons from the case studies above underscore that the common analysis to predict epitope immunogenicity include: 1) molecular mechanical (MM) Poisson-Boltzmann surface area (PBSA) model [48]; 2) root-mean-square deviation (RMSD); 3) root-mean-square fluctuation (RMSF) of each residues, and; 4) radius of gyration (R_g) [28,29,31,47,49-52]. Definitions are given in Table A1.

Predictions of epitope immunogenicity were demonstrated through computation of these parameters and were found to be consistent with experimental immune results. For example, Deng *et al*. developed a multivalent enterovirus subunit vaccine for the prevention of HFMD (hand, foot and mouth disease) [52]. By showing strong binding energy via the PBSA tool and

high stability of formed complex via RMSD and RMSF, the designed multivalent enterovirus vaccine was then able to be expressed and purified. Test results *in vivo* demonstrated significant humoral and cell-mediated immune responses.

Recently, additional parameters have been identified through MDS to predict epitope immunogenicity. For example, 2 constructs designed as potential vaccine for multiple sclerosis disease, namely, Construct 1 and Construct 2, can be considered, Figure A5. By showing that the surface area of myelin oligodendrocyte glycoprotein (MOG) antigen in Construct 1 was similar to antigenic domain in Construct 2, and surface area of myelin basic protein (MBP) in Construct 1 was greater than for Construct 2. Construct 1 was concluded to carry antigenic property for vaccination [53]. This is because the construct with greater antigenic domain area is more likely to induce stronger immune response in the body.

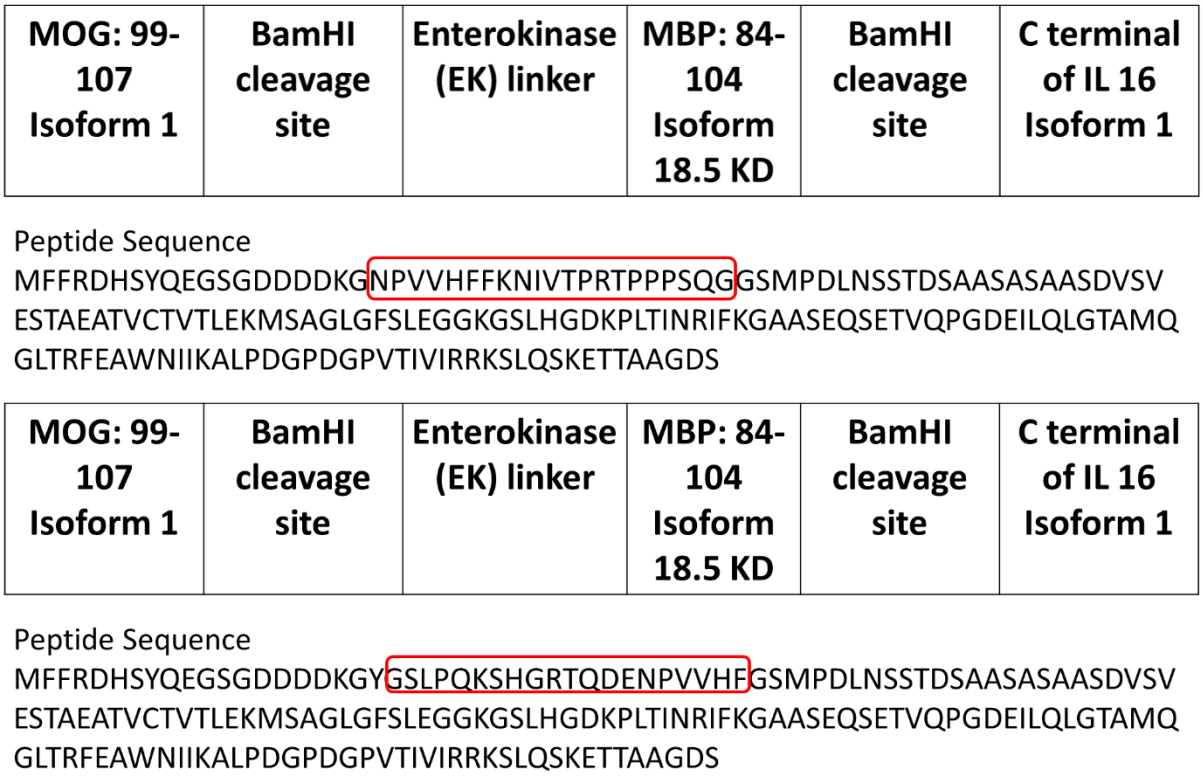


Figure A5 Peptide sequence for Construct 1 and Construct 2 reproduced from Banisharif-Dehkordi et al.[53] Copyright 2019, with permission from PMC, 10.4103/1735-5362.251849.

Additionally, a case study on COVID-19 by Kumar *et al*. employed MDS to assess the potency and specificity of the vaccine with target receptors [54-56].

Rather than computing RMSD, RMSF or R_g, they used a Normal Node Algorithm (NNA) to show that there were few atomic fluctuations in the complex system (vaccine and virus specific membrane receptor TLR-2) [40]. This finding implied low system deformations [54].

An Eigen score was computed of 3×10^{-6}, Figure A6A, that confirmed the rigidity of the complex movements [54]. A low Eigen score suggests the structure is less stable and more easily deformed [57].

However, this is not sufficient to draw any conclusions. This is because Kumar *et al*. did not report comparison of Eigen scores with controls. Through confirming a low deformation index of 0.1 to 1.0 Å, Figure A6B, they concluded that the complex was rigid and stable.

It is concluded that epitope immunogenicity prediction using MDS is therefore critical to understand each analysis, and to be validated by experimental data.

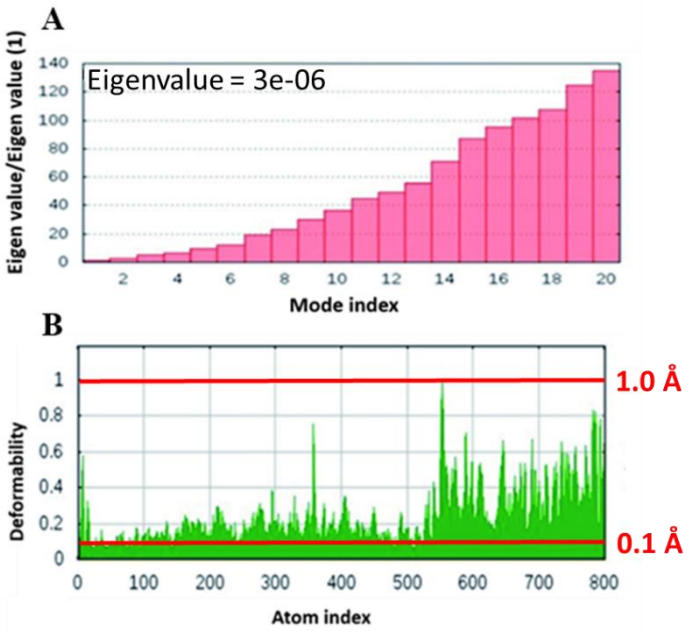


Figure A6 MDS analysis of vaccine construct with TLR-2 (A) Eigen value, (B) Deformability reproduced from Kumar *et al.*[40]. Copyright 2020, with permission from the Royal society of Chemistry, 10.1039/D0RA06849G.

A2.3 Computational limitations and experiments

It is not uncommon to observe weak connection, or even no connection, between simulation predictions and later experimental data. This is because of limitations in computing. These derive from 2 sources: 1) MDS uses shorter time-scales compared with actual biological processes [58]. This can be obviated at some level with development of parallel algorithms, software and hardware (supercomputer) [58-62], and; 2) selection of empirical force fields limiting approaches to actual biological systems. There are limited selections of force fields suitable for modeling actual biological system [58,63,64]. As is stated by Flower *et al.*, 'There is a tension between what we would wish for and what is available to us' [58]. This statement appears practically true when considering the gap between computational systems and actual biology; biological systems for the present are more complex than computational code. For example, existing methods to predict MHC binding typically rely on large-sets of experimentally-

measured binding affinities [65]. The quality of these methods is highly dependent on the available experimental data [66]. In addition, MHC binding data are limited to cover all of MHC alleles existed in the human population [65]. Approaches predicting MHC binding are trying to close this gap, but with mixed success [58].

We conclude however that these case studies highlight the significant practical potential of MDS for predicting epitope immunogenicity, compared with direct testing of immunogenicity of a vaccine via experiment. MDS can be reliably used to predict potency and specificity of selected epitopes through addressing complex binding energy and stability.

Despite some limitations, 'MDS predictions should be used before taking action'. This is because development time and costs for new vaccines are reduced, together with boosted success rates for efficacious vaccines.

A3 Prediction of protein structure

Following careful selection of epitopes with strong immunogenicity, prediction of the structure of epitope-based chimeric vaccine is the MDS following step. Step M2 (Figure A2) is considered an essential preparation before step M3, Protein Stability Study. Without an accurate predicted fusion protein structure, vaccine stability predicted by MDS will not match with experimental data.

A3.1 Prediction methods for protein structure

There are 2 protein structure prediction methods, namely, 1) template-based modeling (TBM), and; 2) template-free modeling (TFM), as is shown schematically in Figure A7.

TBM searches all protein structures determined from prior experiment from the Protein Data Bank (PDB), RCSB [67,68]. Comparative modeling (also called homology modeling) and threading are 2 usual TBM methods [68,69]. Homology modeling is performed by assuming that 2 homologous proteins share similar structures. The accuracy of the model is calculated by the degree of similarity of the sequences [70-72].

As is seen from Figure A7, the black-colored structure represents existing template with known protein information, and the red-colored structure represents the predicted part. The red-colored structure is formed by matching the sequence of black structure and proceeding template alignment. Threading is particularly used for fold recognition. By using certain parts of the target sequence giving high homology, the correct fold can be reliably selected, Figure A7 [73,74].

TFM is more computationally challenging compared with TBM. This is because it predicts protein tertiary structure based on primary amino acid sequence without sufficient structural information (Figure A7) [11,75,76]. *De novo*, or *ab initio*, is TFM modeling that builds 3D protein models based on first principles [75,77,78]. The more successful predictions for protein assembled structure usually contain fragments information from known protein structures [11], such as Rosetta[79,80] and QUARK [81]. Because it requires highly significant computational resources to maintain time-scales, time-length and significant number of interactions for construction of energy function, this predictor is found to work only for small proteins with length up to 120 amino acids [82,83], such as chignolin (10 residues, PDB ID: 1UAO) and Crambin (46 residues; PDB ID: 1CRN) [73]. The design of a fusion protein can be computationally demanding due to greater risks with mis-folding than with single-domain proteins.

Therefore, correct folding, stability and interaction between domains have to be investigated [84]. Two (2) cases are presented here to predict fusion protein structure.

Case-1 is applied when the structure of NPC is known. The construction of NPC inserted with epitopes focused on TEM homology modeling of similar protein structures, and introduction of linkers depends on design purposes. For NPCs that are not available in the PDB, Case-2 is applied to predict NPC, followed by prediction of NPC-epitopes structure. Case-2 therefore requires additional steps compared with Case-1.

