THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Beyond Antibodies: Development of a Novel Molecular Scaffold Based on Human Chaperonin 10

Abdulkarim Mohammed Alsultan

B.Pharm, M.Biotech

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2014

Australian Institute for Bioengineering and Nanotechnology (AIBN)

Abstract

This study reports the development of a new molecular scaffold based on human Chaperonin 10 (**hCpn10**), for the development of protein-based new molecular entities (NMEs) of diagnostic and therapeutic potential. The aims were to establish the fundamental basis of a molecular design for the scaffold, to enable the display of non-native peptides with binding activity to selected targets, while maintaining structural integrity and native heptameric conformation.

Recently there has been much global interest in developing NMEs of protein and peptide origin, as therapeutic agents and diagnostic reagents. A class of NMEs are based on molecular scaffolds, whereby peptides with binding activity to a given target are displayed on a protein backbone or scaffold. The molecular scaffold provides a rigid folding unit which spatially brings together several exposed peptide loops, forming an extended interface that ensures tight binding of the target. Monoclonal antibodies (mAbs) are in essence molecular scaffolds, whereby complementarity determining regions, otherwise known as CDR peptide loops, extend from a framework and contact antigen. mAbs have been widely utilised in the life sciences and biopharmaceutical industries for the development of diagnostic probes and biologic medicines, respectively. However the current patent landscape surrounding mAbs is complex and commercialisation of newly developed antibodies can be hampered by licensing agreements and royalty stacking.

Molecular scaffolds are an alternative to antibodies, and some of the scaffolds in development include lipocalins, fibronectin domain, DARPins consensus repeat domain and avimers, to name a few. hCpn10 is a homo-oligomer composed of seven subunits. Each hCpn10 monomer is composed of a β-barrel core structure that is flanked by two flexible loops, known as the β-hairpin and mobile loop. It is proposed the mobile could be substituted with peptides and so increase the apparent affinity to a given target compared to that of naked peptides, due to avidity conferred by the heptameric structure. This study focussed on assessing the properties of hCpn10 and determining whether this molecule could be utilised as a scaffold for the development of NMEs that have binding activity to selected targets.

In the first example of the utility of hCpn10 as a molecular scaffold, the E-76 peptide, which binds to Factor VIIa (FVIIa) of the extrinsic coagulation cascade pathway, was substituted into the mobile loop region of hCpn10 creating a NME designated **CE76** with anticoagulant activity. Anticoagulants prevent blood from clotting and have clinical applications associated with coronary heart disease and stroke. Common anticoagulants include heparin, warfarin, and dicoumarol. Although these compounds have proved beneficial they are not without complications and side effects. An NME has been created through display of the E-76 peptide on the hCpn10 scaffold, creating the highly avid CE76, which binds FVIIa with sub-nanomolar affinity.

Prior to creation of CE76 NME, molecular dynamic (MD) simulation studies were conducted in order to optimise linkers at the *N*- and *C*-junctures of the mobile loop, so as to optimise mobile loop orientation and conserve native-like heptamer formation. Simulation studies revealed that inserting proline residues at both ends of the mobile loop region stabilized the new CE76 variants (named $CE76_{P1}$) and facilitated heptamer assembly. Characterisation by size exclusion chromatography confirmed $CE76_{P1}$ formed a stable heptamer, and bound specifically to the exosite of FVIIa, and subsequently inhibited the activation of Factor X (IC_{50} 1.5 nM). $CE76_{P1}$ prologed the extrinsic pathway 4-fold higher than E-76 alone, while not affecting the intrinsic coagulation cascade as determined by prothrombin time (PT) and thromboelastography (TEG) assays. $CE76_{P1}$ has the potential to be a new biologic drug with potent anticoagulant properties.

The utility of hCpn10 as a molecular scaffold was further assessed by incorporating P7 peptide into the mobile loop, creating a NME designated **CP7**. P7 peptide binds to CD44, which is overexpressed in cancers of the breast, colon, prostate and brain. Binding studies revealed that CP7 bound specifically to recombinant CD44 (K_D 1.32 nM), while no binding was detected to surface protein CD24. Immunofluorescence imaging and flow cytometry studies showed increased binding activity of CP7 to breast adenocarcinoma MDA-MB-486 cells compared to P7 peptide alone. Further expression studies using peptides that bind VEGFR and FGFR, were incorporated into hCpn10 scaffold also resulted in correct assembly of heptamer.

Based on these studies associated with the incorporation of selected peptides into the hCpn10 scaffold, it is evident that hCpn10 can be utilised as a molecular scaffold for the developing of NMEs with therapeutic and diagnostic potential. The demonstration that human Cpn10 is a useful molecular scaffold for the creation of NMEs in a “beyond antibody” approach offers the opportunity to create new intellectual property in the area of biologic medicines and bioanalytical reagents/diagnostics.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

No publications

Publications included in this thesis

No publications included

Contributions by others to the thesis

Dr. Herbert Treutlein' of Computist Bio-Nanotech, RMIT University, contributed to the research by conducting MD simulations presented in Chapter 3.

Dr. Sharon Du of the Translational Research Institute (TRI), School of Medicine, University of Queensland, contributed in designing clotting assays presented in Chapter 4.

Ms. Dinora Roche Recinos contributed to $CR2_{VEGF\text{-}P2}$ expression in Chapter 6.

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

"*The one who is not thankful for the few blessings will not be thankful for the many, and the one who is not thankful to the people will not be thankful to Allah*"

Prophet Mohammed (Peace be upon him and his progeny)

First and foremost, I am grateful and thankful to Almighty Allah "GOD" for His limitless mercy, as well as to the Messenger of Allah Prophet Mohammed and his progeny peace be upon them all for surrounding me with infinite blessing.

I express my deepest thanks to my guardians, supporters and advisors Associate Professor Stephen Mahler (principal supervisor) and Dr. David Chin (co-supervisor) for their encouragements and continued guidance during my study. It has been a privilege to work with you, to be under your supervisions and to learn more from you both.

To all my colleagues past and current members of Gray Group and National Biologics Facility (NBF) at the Australian Institute for Bioengineering and Nanotechnology (AIBN), especially: Professor Peter Gray (Group leader), Martina Jones, Christopher Howard, Sumukh Kumble, Matthew Smede, Vanessa Sandford, Liam Nolan, Christopher de Bakker, Karin Taylor, Kar Man Leung, Jeff Hou, Kebaneilwe Lebani, Marianne Gillard, Camila Orellana, Dinora Roche Recinos, Emily Chan, Gency Gunasingh, Jay Heise-Seabrook, Jiang Eric Zhu, Kym Hoger, Lyndon Raftery, Michael Song, Michael Yeh, Mohammed AlFaleh, Neetika Arora, Steve Goodall, Tim Ruder, Trent Munro, Veronica Martinez, Xiaoli Chen, Zainab Almansour, Nadya Shale, Edwin Huang, Aidan Bell, and David Crowley, "*a small thing in your eyes is so big in main"*, so thank you all for your help and your advices that you generously gave at various stages in my journey.

Enormous thanks to Sumukh Kumble for assisting in proofreading most of my thesis, and special thanks to Andrea Schaller for helping in proofreading the modelling in Chapter 3, and thank you both for your valuable suggestions.

I express my thanks to Dr. Herbert Treutlein (Computist Bio-Nanotech & RMIT University), for his assistance in conducting MD simulations and ongoing discussions. I deeply thank Associate Professor Paul Masci from the Translational Research Institute (TRI), and School

of Medicine, University of Queensland (UQ), for permitting full access to his haematology laboratory and equipment, and for supplying the important human plasma, blood and reagents. And also, heartfelt thanks to Dr. Sharon Du, for her generous assistance in designing clotting assays, data discussions and valuable feedbacks.

To, Michele Bruschi, Alexandra Depelsenaire, Suriana Sabri, Luke Budianto, Haryadi Sugiarto, Tony Miscamble and to all my friends at AIBN from various disciplined, thanks for sharing your valuable experiences, and your friendship that I will always appreciate.

To my beloved parents Mohammed and Safiya thank you for your limitless love, support and prayers.

To my soul-mate and lovely wife, Amani Jafar, thank you for your endless supports and incredible patience during this tough mission. You are a major part of this journey, and I could not have achieved it without your support and encouragement.

To my in-laws, Jafar and Asmaa, to my brothers and sisters, to my large family members back home in Saudi Arabia, to my friends here in Brisbane and back in Saudi Arabia, I am immensely thankful to you all.

I express my great thanks and appreciations to King Abduallah Bin Abdulaziz Scholarship Program (KASP), Ministry of Higher Education (MOHE), Kingdom of Saudi Arabia (KSA) for the scholarship and financial support. To all members of the Saudi Arabian Cultural Attaché here in Australia, thank you.

Lastly, I am grateful and thankful to all who have prayed, supported, helped, advised or have influenced me and my journey.

Keywords

Chaperonin 10, cpn10, protein engineering scaffold, biologics, biopharmaceutical, modelling and simulation, anticoagulant, Factor VIIa, cancer diagnostic probe, CD44.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 110106, Medical Biochemistry: Proteins and Peptides, 60%

ANZSRC code: 030402, Biomolecular Modelling and Design, 20%

ANZSRC code: 111504, Pharmaceutical Sciences, 20%

Fields of Research (FoR) Classification

FoR code: 1004, Medical Biotechnology, 80%
FoR code: 1115, Pharmacology and Pharmaceutical Sciences, 20%

Invention disclosure

Invention disclosures have been submitted to UniQuest, the commercial arm of the Unversity of Queensland, for protection of intellectual property, generated through the research outcomes achieved throughout PhD enrolment. Accordingly, research outcomes have not been disclosed through public domain publication nor presentation. Internal University of Queensland presentations have been conducted, in accordance with the Intellectual Property policy of UniQuest and the University of Queensland, to maintain confidentiality. Draft manuscripts have been prepared, and will be submitted for publication once intellectual property protection has been secured.

Publication arising from this thesis (draft manuscripts):

Abdulkarim M. Alsultan, Dinora R. Recinos, Christopher B. Howard, David Y. Chin, Martina L. Jones, and Stephen M. Mahler, (2015) **Engineering the Next-Generation of Non-Antibody Molecular Scaffolds.**

Abdulkarim M. Alsultan, David Y. Chin, Martina L. Jones, and Stephen M. Mahler, (2015) **Development of a Factor VIIa Inhibitor Based on a High Affinity Peptide Ligand Displayed on Human Chaperonin-10 Molecular Scaffold.**

Abdulkarim M. Alsultan, Christopher B. Howard, David Y. Chin, Martina L. Jones, and Stephen M. Mahler, (2015) **The Development of a Novel Cell Surface CD44 Detection Probe Based on Displaying P7-peptide Ligand on hCpn10-Molecular Scaffold.**

Dedication

This thesis is dedicated to my beloved young brother, **AbdulRahman**, who sadly passed away, before diagnostic test confirmed he had cancer.

1983-2013

Your smile will always be remembered

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

3D	Three dimensional
A	Adenine
Å	Angstrom
AA	Amino acid
Aβ	β-amyloid peptides
AcN	Acetonitrile
ADC	Antibody drug conjugates
ADCC	Antibody-dependent cellular cytotoxicity
Ala	Alanine
AMD	Age-related macular degeneration
Amp	Ampicillin
APCF	Australian Proteomics Computational Facility
APH	Aminoglycoside phosphotransferase
ApoD	Apolipoprotein D
APPI	β-amyloid protein precursor inhibitor
aPTT	activated Partial Thromboplastin Time
AR	Ankyrin repeat
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATIII	Antithrombin III
β2AR	β2-adrenoceptor
BBP	Biling binding protein
bFGF	Basic fibroblast growth factor
BLI	Bio-Layer Interferometry
bp	base pair
BSA	Bovine serum albumin
C	Cytosine
CD	Cluster of differentiation
CDC	Compliment-directed cytotoxicity
Cdk2	Cyclin-dependent kinase 2
CDR	Complementarity-determining region
CDRs	complementarity determining regions
CH	Heavy chain constant
CHARMM27	Chemistry at HARvard Macromolecular Mechanics 27 program
CHO	Chinese hamster ovary cells

CIAP	calf intestinal alkaline phophatase
CL	Light chain constant
cm	centi metre
Cpn10	Chaperonin 10
CPU	Central Processing Unit
CSC	Cancer stem cell
CTLA-4	Cytotoxic T-lymphocyte antigen-4
CV	Colum volume
Cyc	Cystine
Da	Dalton
DARPin	Designed Ankyrin Repeat Protein
DM	Defined media
DME	Diabetic macular edema
DNA	Desoxyribonucleic acid
Drl	*Drosophila* transmembrane receptor derailed
E. coli	*Escherichia Coli*
EDTA	Ethylenediaminetetraacetic acid
EGF2	Epidermal growth factor 2
ELISA	Enzyme-linked immunosorbent assay
ER	Estrogens receptor
FA	Formic acid
FACS	Fluorescence-activated cell sorting
FBS	Fetal Bovine Serum
Fc	Fragment crystallizable
FCM	Flow cytometry
FDA	Food and Drug Administration
Feb	Fragment antigen-binding
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FITC	Fluorescein isothiocyanate
FIX	Factor IX
FN3	Fibronectin type III
FRET	Fluorescence resonance energy transfer
fs	femto second
Fv	Variable fragment
FVII	Factor VII
FVIIa	Factor VII active

FX	Factor X
FXa	Factor X active
G	Guanine
g	gram
xg	gravity
GBIS	Generalised born implicit solvent
GF	Gel filtration chromatography
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GPCR	G-protein-coupled receptor
HAE	Hereditary Angioedema
hCpn10	Human Chaperonin 10
HEK	Human Embryonic Kidney Cells
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
HER2	Human epidermal growth factor receptor 2
HGF	Hepatocyte growth factor
His	Histidine
HIV	Human immunodeficiency virus
hr	Hour
Hsp10	Heat shock proteins 10
IEX	Ion exchange Chromatography
IgG	Immunoglobulin
IL-2	Interleukin-2
IL-6	Interleukin-6
Ile	Isoleucine
IMAC	Immobilized metal ion affinity chromatography
IP	Intellectual property
K	Kelvin
k_a	association constant
Kan	Kanamycin
K_D	Equilibrium dissociation constant
k_d	dissociation constant
kDa	Kilo dalton
KLK3	Kallikrein-related peptidases 3
L	Litre
LACI	Lipoprotein-associated coagulation inhibitor