In the following, we therefore focus on Case-2 type studies.

In the following, we therefore focus on Case-2 type studies.

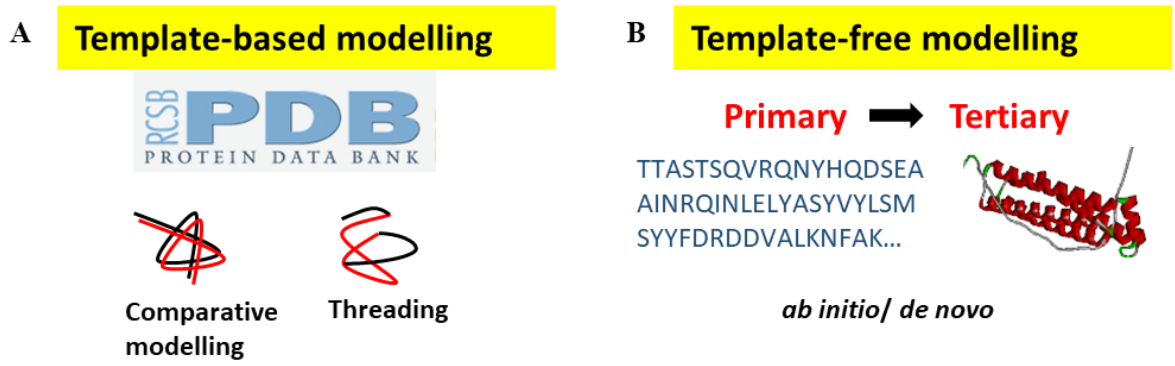


Figure A7 Two example methods for prediction of protein structure. (A) Template-based modelling (TBM). For TBM, for comparative modelling, black-colour represents existing structure in Protein Data Bank (PDB), whilst red is predicted structure. For threading, black represents recognized region, and red, predicted region. (B) Template-free modelling (TFM). For TFM, protein primary information, amino acid sequence, is the only available information. By performing ab initio, or de novo, protein tertiary structure (3D) is predicted.

A3.2 Case-2 type protein structure prediction

Shamriz *et al*. confirmed the advantages of MDS to predict initially unknown structures [84]. This study aimed to design and predict structure for a fusion protein, Plasmodium *falciparum* cell-traversal protein (P*f*CelTOS), inserted with human interleukin-2 (IL-2) as adjuvant and M cell-specific peptide ligand (Co1) as antigen. This fusion protein was developed as vaccine for prevention of malaria infection.

Via molecular simulation of P*f*CelTOS and designed fusion proteins using iterative-threading assembly refinement (I-TASSER), 3D structures were generated. These structures were confirmed in PyMOL by showing majority of residues located on the permitted region of Ramachandran Plot. MDS at an usual condition (300 K) were then carried out for analysis of energy (total, kinetic and potential) and structure stability (RMSD, R_g and H bond formation/deformation).

Designed fusion protein structures were critically compared in MDS. The desirable energy trend is for constant total energy, constant or decreasing kinetic energy, together with constant or increasing potential energy. An increasing kinetic energy level reflects breakdown of folded protein structures [84,85], whilst decreasing potential energy instability of designed protein structures [84]. The authors reported stable designed structures based on low values of RMSD and R_g. A constant number for H bonds confirmed the accuracy of designed fusion protein structures.

Sometimes, there exists limited structural information for NP. A combined study is then applied to predict protein structure. For example, Mobini *et al.* built a VLP-based vaccine using hepatitis B core antigen (HBcAg) as the carrier, with inserted Myrcludex and hepatitis B

surface antigen (HbsAg) [86]. Homologous structures for HBcAg and HBsAg were searched for prediction of protein tertiary structures. It was found that HBcAg was 94 % identical to the control (PDB ID: 1QGT) for domain of the protein in the N-terminal (residues 1-140). There was no available structural information for the nanopeptides (residues 141-149) and C-terminal domain (150-183) in existing database. A combined homology modeling and *ab initio* was applied to predict structure for HBcAg proteins.

A total of 10^4 models were scored based on the discrete optimized potential energy (DOPE) and the top 10 with the lowest DOPE score were readied for conformational analysis. For HbsA, *ab initio* was performed. The quality of the final model was evaluated by Z-score in comparison with experimentally validated structure of proteins.

The conformational behavior of the designed vaccine was confirmed via a low value RMSD and R_g.

A3.3 Current challenges and innovative MDS

In practice, using MDS to predict large protein structure can be a practical challenge. At present *de novo* predictors are difficult for large proteins [75,87,88].

A number of *de novo* approaches for assembling proteins are built on short peptide fragments. These are suitable for small proteins containing less than 1000 amino acids [89,90]. A certain time-length is required to perform umbrella-sampling and proceed to protein folding [91]. Current computer resources are limited for microsecond-based simulations [92]. However, protein folding can take microseconds to tens-of-minutes to complete.

Therefore some slow-folding protein simulations are likely to fail in protein correct folding [93]. However, innovations in MDS are expected to improve accuracy and performance on protein structure prediction.

Cheung *et al*. applied a method titled *NiDelta* with ultra-fast MDS to predict 18 large proteins, with more than 100 residues from sequence to tertiary structures [75]. This is shown schematically as Figure A8. It is seen in the figure that three (3) steps were involved in this method, Data Processing (Figure A8A), Prediction (Figure A8B), and Simulation (Figure A8C).

NiDelta has two (2) restraints, namely, 1) predicted torsion angle, and; 2) residue-contacts for performing a coarse-grained molecular dynamics (CGMD)-*Upside* [75]. There are 2 stages involved to process a given target sequence: Stage-1 is to build *Phsior* by constructing a 'non-redundant sequence' set of data from the PDB library and filter it through PISCES [94], and; Stage-2 is to predict residue-contacts via sequence alignments. Additional detail was provided [75].

An ultra-fast MDS simulation was accomplished with *Upside* in which each model was displayed by a reduced chain representation, but with sufficient structural information, such as, backbone, Cα and C atoms. For each of 18 large proteins, 500 *Upside* simulations were carried out beginning with an unfolded structure. Each was folded in several CPU h. Nine (9) representative residue contacts were compared with native ones by computing Cα-RMSD and TM-score based on experimental reference structures.

It was found that many protein structures were predicted sufficiently, despite 'noises' from incorrect predictions caused by false positive inferences (Figure A9 below is an example of

noises). For example, protein C-H-RAS P21 (PDB ID: 5P21) was sufficiently predicted with lowest Cα-RMSD of 3 Å and the same fold with TM-score of 0.76. These findings gave confidence for practical potential of this innovative approach of *Upside* to predict large protein structures.

This study highlighted however a number of important limitations. These included: 1) Residue-residue contacts were not able to be predicted without alignment of sufficient multiple sequences; 2) Despite sufficient sequences, there were failed predictions; false positive prediction in the pairwise contacts caused incorrect building of protein 3D structure, and; 3) It was difficult to accurately predict torsion angles. It is concluded that overall however this study can be reasonably regarded as a successful demonstration of innovative MDS to predict large protein structures.

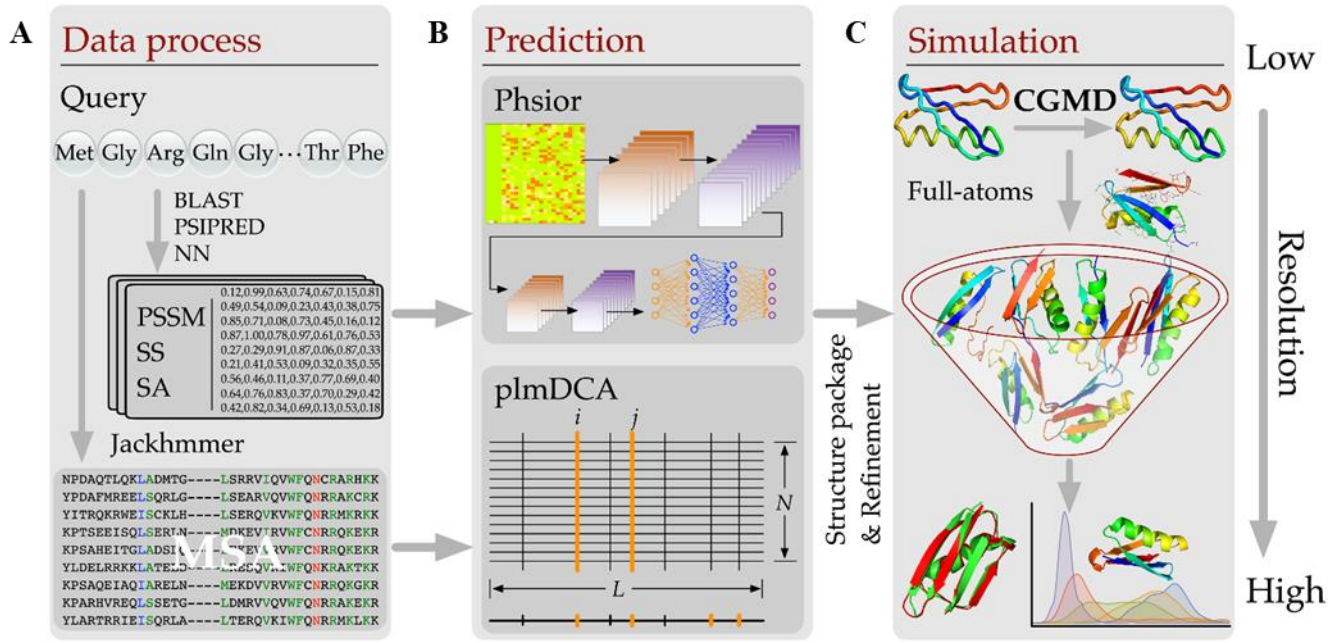


Figure A8 Schematic for innovative approach to predict protein tertiary structure following three (3) steps. (A) Step 1, Data Process (*sic*). MSA stands for multiple sequence alignment. (B) Step 2, Prediction using *Phsior* and plmDCA. (C) Step 3, Simulation. By performing CGMD, protein tertiary structure with highest resolution is refined and filtered. This figure is reprinted from Cheung *et al.*[75] Copyright 2018, with permission from PlOS ONE, 10.1371/journal.pone.0205819.g001.

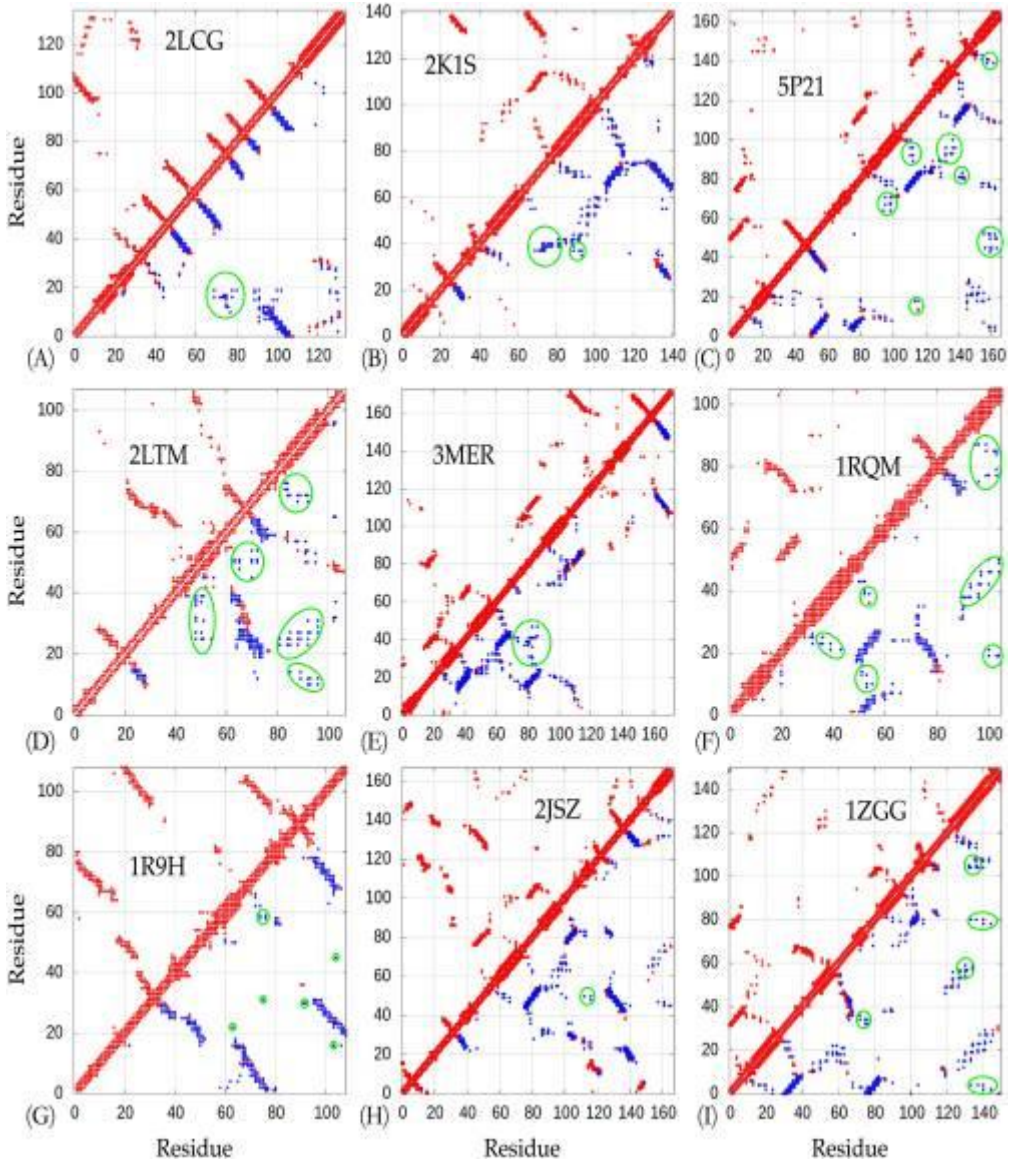


Figure A9 Examples of noises from false positive inferences. Noises are denoted as blue-color dots in green circle. This figure is reprinted from Cheung *et al.*[75] Copyright 2018, with permission from PlOS ONE, 10.1371/journal.pone.0205819.

Recent innovative MDS considers implications of suitable force fields on studying protein folding or mis-folding. There is a special group of NPs called repeat proteins. The simulation for repeat proteins folding is practically difficult to achieve.

Zheng *et al*. applied a coarse-grained protein force field, AWSEM (Associative Memory,

Water-mediated, Structure and Energy Model), into MDS to investigate the folding of fused dimers of SH3 domains and Ig domains from human titin [95]. They reported that the frequency of mis-folding was decreasing by reducing the sequence identity between monomers in the fused dimer. This is supported by the hypothesis that evolution favored low sequence identity between nearby regions in repeat proteins [96].

This finding is practically useful for future epitope-based chimeric vaccine design. For repeating proteins, by altering epitope sequence or length, each monomer can have unequal sequence identity. This is desirable for protein folding and readier downstream processing development.

These studies have shown that overall, MDS has been successfully applied to predict protein structure. A detailed picture of protein folding, which is practically difficult to ascertain from experiment, is gained through simulations. However, contrary to small protein structure prediction, there are additional difficulties involved for large protein application of MDS.

There is a number of recently released, advanced software that apply deep machine learning artificial intelligence (AI) to predict protein structure, with reported high accuracy. DeepMind™ released an open-source version of a deep-learning neural network, AlphaFold2 [97]. This was developed for prediction of accurate protein structure when there are no homologous structures available. The key input is protein primary amino acid sequence, and aligned sequences of homologues (multiple sequence alignments and pairwise features). These are processed by the trunk of the network called Evoformer. The detailed procedure is reported by Jumper *et al*.[97]. AlphaFold has demonstrated a near experimental accuracy in predicting protein 3D structures. Accuracy was comparable with other computing methods with a median

backbone accuracy confirmed 0.96 Å RMSD for Cα (compared with another 'best method' of RMSD for Cα of 2.8 Å).

Another method called RoseTTA fold was developed by Baek *et al.*, and can obtain accurate models of protein-protein complexes [98]. The authors report application to a three-track network to process sequence, distance and coordinate information concurrently to rapidly generate protein structures with an similar to that from DeepMind. This software advantageously speeds up generation of accurate models for protein structures in MDS to predict performance of Ebc vaccine, together with boosted user confidence.

A4 Protein stability

Protein stability is mandatory in epitope-based chimeric vaccine development. Stable proteins ensure the functionality and integrity of vaccines, and avert negative reactions due to aggregates. Unstable proteins result in protein degradation with loss of function, or aggregation. This reduces biological activity [99]. MDS can be used to predict protein dynamics, structures and functions at an atomistic level [17,100,101].

A4.1 Parameters for computation

Analyses commonly computed for determination of protein stability are summarized in Table A1. It is required to make connections between these and experimental data.

RMSD is widely used to reveal protein structural stability via determining conformational stability changes in proteins [102,103]. RMSD varies significantly in the initial timeframes in simulation, and reaches steady-state when the simulation system is stable. Comparing RMSD values at steady-state is important so as to present an accurate estimation of protein

conformation stability. The greater the RMSD, the less stable the protein structure is in MDS [17,18,99,104-106].

R_g is regarded as the fundamental measurement of protein overall size. This serves to show protein structure change in MDS [102,107]. R_g depends on both original protein size, and protein stability change in time. A low R_g value indicates a compacted protein structure that results in greater stability [18,105,108,109].

RMSF is a basic analysis commonly computed to specify the flexibility of individual residues. It is associated with RMSD and R_g. The greater RMSF value, the more flexible movement [106,110,111]. RMSF is related to protein stability. Low average RMSF values indicate individual amino acid residues that help preserve protein stability during MDS[112]. High RMSF values confirm flexible regions on proteins. Therefore, RMSF is a practically useful analysis for mutant designs with stabilized flexible regions [113,114].

Table A1. Common parameters determined in Molecular Dynamics Simulation (MDS) to study protein stability.

Parameter	Role
Root-mean-square (RMSD)deviation	Characterization of conformational change in proteins.
Radius of gyration (R_g)	Basic measurement of overall size of a chain molecule and indication of conformational change in protein rigidity.
Root-mean-square fluctuation (RMSF)	Determination of flexibility differences amongst residues.
Native contacts (Q)	Indication of native contacts preserved under conditions such as hyperthermal or with existence of chemical denaturants (urea).
Hydrogen bonds (HBs)	Maintenance of overall stability of the protein structure.
Solvent-accessible-surface-area (SASA)	Examination of unfolding of proteins exposing buried hydrophobic residues to water.
Secondary structure analysis	Description of conformational change by determination of secondary structure elements change, such as, helical content. Most common element is α-helices.
Ramachandran Plot	A phase diagram of two (2) torsion angles φ and ψ showing if residues are in permitted regions of residue conformations.

The general use of these analyses appears to work well for natural conditions. However, to ensure a detailed study of protein stability at molecular level, it is necessary to run MDS under conditions including, elevated temperature, range of pH, and/or chemical denaturants. The reason for doing this is to understand the mechanism by which proteins unfold from native structure.

The fraction of exhibited native contacts, Q, indicates the stability of proteins under these more-extreme conditions, such as hyperthermal, or in the presence of chemical denaturants [17]. Less stable proteins normally unfold rapidly, resulting in a reduction of native contacts compared with stable proteins over the same simulation time [17,115-117]. Hydrophobic interactions between each residue of the protein and water are quantified by native contacts and hydrogen bond (HB). HB contribute favorably to protein stability [118,119], especially, maintaining secondary structure (α-helices) [108]. The fewer number of HBs in proteins leads to markedly lower stability [17,106].

Another analysis closely related to hydrophobic interactions, is solvent accessible surface area (SASA). SASA of proteins is considered important to observe protein folding and stability, which is stated as 'protein surface surrounded by a hypothetical center of a solvent sphere with the van der Waals contact surface of the molecule' [120]. This value is dependent on the interactions of hydrophilic and hydrophobic residues with solvent, and also, the surface tension around protein-solvent surface. During unfolding, the SASA of the protein will increase because hydration of the hydrophobic core results in disturbance of hydrophobic interactions amongst non-polar residues [17]. It was reported also by Berhanu *et al.* that residues that were located inside protein cage from the solvent had smaller SASA than residues exposed to the solvent [121]. The display of polar residues is energy-favorable, however, exposure of hydrophobic residues increases the free energy of the system, which is unfavorable. This can be used to determine the stability of protein conformations [121].

Zhang *et al*. reported that SASA for apoMb, and its variants E109G and E109A, increased in the presence of urea as unfolding proceeded [17]. E109G exhibited a steeper increase in SASA than wild-type apoMb, whilst E109A showed a reduced increase. Based on this finding, it was

concluded that E109A improved the stability of wild-type apoMb and E109G had a destabilizing impact.

The connection between SASA and protein stability however may not be significantly obvious without more-extreme conditions. For example, Kumar *et al*. deduced a contrary conclusion [106]. They assessed SASA for wild-type and mutant PMP22 structures. They illustrated a greater SASA for the mutant protein structure than for the wild-type protein, making the mutant structure more stable [106,122]. However based on SASAs, the difference between wild-type and mutant was not significant. It is difficult therefore to draw an estimate of protein stability from these SASA data.

Muneeswaran *et al*. computed both RMSD and SASA for wild-type cytochrome (Cyt) c and its mutants (Y67E, K72W) [123]. Low RMSD values for both backbone and Cα showed wild-type Cyt c was the more stable compared to its mutants. However, the value for SASA was greatest amongst these three. It is not possible to confidently link protein stability with computed SASA from native MDS. This however can be more confidently made for more-extreme conditions.

MDS analysis of secondary structure is gaining momentum in studying protein stability. To address why it necessary to analyze protein secondary structures, it is important to understand the limitation(s) of MDS.