LACI-D1	Lipoprotein-associated coagulation inhibitor domain-1
LB	Luria broth
LDLR-A	Low density lipoprotein receptor class A
Leu	Leucine
Lys	Lysine
M	Molar
mAb	Monoclonal antibody
MALDI-TOF	Matrix-assisted laser desorption-ionization time-of-flight
MAPKs	Mitogen-activated protein kinases
MBP	Maltose binding protein
MD	Molecular Dynamic
MES	4-Morpholineethanesulfonic acid
Met	Methionine
mg	milli gram
min	Mintue
mL	milli liter
mm	milli meter
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Mass Spectrum
MS	multiple sclerosis
Mwt	Molecular weight
NAMD2	NAnoscale Molecular Dynamics 2 program
NCS	Neocarzinostatin
nM	nano molar
nm	nano meter
NMDA	*N*-methyl-D-aspartate
NME	New molecular entity
NR1	Neurophilin1
ns	nano second
NSCLC	Non-small-cell lung carcinoma
PBS	Phosphate-Buffered Saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PFA	Parafromaldehyde
Phe	Phenylalanine
pI	Isoelectric point

PIGF	Placental growth factor
pM	pico molar
PMF	Peptide mass fingerprinting
Pro	Proline
ps	pico second
PSA	Prostate-specific antigen
PT	Prothrombin time
PVDF	Polyvinylidene fluoride
RA	Rheumatoid arthritis
RIT	Radioimmunotherapy
RNA	Ribonucleic acid
RT	Room temperature
RT-PCR	Real time polymerase chain reaction
scFv	Single-chain antibody
SDS-PAGE	Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis
sec	Second
SEC	Size exclusion chromatography
SE-HPLC	Size Exclusion-High performance liquid chromatography
Ser	Serine
SOC	Super Optimal Broth with Catabolite repression
T	Thymine
TAE	Tris-Acetate-EDTA
TB	Terrific broth
TBS	Tirs buffer saline
TEG	Thromboelastograph
TFPI	Tissue factor pathway inhibitor
Thr	Threonine
TNF-α	Tumor necrosis factor α
Trp	Tryptophan
TrxA	Thioredoxin
Tyr	Tyrosine
µg	micro gram
UK	United Kingdom
µL	micro liter
µM	micro molar
µm	micro meter
USA	United State of America

V	Voltage
v	velocity
v/v	Volume per volume
Val	Valine
VEGF	Vascular endothelial growth factor
VEGFR2	Vascular endothelial growth factor receptor 2
VH	Heavy chain variable
VHH	Variable domain of heavy chain antibody
VL	Light chain variable
VMD	Visual Molecular Dynamics program
w/v	Weight per volume
WHO	World Health Organization
wt	Wild type
Y2H	Yeast-two-hybrid display

Chapter 1. Engineering the Next-Generation of Non-Antibody Molecular Scaffolds

1.1 General introduction

Therapeutic agents and diagnostic reagents are a crucial component of the commercial activities of the Life Sciences and Biopharmaceutical industries. Monoclonal antibodies (mAbs) are the fastest growing class of biologic medicines, while mAbs are used extensively as components in various forms of diagnostic tests and immunologically-based assays, and as analytical reagents. mAbs immobilised on resins and matrices are also used for purifying recombinant protein biologics, and at the laboratory scale for purifying proteins of interest from tissue lysates or cell culture supernatants. The reason why mAbs are key products of these industries is that they are able to bind potentially to any molecular entity that exists in nature, including macromolecular structures, small organic molecules and some materials. The basic design of an antibody facilitates peptide loops to extend from the antibody scaffold and make contact with a target surface, resulting in binding.

Although mAbs are versatile molecular entities and have evolved to have exquisite specificity as nature's preferred scaffold, there are other molecular entities beyond antibodies that are able to bind specifically to a target, in a similar fashion to the manner that antibodies bind to their target. These molecular scaffolds or macrocycles bind to their target through the physicochemical interaction of a peptide loop or peptide domain with another surface, resulting in binding of the two entities. As with antibodies the weak molecular forces involve salt bridges and ionic interactions, hydrophobic interactions and van der Waals forces, and the overall binding strength is a summation of the number and nature of these physicochemical interactions, and the strength of each individual interaction.

Antibodies have been the cornerstone of the success of the Life Sciences and Biopharmaceutical industries. As antibodies are a crucial part of the immune system and are nature's preferred scaffold, the following paragraphs will incorporate descriptions of the immune system, antibody structure and antibody production. The similarity and differences between antibodies and other molecular scaffolds will be compared and contrasted, and the structures of various molecular scaffolds will be described.

The novel human chaperonin 10 (hCpn10) scaffold, the subject of the research described in this thesis, is introduced and proposed as a new molecular scaffold with application in the therapeutic and diagnostic areas. The hypothesis put forward in this thesis is that the mobile loop of hCpn10 can be substituted with non-hCpn10 peptide sequence, and that new molecular entities (NME) can be created with binding specificity to specific targets. It is also hypothesised that peptide substitution will not hinder the assembly of hCpn10 heptamer from monomer. The hypothesis is tested through the creation of several NMEs based on hCpn10 and demonstration of binding to the target. The NMEs created in this research have both therapeutic and diagnostic value, and the process of intellectual property protection is now underway.

1.1.1 The immune system, antibodies and antibody structure

The immune system, a network of cells, tissues and organs, is the body's defence against infections and diseases. It responds to pathogens and other foreign bodies by producing antibodies, which binds to various molecules with high affinity and specificity. Antibodies are a naturally occurring molecular scaffold that are produced and secreted by B-cells of the acquired immune system. The specificity of an antibody towards a foreign molecule is made possible through immunoglobulin gene rearrangement in hematopoietic stem cells to give rise to the B-cell lineage. Subsequently, antibodies with the highest affinity towards an antigen are naturally selected by the immune system for clonal expansion and propagation in larger numbers (Janeway et al., 2001). Antibodies are either monoclonal or polyclonal depending on the recognition of epitope per antigen. Monoclonal antibodies (mAbs) are derived from a single B-cell clone with specificity towards a single antigenic determinant (Birch and Onakunle, 2005), while polyclonal antibodies are derived from multiple clones of B-cells that produce antibodies binding to different epitopes on single antigen (Burry, 2010).

Antibodies have complex structures composed of two heavy and two light chains linked by disulphide bonds. This in turn gives rise to a quaternary structure that closely resembles a Y shape (Figure 1.1). The structure of the antibody can be further subdivided into two major regions, namely, the Fc (Fragment, crystallisable) region and the Fab (Fragment, antigen-binding) region. The Fc region, which forms part of the heavy chain, has an effector function whereby the complement system and/or antibody dependent cytotoxicity can be initiated. The Fab region or the antigen binding site, which has both heavy and light chains, is responsible

for interacting with the foreign antigen. The antigen binding site is composed of six complementarity determining regions (CDRs), with three CDRs in both the heavy and light chains. Interaction with the antigen is dependent on all six CDRs working in synergy. The CDRs within both chains are interconnected by eight β-strands that serve as a scaffold for the CDR regions. Of all sequence within the variable domain, the CDRs have the highest sequence diversity which is developed through cycles of somatic mutation and clonal expansion (Marks, 2009, Skerra, 2000a, Tramontano and Lesk, 1992). In comparison to CDR1 and CDR2, CDR3 has the longest average peptide length, the largest contact surface and is frequently the greatest contributor towards antigen binding (Rees et al., 1994).