For large protein molecules there are computational time limitations. Experiments, including heating of proteins until denaturation, requires a significant time (min). This is not ideal given the maximum simulation time is μs. Therefore, changes of protein tertiary structure caused by

denaturants without sufficient treatment time are not obvious in MDS, whilst protein secondary structure changes are more visible in given MDS-time length. By analyzing key secondary structure elements (e.g. α helix, β sheet, coil, bend and turn) [18,24,76,124,125], protein stability can be estimated.

For example, Amir *et al*. analyzed the changes in secondary structure in wild-type STN1 protein and its mutants R13T and D157. It was found that D157Y mutant resulted in a significant decrease in secondary structure in comparison with R135T, where there was no significant change [24]. This finding evidenced that D157 mutation impacted conformational stability of STN1. Another study by Schaller *et al*. focused on change in α-helical content in analyzing thermal stability of biosurfactant DAMP4 [22]. By comparing initial helix content (%) and end-point helix content (%) of all variants, loss of secondary structure was identified and stability was assessed. A significant loss of helix content underscored an instability of protein.

The Ramachandran plot is closely related to protein secondary structure analyses. Ramachandran plot is a phase diagram, made by 2 torsion angles φ and ψ for each amino acid residue. It is used to identify whether the conformation of amino acids are in permitted regions [108,126-128]. An example of a Ramachandran plot is presented as Figure A10. Most residues marked as black-color in the figure, were concentrated in the preferred regions (marked in red-color) [108]. This finding evidenced a stable secondary structure conformation.

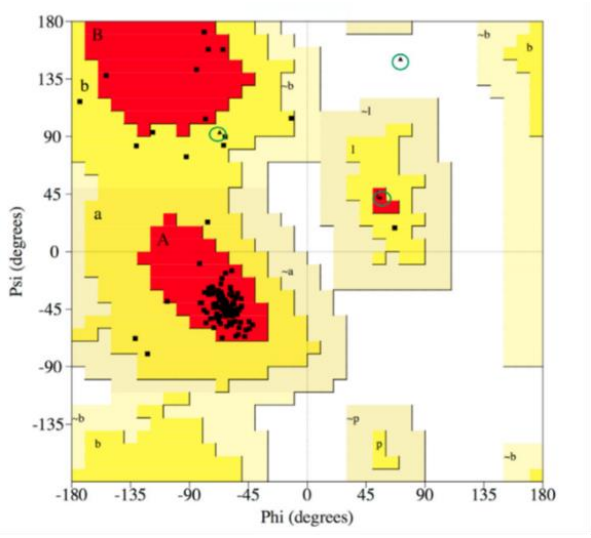


Figure A10 Example Ramachandran plot for Rop protein. Red-colour represents most preferred regions; transition of yellow from dark to light represents most to least favourable conformations. This figure is reprinted from Arnittali *et al*.[108] Copyright 2019, with permission from Elsevier, 10.1016/j.procs.2019.08.181.

A4.2 MDS steps for protein stability

Present application of MDS for determining protein stability involves the three (3) steps: 1) simulation for native/non-native environments, 2) designs via analyzing dynamics of unstable residues, and; 3) simulation of designs to determine impact of mutations to improve protein stability. Figure A11 shows schematically the connections between these steps. The focus in the following is on simulation under more-extreme conditions, including elevated temperature, varying pH and the presence of chemical denaturants.

Modeling of protein structure at elevated temperature is used to determine thermal stability. For example, Manjunath *et al*. simulated 5 functional SAICAR synthetase ligates from different organisms including, mesophilic, thermophilic and hyperthermophilic, at 3 temperature values of 300, 333 and 363 K [105]. To establish stability differences, they

determined RMSD, RMSF, R_g, SASA, HB and salt bridges (SBs). SBs are defined as ion-ion interactions between charged residues, which are important to maintain conformational stability of proteins [129,130].

It was found that mesophilic structures appeared unstable at 363 K. They showed significantly increased RMSD, fluctuating RMSF and R_g at 363 K compared with other groups. There was a gradual collapse of the structure and instability at 363 K for the mesophilic group. Additionally, the hydrophobic SASA more rapidly increased relative to the mesophiles, compared with hyper-thermophiles. In mesophiles, a comparatively higher number of long-lived (with higher percentage existence time) contacts are lost due to increased protein flexibility.

A comparatively greater number of hydrogen bonds are found in the mesophilic proteins at 363 K, compared with that at 300 K, whilst other protein groups remained unchanged. The number of SBs in the mesophilic proteins decreased in comparison with the other groups at 363 K.

These findings led these authors to conclude that mesophilic proteins were not stable at 363 K when compared with other groups.

It is important to note that this research was based on only simulations. Findings from each analysis were supporting each other, however. This confirmed the accuracy of the methodology dependent on MDS. This study therefore highlighted the advantages of MDS in investigating protein stability. Experiments are not necessary therefore to be conducted when sufficient self-reinforcing information has been generated from MDS.

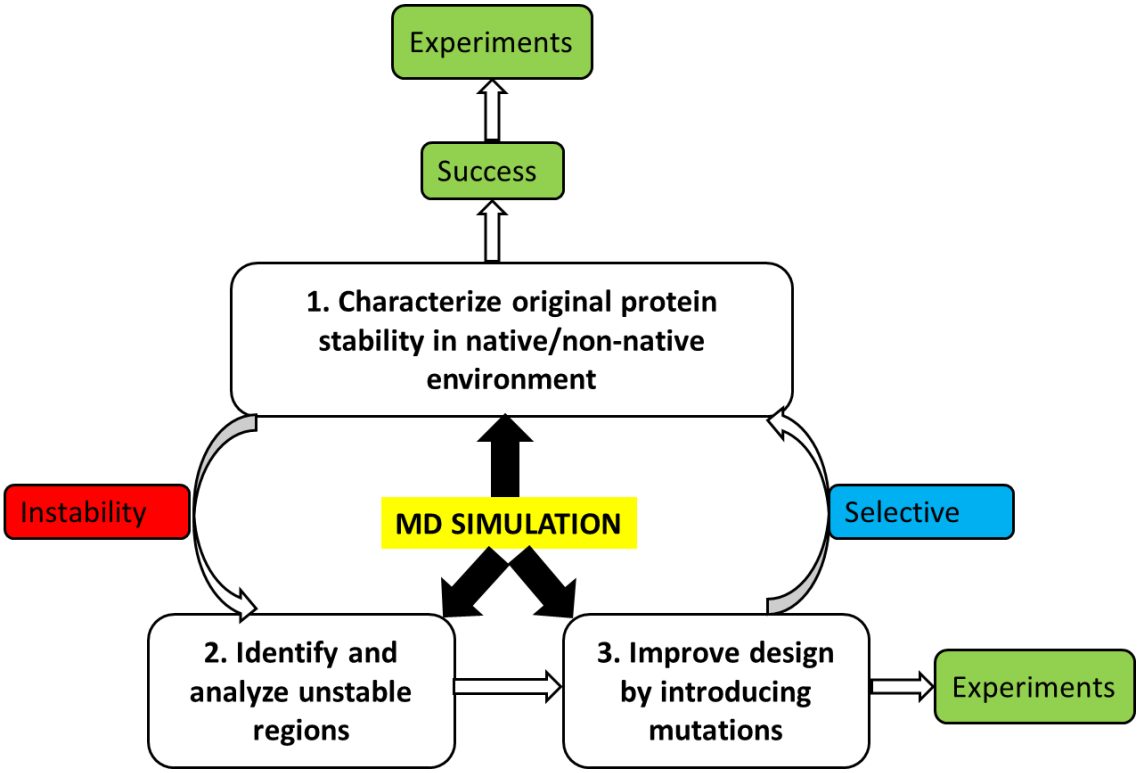


Figure A11 MSD flowchart for protein stability.

There is also potential to study stability using both simulation and experiment. This combined technique sometimes is advantageous because it involves validation of simulation.

A case study reported by Bekker *et al*. for example predicted relative thermal stability of single-domain antibodies (sdAbs) using high temperature MDS and validated the results with melting temperature (T_m) obtained from experiments [23]. Melting temperature measures the temperature where half of a protein is folded and the other half is un-folded.

High-temperature MDS at temperature representing 400 and 500 K were performed using 300 K as a control. The methodology was that high temperature MDS which gave valuable information on protein design. However, they reported that RMSD did not correspond well with T_m. There was no significant difference in RMSD between antibodies at the different

temperatures. The fraction of native atomic contacts (Q) showed closer estimation to experiment. This Q value determines the fluctuation of the protein corresponding to the atomistic interactions.

It was reported that MDS at 400 K combined with Q gave an accurate measure of thermal stability of sdAbs. This was because there was a good correlation between Q value at 400 K and experimental T_m.

Importantly, this study highlighted a challenge with MDS. This is that the connection between simulation and experimental data needs to be carefully interpreted.

Buffer pH is important to protein stability. Thermal stability is closely related to varied pH of solution. Residue charges are determined based on the solution pH. These are involved in both intra- and inter-molecular noncovalent interaction. This results in increasing, or decreasing protein stability.

For example, Sviben *et al*. reported that thermal stability of ovalbumin was dependent on pH. This was confirmed both in MDS and with experiment [99]. Ten (10) simulations were conducted for 2 pH conditions (acidic and alkaline) at 5 temperatures of 37, 42, 57, 74 and 85 °C. It was found that for both acidic and alkaline environment, at temperatures of 37 and 42 °C more stable systems are achieved than at higher temperatures of 74 and 85 °C.

Additionally, instability in the acidic system was more pronounced than that for the alkaline. These findings were consistent with findings from experiment. The key difference between these 2 systems was the interaction of arginine 50 and arginine 58, with surrounding residues.

The alkaline system was found to have stronger and more persistent interactions for both Arg50 and Arg58 compared with the acidic system [99].

In addition to simulation for different temperature and solution pH, the solvent environment with chemical denaturants can be used to provide more information on differences in protein stability.

Zhang *et al*. applied MDS to investigate the stability of 4 apomyoglobin (apoMb) variants, namely, wild-type, E109A, E109G and G65A/G73A, in explicit urea solution [17]. The 2 M urea solution at pH 4.2 acted as a denaturant to drive unfolding of the apoMb variants. Through analyses of RMSD, native contacts and SASA, E109A showed greatest stability, and E109G showed least stability in urea solution.

It was found that the greatest change in conformation was related to destabilization of E109G. In the presence of denaturants, less stable proteins unfold more rapidly than others. This results in significant loss of native contacts against simulation time. Based on RMSD, E109G exhibited greatest drop in native contacts, whilst E109A preserved greatest percentage of native contacts. This led to the most stable structure [17].

These findings were highly consistent with experimental results [131]. This study highlighted therefore that MDS can be used to achieve experiment-matched results for proteins in solvent with denaturant present. This application therefore saved significant time and costs involved with experiment.

With desirable stability results from application of MDS, proteins are ready to be tested in experiment to validate stability and, to check vaccine immunogenicity. However, this model

situation does not always occur. The advantage of MDS compared with experiments is that it computes protein stability, and provides molecular level information to explain any resulting stability difference.

The second-related step in using MDS is to find unstable regions of proteins, Figure A10. However, the second- and the third-related steps using MDS for protein stability are normally bound together. This is shown in Figure A10. Following identification and analyses of flexible regions using MDS, further mutations are investigated with the aim to improve protein stability. Recent studies focus on improving protein thermal stability.

Thermal stability is an important analysis to more fully understand protein structure and to improve protein thermal stability for industrial application. There are a number of methods to improve protein thermal stability, including: introducing disulfide bonds [104,105,132,133]; increasing the number of salt bridges [104,134-137]; optimizing electrostatic interactions [104,133,138,139]; compacting hydrophobic core [104,105,140-142]; enhancing hydrogen bonding [143-147]; and stabilizing α-helices [148]. However it is not necessary to determine all related properties in a given, stable protein.

Combination of one or more of these can significantly impact protein thermal stability [105,143]. Relying on experiment to explore possible combinations is practically time consuming to complete, with at most, likely only mixed success. MDS can therefore be used to speed development via judicious simulation.

Pikkemaat *et al*. used haloalkane dehalogenase (DhlA) as a model to investigate potential of MDS to identify flexible regions in proteins to boost thermal stability [18]. These authors focused

on RMSD of the molecule relative to the X-ray structure. They also determined the effect of thermal denaturants on secondary structure of the molecule. They reported that most of the mean square displacement was concentrated in the region between residues 180 and 210. The secondary structure elements analysis highlighted that this region was highly stable, and indicated less possibilities of unfolding of the structure. They found relatively large motions occurred in the cap domain of DhlA. It was concluded from the MDS, that DhlA had high mobility in helix-loop-helix region involving residues 185-211. The authors reported introducing a disulfide cross-link between residue 201 of this flexible region and residue 16 of the main domain. The mutated enzyme exhibited increasing transition temperature from 47.5 to 52.5 °C. Urea denaturation exhibited a similar trend to experiment. It is concluded that this work demonstrates the advantageous use of MDS to identify mobile protein domains to boost stability.

Gill *et al*. conducted all-atom explicit-solvent MDS for human procarboxypeptidase A2 (AYEwt) with 3 designed variants, namely, AYEdes, AYEwt-4mut and AYEwt-5mut, at both 25 and 100 °C [78]. MDS findings highlighted potential stabilizing factors at atomic-level in a set of engineered proteins.

Through analyses of RMSD, RMSF and SASA, protein stability was determined and compared. It was predicted that Phe18 and Ala23 from AYEdes would stabilize α1 of AYEwt-3mut. The authors therefore further designed AYEwt-F58M and AYEdes-M58F with the aim to validate this prediction. They observed that α1 of AYEwt-3mut was stabilized and that phenylalanine had a minor but meaningful stabilizing impact over methionine at position 58. Importantly, there are a number of examples using MDS to show that mutated structures decrease overall thermal stability.

Meharenna *et al*. investigated the regions of Cytochromes P450 related to thermal stability via high temperature MDS [149]. CYP119 was the most stable P450 compared with P450cam and P450cin [149]. By comparing trajectories at 500 K, the Cys ligand loop was hypothesized as an important region for improving thermal stability. This region in CYP119 was stabilized by rigid, nonpolar interactions between Tyr26 and Leu308. The authors generated Y26A/L308A CYP119 by replacing these 2 amino acids in P450cam with, respectively, Gly and Thr. However, the resulting stability was less than that for the wild-type CYP119. The Cys ligand loop unfolded similarly to P450cam. The melting temperature from experiment exhibited a similar trend for mutated P450cam at 16 °C less than for wild-type CYP119. MDS therefore predicted the Cys ligand loop as a 'hot spot' critically impacting stability.

It is possible to use MDS to predict which mutants boost stability, however, it is practically challenging to mimic exact interactions amongst alternated residues. To model similar interaction between Tyr26 and Leu397 in CYP119, larger nonpolar residues are hypothesized as being replaced in P450cam.

Meharenna *et al*. introduced a disulfide bond, however, this mutant failed to improve stability. One reason is that disulfide bonds are unlikely to mimic non-bonded intramolecular interactions, and unfavorable packing problems likely occur. In a similar investigation. Bae *et al*. reported that introducing a salt bridge improved stability of mesophilic protein [150]. It is possible therefore to mimic non-bonded intramolecular interactions via introducing salt bridges.

MDS has been used to investigate protein stability and to predict likely vaccine efficacy. It is advantageous to: 1) characterize fusion protein stability as impacted by thermal-, pH- and

chemical denaturant factors, 2) give practical feedback to identify flexible regions, and; 3) investigate the likely impact of mutations designed to improve stability. It is concluded that MDS simulations significantly improve vaccine success rates by filtering and/or building highly stable epitope-based chimeric vaccines.

A5 Futher applications and future

There are multiple factors affecting performance of epitope-based chimeric vaccines. The more significant are 1) location of insertion site and 2) linker design. There are however limited studies using MDS to take these 2 factors into account. Most vaccine studies are experiment-based. In the following we suggest future applications using MDS.

The location of the insertion site significantly impacts stability and immunogenicity of the vaccine. For example, human ferritin heavy chain is a protein cage. There are 3 potential insertion sites located on ferritin, 1) N-terminus, 2) C-terminus and 3) the flexible loop region between helix D and E [151]. Our research has shown that for same type of epitopes inserted, N-terminus (E1F1) exhibited a significantly different behavior compared with C-terminus insertions (F1E1) [152]. F1E1 was less hydrophobic and more thermally stable than E1F1 based on the simulation results. These were consistent with experimental data. An immunological study highlighted that C-terminus insertion induced stronger cell-immune response than did N-terminus insertion (to be published). This is mostly likely caused by greater structural stability. Similar results were reported by Han *et al*. The epitopes inserted at C-terminus were processed more efficiently by dendritic cells (DCs) compared with the epitopes inserted at N-terminus [153].

For NPCs that have multiple insertion sites, MDS can be applied as steps M2 and M3 (Figure A2). Protein structure prediction (M2) gives the structures of fusion proteins with inserted epitopes at different sites, whilst protein stability investigation (M3) is used to estimate which insertion site give greatest stability. This is advantageous in the design of epitope-based chimeric vaccine at an early stage without the need to conduct experiments.

Necessary decisions regarding linker type and length can be based on MDS. The method is to construct an accurate fusion protein by careful selection of a protein linker. The selected linkers relate to biological activity [154-156]. There are 3 types of linkers, 1) soft, 2) rigid and 3) cleavable [157]. Rigid and cleavable linkers involve specific functionality based on design aim [155], and are more rigid than soft linkers. Soft linkers are more commonly applied in constructing fusion proteins with only minor impact on final structure. All act as spacers to support protein folding and stability to display epitopes for binding of MHC class I and II molecules [158,159].

Selection of linker type and length are important. Wei *et al*. investigated the impact of linker length and rigidity for fusion proteins, Lumazine synthase (LS) inserted with linear B cell epitopes (PB10) [160]. They used 4 linkers of different length and rigidity, and examined corresponding protein stability and immunogenicity via experiment. This study was limited however to investigation of resulting protein assembly and thermal stability. Different length and rigidity had negligible impact on protein assembly, thermal stability and display of epitopes, and minor impact on protein conformation. Experiments are difficult to determine this sort of information in a short period [161-163]. Previous case studies have demonstrated that MDS is able to estimate protein folding and stability with a near accuracy to experiments but in a faster way.

In practice, selection of linkers with suitable length and flexibility is not straightforward [164]. This can be readily solved using MDS, however. Robinson-Mosher *et al*. for example summarized findings on soft linkers of various lengths by computing distance between proteins, force and binding effect [164]. It was found through simulation that the average distance between the protein elements was significantly shorter than the maximal length of the linker.

It was concluded that fusion protein structure is impacted more by the dynamics of the linker and not the Brownian motion of the protein elements [164]. A maximal binding rate (with cell-surface receptor) was determined via an intermediate linker length and greater stiffness [164].

Additionally, Rezaie *et al*. confirmed that MDS can be reliably used to design linkers [161]. They used computation to design a single linker combining advantages of both rigid linker (EAAAK) and probe (tetracysteine sequences). MDS results (RMSD, RMSF, R_g and contact maps) highlighted the design of a single linker was appropriate for structural stability.

It is concluded that because MDS can reveal the motions of individual atoms it can be reliably used to investigate the impact of linkers without the need for prior experiment. Experimental confirmation is nevertheless necessary for subsequent practical confirmation and progression toward clinic.

A6 Conclusions

Vaccination is widely used to prevent spread of disease. Epitope-based chimeric vaccine has attracted significant attention globally because of efficacy and safety. The selection of efficacious epitopes is important. Traditional experimental developments however are limited because they are time-consuming and expensive. Optimization of vaccine production is predicated on good structural design.

Molecular Dynamics Simulation (MDS) is an emerging alternative for epitope-based chimeric vaccines, and has significant advantages. These include prediction of: 1) performance of epitope-based chimeric vaccines 2) fusion protein structure via considering protein folding and resulting structure stability, and; 3) protein stability. MDS, judiciously used, can filter and establish the most stable vaccine for later validation via experiment with reduced time and cost.

Importantly MDS can be used in both native- or extrema-conditions, and to take account of the impact of temperature, pH and chemical denaturant. Additionally, it can be used to highlight unstable regions, so that targeted mutations can be introduced to boost stability.

Present limitations of MDS prediction however include: 1) restricted time-scale, and; 2) lack of often necessary biological data. Future research could productively explore these limitations to, 1) develop ultra-rapid simulations, and; 2) modify force fields.

We conclude that MDS can significantly boost confidence in vaccine manufacture. MDS can save expensive costs from repeating experiments. Findings will be of immediate interest to a wide range of researchers in simulation methods to guide practical vaccine development.

A7 References

1 López-Sagaseta, J. *et al*. Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **14**, 58-68, doi:https://doi.org/10.1016/j.csbj.2015.11.001 (2016).

2 Lee, N.-H. *et al.* A review of vaccine development and research for industry animals in Korea. *Clinical and experimental vaccine research* **1**, 18-34, doi:10.7774/cevr.2012.1.1.18 (2012).

3 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine: Nanotechnology, Biology and Medicine* **2**, 137-149, doi:https://doi.org/10.1016/j.nano.2006.07.009 (2006).

4 Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnology Advances* **41**, 107547, doi:https://doi.org/10.1016/j.biotechadv.2020.107547 (2020).

5 Bachmann, M. F. & Jennings, G. T. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **10**, 787-796, doi:10.1038/nri2868 (2010).

6 Mohsen, M. O. *et al.* Major findings and recent advances in virus–like particle (VLP)-based vaccines. *Seminars in Immunology* **34**, 123-132, doi:https://doi.org/10.1016/j.smim.2017.08.014 (2017).

7 Pati, R. *et al.* Nanoparticle vaccines against infectious diseases. *Frontiers in immunology* **9**, 2224-2224, doi:10.3389/fimmu.2018.02224 (2018).

8 Lung, P. *et al.* Nanoparticle formulated vaccines: opportunities and challenges. *Nanoscale* **12**, 5746-5763, doi:10.1039/C9NR08958F (2020).