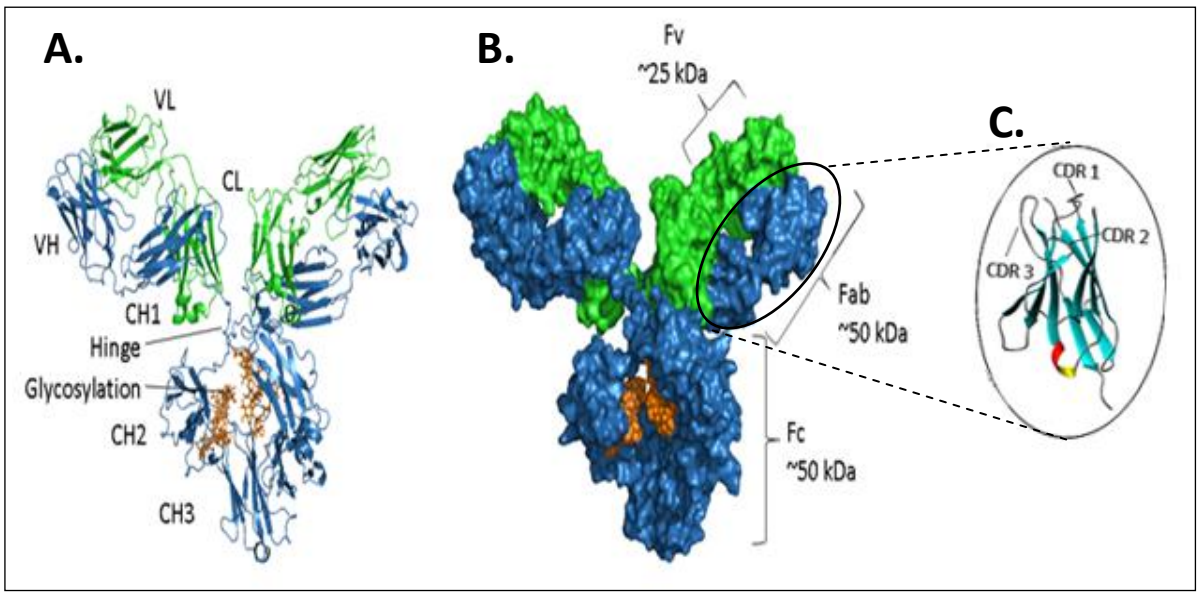


Figure 1.1 Structural and schematic representations of an antibody

Structure of a mouse IgG1 (PDB ID 1IGY). The heavy chain is shown in blue, light chain in green and glycosylation in orange in both representations of the antibody. **(A)** A ribbon representation of the secondary structure of the antibody. **(B)** A space-filled model representation of the same molecule. **(C)** An enlargement of the variable domain shown in ribbon representation of the β-sheet framework and antigen binding CDR1, 2 and 3 loops. Figure adapted from AbsoluteAntibody (2014).

1.1.2 Monoclonal antibodies (mAbs) and the biopharmaceutical industry

mAbs are the largest class of biologics, and are able to treat a wide variety of diseases indications including chronic conditions such as cancer and inflammatory disease, and also infectious disease. mAbs may be produced through conventional hybridoma technology, or through recombinant DNA methodologies that include antibody engineering, phage display

and transgenic mouse technologies. A number of mAbs have already been approved for cancer therapy and for the treatment of some autoimmune and infectious diseases with varying degrees of success. Examples of these mAbs are Herceptin® (trastuzumab) for breast and lung cancer treatment (Nakamura et al., 2005) and Avastin® (Bevacizumab) for the inhibition of angiogenesis in non-small cell lung cancer (Cohen et al., 2007). In some cases due to their large size and complex molecular structure, mAbs, have some limitations as therapeutic agents. Furthermore, the cost of discovery to actual usage of novel mAbs is quite high, exceeding $US 1.2 billion (Nelson and Reichert, 2009, Steinmeyer and McCormick, 2008). This is because mAbs are manufactured by biopharmaceutical companies using large scale mammalian cell culture aiming for a high gross margins of between 80-85 %, in order to recoup the high cost of mAb development of $US 300-500 million per annum over a significant period of time (7-10 years). Therapeutic mAbs cost around $US 2,000-10,000 per treatment with an annual cost of $US 20,000-100,000 per patient being treated for serious illness (Hassett and Elkin, 2013, Samaranayake et al., 2009).

In addition to having a structure of two sets of heavy and light chains stabilised through disulphide bonds, complete, full length mAbs are glycosylated, which is critical for its functionality and therefore, can only be expressed and purified from mammalian systems such as Chinese hamster ovary (CHO) or Human Embryonic Kidney (HEK) cells, or in yeasts such as *Pichia pastoris*. As previously mentioned, large scale manufacturing of mAbs using eukaryotic expression systems incurs high costs and is labour intensive (Steinmeyer and McCormick, 2008). In addition, the therapeutic efficacy of mAbs can be limited in some cases due to their relatively large size, (≥ 150 kDa) since full length mAbs cannot penetrate deep into the tissues or into solid tumours. Moreover, antibodies cannot target critical intracellular molecules that are implicated in uninhibited cancer cell proliferation (Beck et al., 2008). Regarding thermodynamic stability, mAbs are sensitive to higher temperatures and require optimum storage conditions whereby temperatures below 4 °C are ideal in order to retain therapeutic efficacy (Gebauer and Skerra, 2009). Importantly, acquiring intellectual property (IP) protection for new mAbs is complicated by the relatively large number of patents that have already been granted (Beck et al., 2008, Kontermann, 2010). Therefore, it is important to look beyond conventional antibodies for the development of new molecular entities (NMEs) with target binding capabilities that are not only scalable for production of kilogram quantities, but also provide opportunities for a robust IP portfolio.

1.1.3 Non-antibody molecular scaffolds

For the past 15-20 years, researchers have been developing alternative binding entities using different antibody and non-antibody scaffolds. Antibody-derived scaffolds are beyond the scope of this study and have been reviewed elsewhere (Holliger and Hudson, 2005, Nelson and Reichert, 2009, DS, 2012). Through advancements in recombinant DNA technology, next generation biologics have been derived from endogenous protein scaffolds and have been modified in such a manner that they possess the antigen binding properties of conventional mAbs but with the added advantage in some cases of lower cost of production and thermal stability.

Non-antibody scaffolds are defined by Binz et al. (2005), as "a framework that can carry altered amino acid or sequence insertions that confer on protein variants different functions, usually for binding specific targets". The ideal molecular scaffold must retain the properties of conventional mAbs, i.e. target specificity and affinity, and secondly, have a higher degree of structural stability with a lower molecular weight (<100 kDa). This is particularly important in the field of oncology for example, since an enhancement of the tissue penetrating properties of the molecular scaffold ensures a higher therapeutic efficacy against solid tumours. From a commercial perspective, non-antibody scaffolds have the potential to generate new IP that will not be hindered by the current complex landscape in the development of conventional mAbs.

More than 50 non-antibody derived proteins have already been identified as alternative scaffolds for various basic and applied research applications in the field of biopharmaceutical development (Wurch et al., 2012, Lofblom et al., 2011, Gebauer and Skerra, 2009, Nuttall and Walsh, 2008, Gill and Damle, 2006). The antibody's genetic engineering platforms have facilitated further design and development of such molecular scaffolds and there are currently many more platforms used for the research and development of such alternative-binding molecules.

Two most commonly used protein engineering platforms are the combinatorial and rational engineering approaches, summarised in Table 1.1. Utilising either of these two platforms, naturally occurring proteins can be genetically engineered to specifically recognize and bind to a potential target though modification of the binding site (Kontermann, 2010, Wurch et al., 2012, Gebauer and Skerra, 2009). Table 1.1 shows that for each engineering platform, there are various possible locations of engineered binding-sites, associated screening/selection methodologies and selected examples for each scaffold platform technology.

Table 1.1 Non-antibody scaffold platform

	Engineering Platform	Engineering Site(s)	Selection Method(s)	Scaffold Example(s)	Reference(s)
I.	Combinatorial Engineering Approaches:				
	Flat surface randomisation	**α-helices**	Phage display.	Affibody (Protein Z)	(Nord et al., 1997)
	Loop randomisation	**Loop(s)**	Phage display; mRNA display; & Yeast two-hybrid display.	^{10}FN3 (Fibronectin)	(Koide et al., 1998, Xu et al., 2002, Hackel et al., 2008)
	Cavity randomisation	**β-barrel**	Phage display, & Ribosome display.	Anticalin (Lipocalin)	(Beste et al., 1999, Liu et al., 2012)
	Multiple-sites randomisation	**β-hairpin & α-helices**	Ribosome display.	DARPins (Ankyrin repeat)	(Binz et al., 2004)
	Random peptide insertion	**Random sites**	Flagellin fusion; & Yeast two-hybrid display.	Aptamers (Thioredoxin)	(Lu et al., 1995, Colas et al., 1996)
	Error-prone PCR	**Undefined position(s)**	Yeast two-hybrid.	PDZ domain	(Schneider et al., 1999)
II.	Rational Engineering Approaches:				
	Grafting	**Loop**	Traditional ELISA assays, SPR measurements , & etc.	Neocarzinostatin (NCS)	(Nicaise et al., 2004)
		Loop	Computational modelling.	**CE76$_{FVIIa}$** **CP7$_{CD44}$**	**Unpublished**/ Chapters 4 & 5
	"*De novo*" Rational design	***De novo* binding-site**	Automated design procedures.	Ribose-binding protein (RBP)	(Looger et al., 2003)
		***De novo* α-helices**	Computational strategies	(Fe_4S_4)-Cluster protein	(Grzyb et al., 2010)
		***De novo* α-helices & β-barrel**	*De Novo* protein Computational strategy	Top7	(Kuhlman et al., 2003)

1.2 Alternative molecular scaffolds

While mAbs have been spectacularly successful in the clinic, there is much global research activity at present associated with the discovery of NMEs based on conferring functionality within a protein scaffold, whereby binding domains can be engineered within loop structures or flat surfaces, for example, without interrupting the folding or structural integrity of the protein scaffold. These surface exposed residues can be utilised as binding entities towards various antigens and possess different secondary structures such as α-helical bundles, β-strands configurations or spatial loops. These sites can be modified by a variety of protein engineering strategies without significantly affecting the core protein framework (Table 1.1). There are a number of molecular scaffolds in various stages of clinical development and are described in Tables 1.2 to 1.4. These scaffolds are not only diverse in size, topology and modes of interaction but they also demonstrate the properties critical to designing proteins with synthetic binding-sites to be used in a variety of applications (Figure 1.2). Non-antibody based scaffolds for use in therapeutics and diagnostics will serve as an alternative to naturally occurring antibodies and potentially overcome some of the structural limitations of mAbs.