9 Gregory, A. E. *et al.* Vaccine delivery using nanoparticles. *Frontiers in Cellular Infection Microbiology* **3**, 13-13, doi:10.3389/fcimb.2013.00013 (2013).

10 Pronker, E. S. *et al.* Risk in vaccine research and development quantified. *PloS one* **8**, e57755-e57755, doi:10.1371/journal.pone.0057755 (2013).

11 Geng, H. *et al.* Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Computational and Structural Biotechnology Journal* **17**, 1162-1170, doi:https://doi.org/10.1016/j.csbj.2019.07.010 (2019).

12 Shaw, D. E. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **330**, 341-346, doi:10.1126/science.1187409 (2010).

13 Gront, D., Kmiecik, S., Blaszczyk, M., Ekonomiuk, D. & Koliński, A. Optimization of

protein models. *WIREs Computational Molecular Science* **2**, 479-493, doi:https://doi.org/10.1002/wcms.1090 (2012).

14 Fukunishi, Y. Structural ensemble in computational drug screening. *Expert Opinion on Drug Metabolism & Toxicology* **6**, 835-849, doi:10.1517/17425255.2010.486399 (2010).

15 Ivetac, A. & McCammon, J. A. A molecular dynamics ensemble-based approach for the mapping of druggable binding sites. *Methods in Molecular Biology* **819**, 3-12, doi:10.1007/978-1-61779-465-0_1 (2012).

16 Hospital, A. *et al.* Molecular dynamics simulations: advances and applications. *Advances and Application in Bioinformatics and Chemistry* **8**, 37-47, doi:10.2147/AABC.S70333 (2015).

17 Zhang, D. & Lazim, R. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports* **7**, 44651, doi:10.1038/srep44651 (2017).

18 Pikkemaat, M. G. *et al.* Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering, Design and Selection* **15**, 185-192, doi:10.1093/protein/15.3.185 (2002).

19 Ramírez-Salinas, G. L. *et al.* Bioinformatics design and experimental validation of influenza A virus multi-epitopes that induce neutralizing antibodies. *Archives of Virology* **165**, 891-911, doi:10.1007/s00705-020-04537-2 (2020).

20 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).

21 Shamriz, S. & Ofoghi, H. Design, structure prediction and molecular dynamics simulation of a fusion construct containing malaria pre-erythrocytic vaccine candidate, PfCelTOS, and human interleukin 2 as adjuvant. *BMC Bioinformatics* **17**, 71-71, doi:10.1186/s12859-016-0918-8 (2016).

22 Schaller, A. *et al.* Computational study of elements of stability of a four-helix bundle protein biosurfactant. *Journal of Computer Aided Molecular Design* **29**, 47-58, doi:10.1007/s10822-014-9803-6 (2015).

23 Bekker, G.-J., Ma, B. & Kamiya, N. Thermal stability of single-domain antibodies estimated by molecular dynamics simulations. *Protein Science* **28**, 429-438, doi:10.1002/pro.3546 (2019).

24 Amir, M. *et al.* Structural Analysis and Conformational Dynamics of STN1 Gene Mutations Involved in Coat Plus Syndrome. *Frontiers in Molecular Biosciences* **6**,

doi:10.3389/fmolb.2019.00041 (2019).

25 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).

26 Kelly, S. M. *et al.* How to study proteins by circular dichroism. *Biochimica et Biophysica Acta* **1751**, 119-139, doi:10.1016/j.bbapap.2005.06.005 (2005).

27 Hasan, M. *et al.* Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infection, Genetics and Evolution* **70**, 140-157, doi:https://doi.org/10.1016/j.meegid.2019.03.003 (2019).

28 Gupta, A. *et al.* Vaccine candidate designed against carcinoembryonic antigen-related cell adhesion molecules using immunoinformatics tools. *Journal of Biomolecular Structure and Dynamics*, 1-15, doi:10.1080/07391102.2020.1797539 (2020).

29 Rostamian, M. *et al.* Immunoinformatics and molecular dynamics studies to predict T-cell-specific epitopes of four Klebsiella pneumoniae fimbriae antigens. *Journal of Biomolecular Structure and Dynamics*, 1-11, doi:10.1080/07391102.2020.1810126 (2020).

30 Gupta, N. & Kumar, A. Identification of Potent vaccine candidates against campylobacter jejuni using immunoinformatics approach. *International Journal of Peptide Research and Therapeutics* **26**, 1303-1312, doi:10.1007/s10989-019-09933-0 (2020).

31 Samad, A. *et al.* Designing a multi-epitope vaccine against SARS-CoV-2: an immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1-17, doi:10.1080/07391102.2020.1792347 (2020).

32 Calis, J. J. A. *et al.* Properties of MHC class I presented peptides that enhance immunogenicity. *PLOS Computational Biology* **9**, e1003266, doi:10.1371/journal.pcbi.1003266 (2013).

33 Rammensee, H. *et al.* SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219, doi:10.1007/s002510050595 (1999).

34 Singh, H. & Raghava, G. P. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* **19**, 1009-1014, doi:10.1093/bioinformatics/btg108 (2003).

35 Solanki, V. & Tiwari, V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Scientific Reports* **8**, 9044, doi:10.1038/s41598-018-26689-7 (2018).

36 Meng, X.-Y. *et al.* Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design* **7**, 146-157,

doi:10.2174/157340911795677602 (2011).

37 Huey, R. *et al*. A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry* **28**, 1145-1152, doi:https://doi.org/10.1002/jcc.20634 (2007).

38 Maupetit, J. *et al.* A fast method for large-scale De Novo peptide and miniprotein structure prediction. *Journal of Computational Chemistry* **31**, 726-738, doi:https://doi.org/10.1002/jcc.21365 (2010).

39 Wang, Z. *et al.* APOLLO: a quality assessment service for single and multiple protein models. *Bioinformatics* **27**, 1715-1716, doi:10.1093/bioinformatics/btr268 (2011).

40 Lopéz-Blanco, J. R. *et al.* iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* **27**, 2843-2850, doi:10.1093/bioinformatics/btr497 (2011).

41 Bai, Q. & Yao, X. Investigation of allosteric modulation mechanism of metabotropic glutamate receptor 1 by molecular dynamics simulations, free energy and weak interaction analysis. *Scientific Reports* **6**, 21763, doi:10.1038/srep21763 (2016).

42 Bello, M. & Correa-Basurto, J. Molecular dynamics simulations to provide insights into epitopes coupled to the soluble and membrane-bound MHC-II complexes. *PLOS ONE* **8**, e72575, doi:10.1371/journal.pone.0072575 (2013).

43 Chen, J. *et al.* Molecular mechanism and energy basis of conformational diversity of antibody SPE7 Revealed by molecular dynamics simulation and principal component analysis. *Scientific Reports* **6**, 36900, doi:10.1038/srep36900 (2016).

44 Knapp, B. *et al.* Current status and future challenges in T-cell receptor/peptide/MHC molecular dynamics simulations. *Briefings in Bioinformatics* **16**, 1035-1044, doi:10.1093/bib/bbv005 (2015).

45 Usman Mirza, M. *et al.* Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific Reports* **6**, 37313, doi:10.1038/srep37313 (2016).

46 Singaravelu, M. *et al.* Molecular dynamics simulations of lectin domain of FimH and immunoinformatics for the design of potential vaccine candidates. *Computational Biology and Chemistry* **52**, 18-24, doi:https://doi.org/10.1016/j.compbiolchem.2014.08.002 (2014).

47 Enayatkhani, M. *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *Journal of Biomolecular Structure and Dynamics*, 1-16, doi:10.1080/07391102.2020.1756411 (2020).

48 Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **10**, 449-461, doi:10.1517/17460441.2015.1032936 (2015).

49 Nain, Z. *et al.* Immunoinformatic and dynamic simulation-based designing of a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis*. *bioRxiv*, 758219, doi:10.1101/758219 (2019).

50 Gupta, N. & Kumar, A. Designing an efficient multi-epitope vaccine against Campylobacter jejuni using immunoinformatics and reverse vaccinology approach. *Microbial Pathogenesis* **147**, 104398, doi:https://doi.org/10.1016/j.micpath.2020.104398 (2020).

51 Pavitrakar, D. V. *et al.* Design of a multi-epitope peptide vaccine candidate against chandipura virus: an immuno-informatics study. *Journal of Biomolecular Structure and Dynamics*, 1-12, doi:10.1080/07391102.2020.1816493 (2020).

52 Deng, H. *et al.* Development of a multivalent enterovirus subunit vaccine based on immunoinformatic design principles for the prevention of HFMD. *Vaccine* **38**, 3671-3681, doi:https://doi.org/10.1016/j.vaccine.2020.03.023 (2020).

53 Banisharif-Dehkordi, F. *et al.* Design and molecular dynamic simulation of a new double-epitope tolerogenic protein as a potential vaccine for multiple sclerosis disease. *Research in Pharmaceutical Science* **14**, 20-26, doi:10.4103/1735-5362.251849 (2019).

54 Kumar, N. *et al.* Design and optimization of a subunit vaccine targeting COVID-19 molecular shreds using an immunoinformatics framework. *RSC Advances* **10**, 35856-35872, doi:10.1039/D0RA06849G (2020).

55 Kumar, N. *et al.* Antimicrobial Peptide designing and optimization employing large-scale flexibility analysis of protein-peptide fragments. *ACS Omega* **4**, 21370-21380, doi:10.1021/acsomega.9b03035 (2019).

56 Sood, D. *et al.* Mechanistic interaction study of bromo-noscapine with bovine serum albumin employing spectroscopic and chemoinformatics approaches. *Scientific Reports* **8**, 16964, doi:10.1038/s41598-018-35384-6 (2018).

57 López-Blanco, J. R. *et al.* iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Research* **42**, W271-W276, doi:10.1093/nar/gku339 (2014).

58 Flower, D. R. *et al.* T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. *Immunome Research* **6 Suppl 2**, S4-S4, doi:10.1186/1745-7580-6-S2-S4 (2010).

59 Dror, R. O. *et al.* Exploring atomic resolution physiology on a femtosecond to

millisecond timescale using molecular dynamics simulations. *Journal of General Physiology* **135**, 555-562, doi:10.1085/jgp.200910373 (2010).

60 Abhinav, B. *et al.* in *2008 IEEE International Symposium on Parallel and Distributed Processing*. 1-12.

61 Hess, B. *et al.* GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **4**, 435-447, doi:10.1021/ct700301q (2008).

62 Klepeis, J. L. *et al.* Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* **19**, 120-127, doi:10.1016/j.sbi.2009.03.004 (2009).

63 Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197, doi:10.1021/ja00124a002 (1995).

64 Guvench, O. & MacKerell, A. D., Jr. Comparison of protein force fields for molecular dynamics simulations. *Methods in Molecular Biology* **443**, 63-88, doi:10.1007/978-1-59745-177-2_4 (2008).

65 Zhang, H. *et al.* Limitations of *ab initio* predictions of peptide binding to MHC class II molecules. *PLOS ONE* **5**, e9272, doi:10.1371/journal.pone.0009272 (2010).

66 Peters, B. *et al.* A Community Resource Benchmarking Predictions of Peptide Binding to MHC-I Molecules. *PLOS Computational Biology* **2**, e65, doi:10.1371/journal.pcbi.0020065 (2006).

67 John, B. & Sali, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research* **31**, 3982-3992, doi:10.1093/nar/gkg460 (2003).

68 Qu, X. *et al.* A guide to template based structure prediction. *Current Protein and Peptide Science* **10**, 270-285, doi:10.2174/138920309788452182 (2009).

69 Fiser, A. Template-based protein structure modeling. *Methods in Molecular Biology* **673**, 73-94, doi:10.1007/978-1-60761-842-3_6 (2010).

70 Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

71 Ji, X. *et al.* Homology modeling and molecular dynamics simulation studies of a marine alkaline protease. *Bioinformatics and Biology Insights* **6**, 255-263, doi:10.4137/BBI.S10663 (2012).

72 Yang, H.-C. *et al.* Homology modeling and molecular dynamics simulation combined

with X-ray solution scattering defining protein structures of thromboxane and prostacyclin synthases. *The Journal of Physical Chemistry B* **121**, 11229-11240, doi:10.1021/acs.jpcb.7b08299 (2017).

73 Lemer, C. M. *et al.* Protein structure prediction by threading methods: evaluation of current techniques. *Proteins* **23**, 337-355, doi:10.1002/prot.340230308 (1995).

74 Peng, J. & Xu, J. Boosting protein threading accuracy. *Research in Computational Molecular Biology* **5541**, 31-45, doi:10.1007/978-3-642-02008-7_3 (2009).

75 Cheung, N. J. & Yu, W. *De novo* protein structure prediction using ultra-fast molecular dynamics simulation. *PLOS ONE* **13**, e0205819, doi:10.1371/journal.pone.0205819 (2018).

76 Childers, M. C. & Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering* **2**, 9-33, doi:10.1039/C6ME00083E (2017).

77 Li, J. *et al.* An improved integration of template-based and template-free protein structure modeling methods and its assessment in CASP11. *Protein Peptide Letter* **22**, 586-593, doi:10.2174/0929866522666150520145717 (2015).

78 Gill, M. & McCully, M. E. Molecular dynamics simulations suggest stabilizing mutations in a *de novo* designed α/β protein. *Protein Engineering, Design and Selection* **32**, 317-329, doi:10.1093/protein/gzaa005 (2019).

79 Bradley, P. *et al.* Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309**, 1868, doi:10.1126/science.1113801 (2005).

80 Rohl, C. A. *et al.* in *Methods in Enzymology* Vol. 383 66-93 (Academic Press, 2004).

81 Xu, D. & Zhang, Y. *Ab initio* protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* **80**, 1715-1735, doi:10.1002/prot.24065 (2012).

82 Kato, K. *et al.* Validation of molecular dynamics simulations for prediction of three-dimensional structures of small proteins. *Molecules* **22**, doi:10.3390/molecules22101716 (2017).

83 Vallat, B. *et al.* Modularity of protein folds as a tool for template-free modeling of structures. *PLOS Computational Biology* **11**, e1004419, doi:10.1371/journal.pcbi.1004419 (2015).

84 Shamriz, S. & Ofoghi, H. Design, structure prediction and molecular dynamics simulation of a fusion construct containing malaria pre-erythrocytic vaccine candidate, PfCelTOS, and human interleukin 2 as adjuvant. *BMC Bioinformatics* **17**, 71,

doi:10.1186/s12859-016-0918-8 (2016).

85 von Holst, H. & Li, X. Numerical impact simulation of gradually increased kinetic energy transfer has the potential to break up folded protein structures resulting in cytotoxic brain tissue edema. *Journal of Neurotrauma* **30**, 1192-1199, doi:10.1089/neu.2012.2730 (2013).

86 Mobini, S. *et al.* Computational design of a novel VLP-based vaccine for hepatitis B virus. *Frontiers in Immunology* **11**, doi:10.3389/fimmu.2020.02074 (2020).

87 Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annual Review of Biochemistry* **77**, 363-382, doi:10.1146/annurev.biochem.77.062906.171838 (2008).

88 Shen, Y. & Bax, A. Homology modeling of larger proteins guided by chemical shifts. *Nature Methods* **12**, 747-750, doi:10.1038/nmeth.3437 (2015).

89 Kim, D. E. *et al.* Sampling bottlenecks in *de novo* protein structure prediction. *Journal of Molecular Biology* **393**, 249-260, doi:https://doi.org/10.1016/j.jmb.2009.07.063 (2009).

90 Söding, J. Big-data approaches to protein structure prediction. *Science* **355**, 248, doi:10.1126/science.aal4512 (2017).

91 Freddolino, P. L. *et al.* Challenges in protein folding simulations: Timescale, representation, and analysis. *Nature Physics* **6**, 751-758, doi:10.1038/nphys1713 (2010).

92 Schaeffer, R. D. *et al.* Combining experiment and simulation in protein folding: closing the gap for small model systems. *Current opinion in structural biology* **18**, 4-9, doi:10.1016/j.sbi.2007.11.007 (2008).

93 Gershenson, A. *et al.* Successes and challenges in simulating the folding of large proteins. *The Journal of biological chemistry* **295**, 15-33, doi:10.1074/jbc.REV119.006794 (2020).

94 Wang, G. & Dunbrack, R. L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589-1591, doi:10.1093/bioinformatics/btg224 (2003).

95 Zheng, W. *et al.* Frustration in the energy landscapes of multidomain protein misfolding. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1680-1685, doi:10.1073/pnas.1222130110 (2013).

96 Han, J. H. *et al.* The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology* **8**, 319-330, doi:10.1038/nrm2144 (2007).

97 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

98 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-

track neural network. *Science* **373**, 871-876, doi:10.1126/science.abj8754 (2021).

99 Sviben, D. *et al.* Investigation of the thermal shift assay and its power to predict protein and virus stabilizing conditions. *Journal of Pharmaceutical and Biomedical Analysis* **161**, 73-82, doi:https://doi.org/10.1016/j.jpba.2018.08.017 (2018).

100 Blundell, T. L. *et al.* Protein engineering and design. *Philosophical Transactions of the Royal Society of London B Biological Science* **324**, 447-460, doi:10.1098/rstb.1989.0059 (1989).

101 Hellinga, H. W. Computational protein engineering. *Nature Structural Biology* **5**, 525-527, doi:10.1038/776 (1998).

102 Jiang, Z. *et al.* Effects of an Electric Field on the Conformational Transition of the Protein: A Molecular Dynamics Simulation Study. *Polymers (Basel)* **11**, 282, doi:10.3390/polym11020282 (2019).

103 Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology* **235**, 625-634, doi:10.1006/jmbi.1994.1017 (1994).

104 Chen, Z. *et al.* Molecular Dynamics Simulation of Barnase: Contribution of Noncovalent Intramolecular Interaction to Thermostability. *Mathematical Problems in Engineering* **2013**, 504183, doi:10.1155/2013/504183 (2013).

105 Manjunath, K. & Sekar, K. Molecular Dynamics Perspective on the Protein Thermal Stability: A Case Study Using SAICAR Synthetase. *Journal of Chemical Information and Modeling* **53**, 2448-2461, doi:10.1021/ci400306m (2013).

106 Kumar, C. V. *et al.* Computational Analysis Reveals the Association of Threonine 118 Methionine Mutation in PMP22 Resulting in CMT-1A. *Adv Bioinformatics* **2014**, 502618-502618, doi:10.1155/2014/502618 (2014).

107 Lobanov, M. Y. *et al.* Radius of gyration as an indicator of protein structure compactness. *Molecular Biology* **42**, 623-628, doi:10.1134/S0026893308040195 (2008).

108 Arnittali, M. *et al.* Structure of biomolecules through molecular dynamics simulations. *Procedia Computer Science* **156**, 69-78, doi:https://doi.org/10.1016/j.procs.2019.08.181 (2019).

109 Amir, M. *et al.* Structural analysis and conformational dynamics of stn1 gene mutations involved in coat plus syndrome. *Frontiers in Molecular Biosciences* **6**, 41-41, doi:10.3389/fmolb.2019.00041 (2019).

110 Zhao, Y., *et al.* Molecular dynamics simulation reveals insights into the mechanism of

unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. *PLOS ONE* **10**, e0124377, doi:10.1371/journal.pone.0124377 (2015).

111 Burton, B. *et al.* A computational investigation on the connection between dynamics properties of ribosomal proteins and ribosome assembly. *PLOS Computational Biology* **8**, e1002530, doi:10.1371/journal.pcbi.1002530 (2012).

112 Abdulazeez, S. Molecular simulation studies on B-cell lymphoma/leukaemia 11A (BCL11A). *American Journal of Translational Research* **11**, 3689-3697 (2019).

113 Chapman, A. D. *et al.* Structural basis of substrate specificity in malate dehydrogenases: crystal structure of a ternary complex of porcine cytoplasmic malate dehydrogenase, alpha-ketomalonate and tetrahydoNAD. *Journal of Molecular Biology* **285**, 703-712, doi:10.1006/jmbi.1998.2357 (1999).

114 Dong, Y. *et al.* Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences* **115**, 1274, doi:10.1073/pnas.1718910115 (2018).

115 Best, R. B. *et al.* Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of National Academy of Sciences of the United States of America* **110**, 17874-17879, doi:10.1073/pnas.1311599110 (2013).

116 Xu, Y. *et al.* Stabilizing effect of inherent knots on proteins revealed by molecular dynamics simulations. *Biophysical Journal* **115**, 1681-1689, doi:https://doi.org/10.1016/j.bpj.2018.09.015 (2018).

117 Roy, S. *et al.* The unfolding MD simulations of cyclophilin: Analyzed by surface contact networks and their associated metrics. *PLOS ONE* **10**, e0142173, doi:10.1371/journal.pone.0142173 (2015).

118 Pace, C. N. *et al.* Contribution of hydrogen bonds to protein stability. *Protein Science* **23**, 652-661, doi:10.1002/pro.2449 (2014).

119 Chikalov, I. *et al.* Learning probabilistic models of hydrogen bond stability from molecular dynamics simulation trajectories. *BMC Bioinformatics* **12 Suppl 1**, S34-S34, doi:10.1186/1471-2105-12-S1-S34 (2011).

120 Ali, S. A. *et al.* A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. *Curr Protein Pept Sci* **15**, 456-476, doi:10.2174/1389203715666140327114232 (2014).

121 Berhanu, W. M. & Hansmann, U. H. E. Side-chain hydrophobicity and the stability of $A\beta_{16-22}$ aggregates. *Protein Sci* **21**, 1837-1848, doi:10.1002/pro.2164 (2012).

122 Shy, M. E. *et al.* T118M PMP22 mutation causes partial loss of function and HNPP-

like neuropathy. *Annals of Neurology* **59**, 358-364, doi:10.1002/ana.20777 (2006).