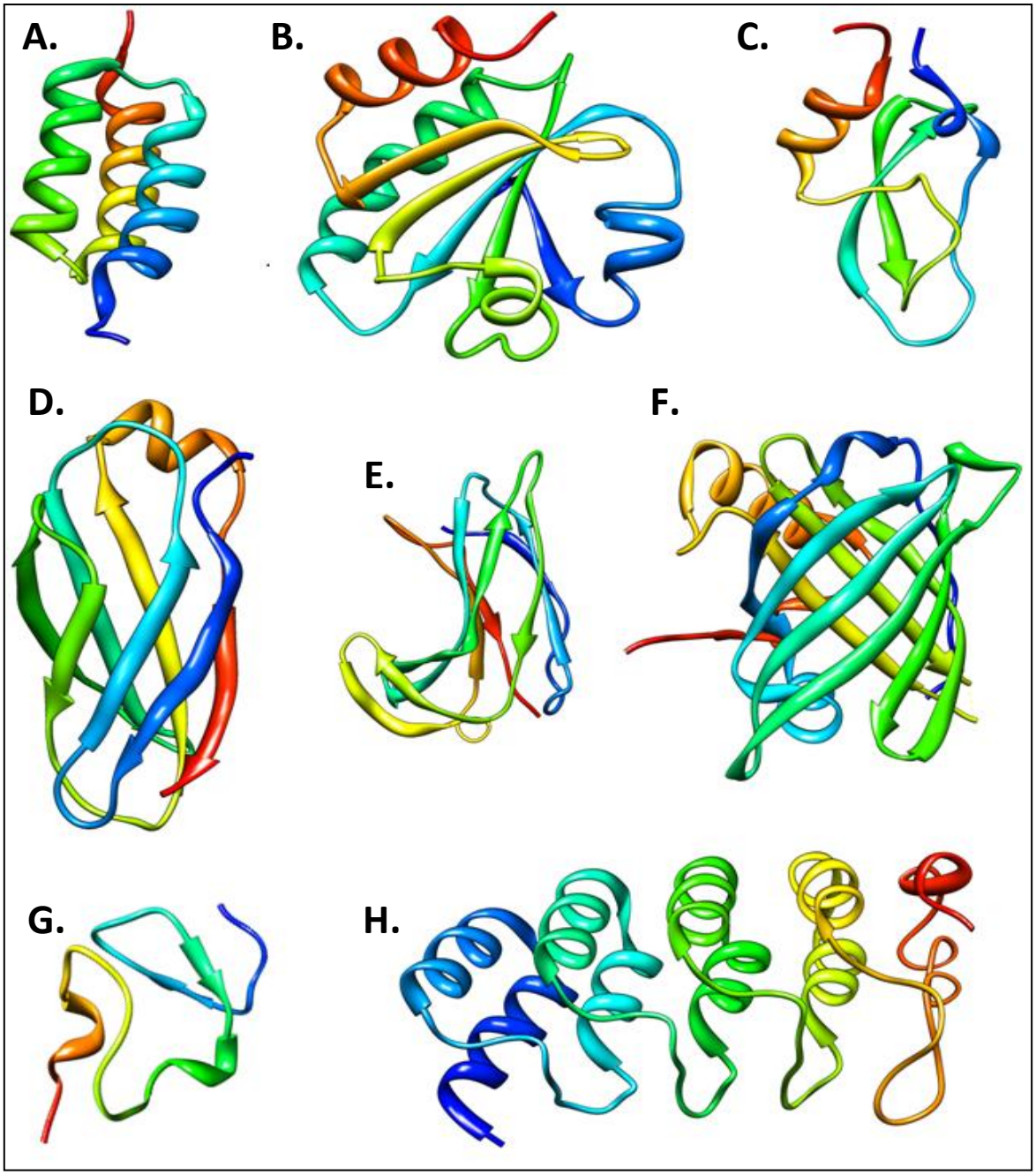


Figure 1.2 Non-antibody molecular scaffolds

Schematic representations of protein backbones used as scaffolds for generating protein-binding agents: **(A)** α-helix bundle, **Affibody** (PDB ID. 2B89); **(B)** α-helix bundle & β-barrel, **Aptamer** (PDB ID. 1TXX); **(C)** Loop, **Kuntiz** (PDB ID. 4DTG); **(D)** β-sandwich & loop, **Fibronectin** (FN3) (PDB ID. 1TTG); **(E)** β-barrel bundle & loop, **Neocarzinostatin** (PDB ID. 2G0K); **(F)** β-barrel & loop, **Anticalin** (PDB ID. 1LNM) of lipocalin; **(G)** bundle & loop, **Avimer** (PDB ID. 1LDL); and **(H)** α-helix bundle & β-turn repeat unit, **DARPin** (PDB ID. 2BKK) of ankyrin repeat protein. These are examples of molecular based-scaffold; however, they differ from each other in terms of size, topology and modes of interaction with target protein/antigen. All molecular structures are shown in ribbon representation and were modelled using UCSF Chimera1.5.3.

The general design principle in engineering protein scaffolds is to introduce residue(s)/ peptide sequence(s) with a specific function(s) into a specific region of the protein scaffold without compromising protein folding and overall stability (Skerra, 2007a, Skerra, 2000a, Mathonet and Fastrez, 2004). It has been observed that many molecular entities exist whereby for example, certain endogenous loop structures can be replaced with non-native peptide sequence. This provides an opportunity to develop NMEs using these non-antibody based scaffolds. Ideally, the non-antibody molecular scaffold should be small, robust and economical when scaled in prokaryotic expression systems.

Despite the favourable outcomes expected in utilising non-antibody scaffolds for therapeutic or diagnostic purposes, only a few such biological entities have either reached clinical trials (Table 1.2). As an example, Ecallantide (a DX-88 derived from the kuntiz domain), which is a kallikrein inhibitor, is the first non-antibody, scaffold-based biologic to be approved by the US Food and Drug Administration (FDA) in 2009 for the treatment of acute hereditary angioedema (HAE) and the prevention of blood loss in cardiothoracic surgery (Bernstein and Moellman, 2012). The success of Ecallantide demonstrates that non-antibody scaffolds can be used in the treatment of various diseases. Others are being investigated for their application as *in vivo/in vitro* diagnostic agents for various diseases (Table 1.3). A number of molecular scaffolds have been extensively studied for their use in a wide range of biotechnological applications (Table 1.4), including affibodies, fibronectin, lipocalin, ankyrin, kunitz domain and others. In the following sections, each of these non-antibody scaffolds isare described in detail.

1.2.1 Protein A (Affibody)

Affibodies are derived from the *Staphylococcus aureus* protein A which is known to bind to the Fc domain of immunoglobulins (Nord et al., 1997). Affibodies consist of three α-helices with no disulfide linkage and have the capacity to display reversible folding (Figure 1.2). Affibodies have been developed through combinatorial protein design and selected via phage display for a range of therapeutic applications such as tumour targeting (Friedman et al., 2008). Affibody Z_{HER}, for example, is a high affinity affibody that binds to the cell surface of human epidermal growth factor receptor 2 (HER2) and is now undergoing Phase I clinical trial for the treatment of breast cancer (Wikman et al., 2004) (Table 1.2). Affibodies have also shown promise as *in vitro* diagnostic reagents and for tumour imaging for detecting

positive HER2 expression (Nilsson and Tolmachev, 2007) (Table 1.3). As summarised in Table 1.4, Affibodies have also been utilised in bio-purifications (Nord et al., 2000) and as analytical reagents (Karlstrom and Nygren, 2001).

1.2.2 Fibronectin (Monobody)

Fibronectin type III (FN3), derived from the tenth FN type III domain in human fibronectin (^{10}FN3) (Koide and Koide, 2007, Koide et al., 1998) has an antibody-like structure that displays CDR-like loops (Figure 1.2). In contrast to antibodies, FN3 does not rely on disulfide cross-linking for folding, and thus has a higher thermodynamic stability (Koide et al., 1998). The use of ^{10}FN3, alternatively known as a Monobody or Adnectine, was first proposed as a scaffold as it has the combined features of having a small size and high structural stability, all of which are favourable properties for an NME intended for therapeutic use (Bloom and Calabro, 2009). A library of diverse Monobodies have been developed through combinatorial engineering approaches, whereby individual binding clones can be isolated through various selection strategies such as phage, mRNA, or yeast displays approaches (Table 1.1). Overall, due to its structural stability combined with the absence of disulfides, fibronectin is speculated to have similar binding capabilities as that of conventional antibodies, and can therefore be applied to cellular targeting (Table 1.2), (Getmanova et al., 2006, Karatan et al., 2004, Koide et al., 2002). Monobodies have been successfully applied to the treatment of cancer.; Ffor example, a Monobody reactive to vascular endothelial growth factor receptor 2 (VEGFR2), has been used to inhibit angiogenesis in tumours (Table 1.2), (Getmanova et al., 2006).

1.2.3 Lipocalin (Anticalin)

The Anticalin scaffold is derived from the lipocalin family proteins of various species such as humans and insects (Skerra, 2000b, Skerra, 2001, Beste et al., 1999). The term Anticalin is derived from the combination of antibody and lipocalin (Beste et al., 1999). Anticalins consist of a β-barrel core that is stabilized by disulfide bridges with an interconnection of two to three mobile loops (Figure 1.2) (Beste et al., 1999). Lipocalins typically bind to small molecules, such as water-insoluble vitamin A, pheromones and odorants for example, (Schlehuber et al., 2000), through their β-barrel surface or through the mobile loop regions. Anticalins, on the other hand, are designed to bind with high affinity and specificity to various antigens through loop-substituted peptides (Table 1.1). Anticalin libraries have been

developed where peptides with wide sequence diversity were introduced within the mobile loops of the β-strands and potential binders were selected via phage or ribosomal display technologies (Table 1.1), (Schlehuber and Skerra, 2002). As shown in Table 1.2, Anticalins have been developed for a number of uses such as an antidote for overdose of digoxigenin (Schlehuber et al., 2000), an immunomodulator by targeting cytotoxic T-lymphocyte antigen-4 (CTLA-4) and thus, blocking cells signalling (Schlehuber and Skerra, 2005) and as an antagonist for vascular endothelial growth factor (VEGF) for cancer treatment (Skerra, 2008). Anticalins, have also been proposed as *in vivo/in vitro* diagnostic agents (Table 1.3), as well as being utilised as reagents in various biomedical research field settings (Table 1.4).