123 Muneeswaran, G. *et al.* Molecular dynamics simulation approach to explore atomistic molecular mechanism of peroxidase activity of apoptotic cytochrome c variants. *Informatics in Medicine Unlocked* **11**, 51-60, doi:https://doi.org/10.1016/j.imu.2018.04.003 (2018).

124 Ali, S. K. *et al.* Molecular dynamics-based analyses of the structural instability and secondary structure of the fibrinogen gamma chain protein with the D356V mutation. *Journal of Biomolecular Structure and Dynamics* **35**, 2714-2724, doi:10.1080/07391102.2016.1229634 (2017).

125 Medvedev, K. E. *et al.* Molecular dynamics simulations of the Nip7 proteins from the marine deep- and shallow-water Pyrococcus species. *BMC Structural Biology* **14**, 23, doi:10.1186/s12900-014-0023-z (2014).

126 Janin, J. *et al.* Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* **125**, 357-386, doi:https://doi.org/10.1016/0022-2836(78)90408-4 (1978).

127 Anderson, R. J. *et al.* Main-chain conformational tendencies of amino acids. *Proteins: Structure, Function, and Bioinformatics* **60**, 679-689, doi:10.1002/prot.20530 (2005).

128 Hollingsworth, S. A. & Karplus, P. A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *BioMolecular Concepts* **1**, 271-283, doi:10.1515/bmc.2010.022 (2010).

129 Matsuura, Y. *et al.* Evaluating the strengths of salt bridges in the CutA1 protein using molecular dynamic simulations: a comparison of different force fields. *FEBS Open Bio* **9**, 1939-1956, doi:10.1002/2211-5463.12731 (2019).

130 Pendley, S. S. *et al.* Molecular dynamics guided study of salt bridge length dependence in both fluorinated and non-fluorinated parallel dimeric coiled-coils. *Proteins* **74**, 612-629, doi:10.1002/prot.22177 (2009).

131 Luo, Y. & Baldwin, R. L. How Ala-->Gly mutations in different helices affect the stability of the apomyoglobin molten globule. *Biochemistry* **40**, 5283-5289, doi:10.1021/bi010122j (2001).

132 Beeby, M. *et al.* The Genomics of Disulfide Bonding and Protein Stabilization in Thermophiles. *PLOS Biology* **3**, e309, doi:10.1371/journal.pbio.0030309 (2005).

133 Dürrschmidt, P. *et al.* Differentiation between conformational and autoproteolytic stability of the neutral protease from Bacillus stearothermophilus containing an engineered disulfide bond. *European Journal of Biochemistry* **268**, 3612-3618, doi:10.1046/j.1432-1327.2001.02270.x (2001).

134 Chen, J. *et al.* Improving stability of nitrile hydratase by bridging the salt-bridges in specific thermal-sensitive regions. *Journal of Biotechnology* **164**, 354-362, doi:https://doi.org/10.1016/j.jbiotec.2013.01.021 (2013).

135 Makhatadze, G. I. *et al.* Contribution of Surface Salt Bridges to Protein Stability: Guidelines for Protein Engineering. *Journal of Molecular Biology* **327**, 1135-1148, doi:https://doi.org/10.1016/S0022-2836(03)00233-X (2003).

136 Missimer, J. H. *et al.* Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Science* **16**, 1349-1359, doi:10.1110/ps.062542907 (2007).

137 Thomas, A. S. & Elcock, A. H. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *Journal of the American Chemical Society* **126**, 2208-2214, doi:10.1021/ja039159c (2004).

138 Papaleo, E. *et al.* Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. *Journal of Molecular Graphics and Modelling* **26**, 93-103, doi:https://doi.org/10.1016/j.jmgm.2006.09.012 (2007).

139 Vinther, J. M. *et al.* Enhanced stability of a protein with increasing temperature. *Journal of the American Chemical Society* **133**, 271-278, doi:10.1021/ja105388k (2011).

140 Dong, H. *et al.* Hydrophobic effect on the stability and folding of a hyperthermophilic protein. *Journal of Molecular Biology* **378**, 264-272, doi:https://doi.org/10.1016/j.jmb.2008.02.039 (2008).

141 Pace, C. N. *et al.* Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology* **408**, 514-528, doi:https://doi.org/10.1016/j.jmb.2011.02.053 (2011).

142 Paiardini, A. *et al.* "Hot cores" in proteins: Comparative analysis of the apolar contact area in structures from hyper/thermophilic and mesophilic organisms. *BMC Structural Biology* **8**, 14, doi:10.1186/1472-6807-8-14 (2008).

143 Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Molecular Biology Reviews* **65**, 1-43, doi:10.1128/MMBR.65.1.1-43.2001 (2001).

144 Jaenicke, R. & Böhm, G. The stability of proteins in extreme environments. *Current Opinion in Structural Biology* **8**, 738-748, doi:https://doi.org/10.1016/S0959-440X(98)80094-8 (1998).

145 Vogt, G. & Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Folding and Design* **2**, S40-S46, doi:https://doi.org/10.1016/S1359-0278(97)00062-X

(1997).

146 Vogt, G. *et al.* Protein thermal stability, hydrogen bonds, and ion pairs. Edited by F. E. Cohen. *Journal of Molecular Biology* **269**, 631-643, doi:https://doi.org/10.1006/jmbi.1997.1042 (1997).

147 Zhou, H.-X. Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophysical Journal* **83**, 3126-3133, doi:10.1016/S0006-3495(02)75316-2 (2002).

148 Li, J. *et al.* A stable α-helix-rich intermediate is formed by a single mutation of the β-sheet protein, src SH3, at pH 3. *Journal of Molecular Biology* **372**, 747-755, doi:https://doi.org/10.1016/j.jmb.2007.07.001 (2007).

149 Meharenna, Y. T. & Poulos, T. L. Using molecular dynamics to probe the structural basis for enhanced stability in thermal stable cytochromes P450. *Biochemistry* **49**, 6680-6686, doi:10.1021/bi100929x (2010).

150 Bae, E. & Phillips, G. N., Jr. Identifying and engineering ion pairs in adenylate kinases. Insights from molecular dynamics simulations of thermophilic and mesophilic homologues. *Journal of Biological Chemistry* **280**, 30943-30948, doi:10.1074/jbc.M504216200 (2005).

151 Wang, Z. *et al.* Ferritin nanocage-based antigen delivery nanoplatforms: epitope engineering for peptide vaccine design. *Biomaterials Science* **7**, 1794-1800, doi:10.1039/C9BM00098D (2019).

152 Qu, Y. *et al.* Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments. *The Journal of Physical Chemistry B* **125**, 3830-3842, doi:10.1021/acs.jpcb.1c00276 (2021).

153 Han, J. A. *et al.* Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. *Nanomedicine* **10**, 561-569, doi:10.1016/j.nano.2013.11.003 (2014).

154 Bai, Y. *et al.* Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7292-7296, doi:10.1073/pnas.0500062102 (2005).

155 Klein, J. S. *et al.* Design and characterization of structured protein linkers with differing flexibilities. *Protein Engineering Design and Selection* **27**, 325-330, doi:10.1093/protein/gzu043 (2014).

156 Zhang, J. *et al.* Design and optimization of a linker for fusion protein construction. *Progress in Natural Science* **19**, 1197-1200,

doi:https://doi.org/10.1016/j.pnsc.2008.12.007 (2009).

157 Chen, X. *et al.* Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

158 Mahapatra, S. R. *et al.* Designing an efficient multi-epitope vaccine displaying interactions with diverse HLA molecules for an efficient humoral and cellular immune response to prevent COVID-19 infection. *Expert Review of Vaccines*, 1-15, doi:10.1080/14760584.2020.1811091 (2020).

159 Foged, C. Subunit vaccines of the future: the need for safe, customized and optimized particulate delivery systems. *Therapeutic Delivery* **2**, 1057-1077, doi:10.4155/tde.11.68 (2011).

160 Wei, Y. *et al.* Evaluation of lumazine synthase from Bacillus anthracis as a presentation platform for polyvalent antigen display. *Protein Science* **26**, 2059-2072, doi:10.1002/pro.3243 (2017).

161 Rezaie, E. *et al.* Application of molecular dynamics simulations to design a dual-purpose oligopeptide linker sequence for fusion proteins. *Journal of Molecular Modeling* **24**, 313, doi:10.1007/s00894-018-3846-x (2018).

162 Mohammadi, M. *et al. In silico* analysis of three different tag polypeptides with dual roles in scFv antibodies. *Journal of Theoretical Biology* **402**, 100-106, doi:https://doi.org/10.1016/j.jtbi.2016.04.016 (2016).

163 Caflisch, A. & Paci, E. *Molecular dynamics simulations to study protein folding and unfolding*. (2005).

164 Robinson-Mosher, A. *et al.* Dynamics simulations for engineering macromolecular interactions. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**, 025110, doi:10.1063/1.4810915 (2013).

PUBLICATIONS DURING CANDIDATURE

Published

1. **Qu, Y.**; Wang, L.; Yin, S.; Zhang, B.; Jiao, Y.; Sun, Y.; Middelberg, A.; Bi, J. Stability of Engineered Ferritin Nanovaccines Investigated by Combined Molecular Simulation and Experiments. *The Journal of Physical Chemistry B* 2021, 125, 3830-3842.

2. **Qu, Y.**; Zhang, B.; Wang, Y.; Yin, S.; Pederick, J. L.; Bruning, J. B.; Sun, Y.; Middelberg, A.; Bi, J. Immunogenicity study of engineered ferritins with C- and N-terminus insertion of Epstein-Barr nuclear antigen 1 epitope. *Vaccine* 2021, 39, 4830-4841.

3. **Qu, Y.**; Zhang, B.; Wang, Y.; Yin, S.; Sun, Y.; Middelberg, A.; Bi, J. Immunogenicity and Vaccine Efficacy Boosted by Engineering Human Heavy Chain Ferritin and Chimeric Hepatitis B Virus Core Nanoparticles. *ACS Applied Bio Materials* 2021, 4, 7147-7156.

4. Yin, S., Wang, Y., Zhang, B., **Qu, Y.**, Liu, Y., Dai, S., Zhang, Y., Wang, Y. and Bi, J. Engineered ferritin with half-life extension and tumor targeting by PAS and RGDK peptide functionalization. *Pharmaceutics* 2021, 13(4), 521.

5. Yin, S., Liu, Y., Zhang, B., **Qu, Y.**, Zhang, Y., Choe, W., Dai, S. and Bi, J. Mechanism study of thermally induced anti-tunor drug loading to engineered human heavy-cchain ferritin nanocages aided by computational analysis. *Biosensors* 2021, 11(11), 444.

In progress

6. **Qu, Y.**; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Impact of flexible linker length on protein stability of engineered ferritin as a vaccine carrier. Submitted to *International Journal of Biological Macromolecules*.

7. **Qu, Y.**; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Engineered design of helix E structure on ferritin nanoparticles. Submitted to *ACS Applied Bio Materials*.

8. **Qu, Y.**; Davey, K; Sun, Y.; Middelberg, A.; Bi, J. Using molecular dynamics simulation (MDS) to predict performance of Epitope-based chimeric vaccine: a critical review. Submitted to *Journal of Chemical Information and Modeling*.