1.2.4 Designed Ankyrin Repeat Protein (DARPin)

DARPins are another class of molecular scaffolds created by fusing the ankyrin repeat sequence of multiple α-helices and β-turns (Figure 1.2), (Forrer et al., 2003). DARPins have the unique capability of forming large surface interaction interfaces by adjusting the number of the ankyrin repeats (Table 1.1). DARPins have a high thermodynamic stability and reversible folding capacity, thus they have been successfully used to develop a library of high-affinity binders against a range of targets (Table 1.2), (Stumpp et al., 2008, Binz et al., 2004, Mosavi et al., 2004). DARPins can be expressed at high levels as functional and soluble proteins using bacterial expression systems (Stumpp et al., 2008). Conjugated-DARPins have also been designed to bind to intracellular and extracellular targets (Stumpp et al., 2008). Therapeutic applications of high affinity DARPins include targeting maltose binding protein (MBP) (Binz et al., 2004), caspase-2 (Schweizer et al., 2007), intracellular proteinase (Kawe et al., 2006), human epidermal growth factor 2 (EGF2) and HER2 (Zahnd et al., 2007) (Table 1.2). Also, anti-idiotypic IgE-DARPins have been developed as vaccine candidates (Table 1.2), (Vogel et al., 2007). Moreover, DARPins have been utilised for *in vivo/in vitro* detection of HER2 expression in breast cancer (Table 1.3). The unique structural properties of DARPins have been utilised in their co-crystallisation with unstable proteins and subsequent determination of protein structure by X-ray crystallography (Table 1.4).

1.2.5 Kuntiz Domain

The Kuntiz Domains are another class of non-antibody scaffold based on human serine protease inhibitors that have been engineered as an alternative scaffold for a variety of important protease targets (Figure 1.2), (Dennis and Lazarus, 1994b, Dennis and Lazarus, 1994a, Nixon and Wood, 2006). An LACI-D1 is a Kuntiz Domain-derivative based on domain-1 of human lipoprotein-associated coagulation inhibitor, and is the first non-antibody scaffold derivative approved by FDA for medical applications in the USA; LACI-D1 (or DX-88/ Escallantide) is approved for treatment of acute HAE as a potent inhibitor of the plasma kallikrein, and also used for prevention of blood loss in cardiothoracic surgery (Table 1.2), (Lehmann, 2008, Lunn and Banta, 2011). DX-1000 is another example of a Kunitz Domain derivative based on tissue factor pathway inhibitor (TFPI). This biologic drug is currently being studied in animal models for blocking the growth of breast cancer and preventing metastasis by inhibiting plasminogen activation (Table 1.2).

1.2.6 Avimers

Avimer molecules (named from Avidity and multimer) are another example of a protein scaffold. Avimers are derived from the multivalent low density lipoprotein receptor class A (LDLR-A), (Figure 1.2). An avimer is capable of neutralising interleukin-6 (IL-6) and has been developed for the treatment of Crohn's disease (Table 1.2), (Silverman et al., 2005). Recently, the anti-IL6 Avimer was trialled (Phase I) for safety and toxicity studies (Table 1.2).

1.2.7 Neocarzinostatin (NCS)

Neocarzinostatin (NCS) is a small protein and member of endiyne-chromoprotein family with a well-defined hydrophobic β-barrel core connected with two loops structurally equivalent to CDR1 & CDR3 of immunoglobulin (Figure 1.2), (Nicaise et al., 2004). NCS was developed as a molecular scaffold by grafting an immunoglobulin-CDR3 sequence to one of the NCS loops (Table 1.1). The scaffold NCS-CDR3 specifically bound with high affinity to hen egg white lysozyme and has been exploited for future drug and biologic development (Nicaise et al., 2004).

Table 1.2 Non-antibody molecular scaffolds utilised for therapeutic applications

Scaffold Name	Scaffold Format	Target	Application(s)/ description(s)	Clinical Trial Phase	Reference(s)
$Z_{HER2:342}$	Affibody	HER2	Radioimmunotherapy (RIT); 177Lu-labeled HER2-Specific Affibody® For treatment of disseminated tumours with a high level of HER2 expression on breast, ovary, and urinary bladder.	1	(Tolmachev et al., 2007)
Z_{CD25}	Affibody	IL-2Rα/CD25	Affibody® binding to interlekin-2 (IL-2) receptor α subunit (known as CD25) used for prevention of organ rejection or autoimmune disease. Also, as therapeutic depletion of $CD25^+$ cells in T-cell leukemia.	--	(Gronwall et al., 2008)
Zgp120	Affibody	HIV-I gp120	Treatment for HIV-I through virus neutralization, inhibiting attachment or entry.	--	(Wikman et al., 2006)
Z_{AB}	Affibody	β-Amyloid	Slow down plaque formation for Alzheimer patients, by bind to amyloid-beta (Aβ) peptide and reduce amyloid plaque formation.	--	(Gronwall et al., 2007)
Monobody	^{10}FN3	VEGFR-2	Anti-angiogenesis For treatment of non-Hodgkin's lymphoma. Antagonists of human VEGFR2.	1	(Getmanova et al., 2006).

Scaffold Name	Scaffold Format	Target	Application(s)/ description(s)	Clinical Trial Phase	Reference(s)
Adnectin™ CT-322	^{10}FN3	VEGFR-2	Glioblastoma, Non-small-cell lung carcinoma (NSCLC) VEGFR-2 inhibitor Treatment of glioblastoma multiforme, colorectal cancer, pancreatic cancer, and non-small cell lung cancer.	2	(Dineen et al., 2008, Mamluk et al., 2010)
Anticalin (FluA)	Lipocalin	Fluorescein	Antidote Engineered lipocalin, from bilin-binding protein (BBP), against small molecules and target proteins coupled with fluorescein as a nonradioactive label	--	(Beste et al., 1999).
Anticalin (DigA)	Lipocalin	Digoxigenin	Antidote Engineered lipocalin, from BBP, for binding to Digoxigenin group, for overdoses or poisons.	--	(Schlehuber et al., 2000)
Anticalin (Lcn2)	Lipocalin		Radioimmunotherapy (RIT), Metal-chelator Used in RIT, and used as radionuclide-chelate capturing (e.g. $Y3^{+}$-DTPA), preventing the release of free metal ion and preventing kidney toxicity.	--	(Kim et al., 2009)

Scaffold Name	Scaffold Format	Target	Application(s)/ description(s)	Clinical Trial Phase	Reference(s)
Anticalin (PRS-050)	Lipocalin	VEGF	Angiogenesis inhibitor, oncology A potential treatment for certain solid tumors, act as inhibitor of tumor angiogenesis, by neutralizes human VEGF.	1	(Skerra, 2007b)
Anticalin (PRS-010)	Lipocalin	CTLA-4	Cancer immunotherapy, blocking CTLA-4, a receptor with attenuating function expressed on activated T cells, and stimulating the cellular immune response.	Pre-clinical	(Schonfeld et al., 2009)
TF7I-C	Kuntiz domain	TF-FVIIa	Anticoagulant Derived from β-amyloid protein precursor inhibitor (APPI), used as potent inhibitors of the active site of the human TF.FVIIa complex.	--	(Dennis and Lazarus, 1994b).
Escallantide (DX-88)	Kuntiz domain	Plasma kallikrein	Kallikrein inhibitor Approved for treating acute HAE as potent inhibitor of the plasma kallikrein. Also, for prevention of blood loss in cardiothoracic surgery. It is derived from human lipoprotein-associated coagulation inhibitor (LACI).	FDA approved (2009)	(Lehmann, 2008, Lunn and Banta, 2011)

Scaffold Name	Scaffold Format	Target	Application(s)/ description(s)	Clinical Trial Phase	Reference(s)
DX-1000	Kuntiz domain	Plasmin (pln)	Anticancer and plasmin inhibitor Blocking breast cancer growth and metastasis by inhibiting plasminogen activation. It is derived from tissue factor pathway inhibitor (TFPI).	Pre-clinical	(Devy et al., 2007)
EPIhNE4 (DX-890)	Kuntiz domain	Neutrophil elastase (HNE).	Proteolysis-resistant inhibitor Anti-inflammatory drug for cystic fibrosis. It is derived from the light chain (bikunin) of the human protease inhibitor, inter-α-trypsin inhibitor.	--	(Attucci et al., 2006)
CD4-DARPin	DARPins	CD4	Microbicide Blocking mucosal HIV transmission, by specifically bind to the cellular CD4 receptor, the main entry receptor of HIV.	--	(Schweizer et al., 2008)
DARPin	DARPins	IgE & IgE receptor	Anti-idiotypic antibody, regulators of immune responses, inhibits mast cell degeneration by blocking IgE and IgE receptor. Potential vaccine, used to induce an anti-IgE response preventing the binding of IgE to FcεRI and the degranulation of effector cells implicating in allergic disease.	--	(Vogel et al., 2007)

Scaffold Name	Scaffold Format	Target	Application(s)/ description(s)	Clinical Trial Phase	Reference(s)
DARPin G3	DARPins	HER2	Anti-HER2, for HER2-overexpressing cancers e.g. breast and ovarian cancers.	Pre-clinical	(Zahnd et al., 2010)
MP0112	DARPins	VEGF-A, VEGFR-1	Ophthalmologic, AMD, DME Ophthalmological diseases, against VEGR cellular receptors, as a possible treatment for various ophthalmologic disease indications	2	(Campochiaro et al., 2013)
C326 (AMG-220) C426 C2810 C65 C2	Avimer	Interleukin-6 cMet, CD28, CD40L, & BAFF	Anti-IL6 Neutralizes IL-6 (a pro-inflammatory cytokine) for treating Crohn's disease. Different designed Avimers bind to and antagonise for: cMet, CD28, CD40L and BAFF.	1 Completed --	(Tawara et al., 2011, Silverman et al., 2005)
$CE76_{FVIIa}$	Cpn10	FVII/FVIIa	Anticoagulant, act as Factor X activity inhibitor by blocking the extrinsic-dependent TF-FVIIa complex. Designed based on human Cpn10.	--	**Unpublished**/ Chapter 4

Table 1.3 Non-antibody molecular scaffolds utilised for diagnostic applications

Scaffold Name	For	Application(s)	Description(s)	Reference(s)
Affibodies				
Z_{HER2}	*In vivo*	Tumor imaging/ localization	HER2 expression in breast and ovarian cancer. In phase (I) clinical trial for breast cancer imaging.	(Orlova et al., 2006, Tolmachev et al., 2007, Wikman et al., 2004)
Z_{EGFR}	*In vivo*	Tumor imaging/ localization	Imaging of expression of EGFR-1 carcinoma.	(Friedman et al., 2008)
$Z_{TNF\text{-}\alpha:185}$	*In vivo*	Inflammatory imaging	Binding to human tumor necrosis factor α (TNF-α), aids in the diagnosis and staging of patients.	(Jonsson et al., 2009)
Z_{CD25}	*In vitro*	Medical imaging	Visualizing of over expression of IL-2α/CD25 in organ rejections, autoimmune diseases and T-cell malignancies.	(Gronwall et al., 2008)
Z_{Taq}	*In vitro*	Protein microarrays	*Taq* DNA polymerase and human IgA.	(Renberg et al., 2007)
Anticalins				
Lcn2	*In vivo*	Imaging	Radioimmunodiagnosis (RID)	(Kim et al., 2009)
DigA16	*In vivo/ in vitro*	Detector reagent	For studying digoxigenin-binding function and enzymatic activity.	(Schlehuber et al., 2000)
CTLA-4 specific	*In vitro*	Immunocytochemicals staining	Immunological reagent for cytotoxic T lymphocyte-associated antigen (CTLA)-4.	(Schlehuber and Skerra, 2005)
DARPins	*In vivo*	Tumor imaging/ localization	HER2 imaging of breast cancer patients	(Stumpp et al., 2008)
	In vitro/ In vivo	Tumor imaging/ stage detections	HER2 expression state of breast and colon carcinoma	(Zahnd et al., 2007)
PDZ domain	*In vivo/ in vitro*	Discovery reagent	For peptide C-terminus recognitions of drugs & for studying protease activities.	(Ferrer et al., 2005)
	In vitro	Detector	For time-resolved fluorescence resonance energy transfer (TR-RET) method.	(Ferrer et al., 2002)
$CP7_{CD44}$	*In vitro/ in vivo*	Tumor imaging/ localization	Targeting CD44 overexpression in breast, colon, & prostate cancer. Designed based on human Cpn10.	**Unpublished**/ Chapter 5

Table 1.4 Non-antibody molecular scaffolds utilised for bio-research applications

Scaffold	Application(s)	Relevance(s)	Reference(s)
Affibodies	Microarray detection	Protein capturing in microarray, of *Taq* DNA polymerase, human IgA, IgE, IgG, TNFα, and insulin.	(Renberg et al., 2007, Nord et al., 2001, Ronnmark et al., 2002, Gronwall et al., 2007)
	Biopurification & Bioseparation	Affinity chromatography capturing recombinant human Factor VIII from CHO superannuant, apolioprtoein A from fermentation media, *Taq* DNA polymerase or recombinant human IgA from bacterial lysate, and human β-amyloid peptide from human body fluid.	
FN3	Biopurification	Purify the active estrogens receptor (ER) from lysate mixture containing misfolded receptors.	(Huang et al., 2006, Karatan et al., 2004)
	Affinity reagent	Used in immunoprecipitation experiments for c-Src from murine fibroblast cell extracts.	
	Research diagnostic	Valuable tool for studying intercellular protein functions, due to expression inside cells.	
Anticalins (DigA/ FluA)	Bioanalytics	Fusion with alkaline phosphatase (reporter enzyme).	(Schlehuber et al., 2000, Beste et al., 1999)
	Biochemical reagent	For the detection of biomolecules coupled with a nonradioactive label e.g. fluorescein	
DARPins	Protein crystallography	Facilitate crystallisation studies of; trimeric AcrB (a multidrug exporter in *E. coli*); & human polo-like kinase 1 (Plk-1) proteins, a well validated drug target in cancer therapy; the *E. coli* maltose binding protein (MBP); and two eukaryotic mitogen-activated protein kinases (MAPKs).	(Sennhauser et al., 2007, Bandeiras et al., 2008, Binz et al., 2004, Schweizer et al., 2007)
(AR)	Biopurification	Facilitate purification of MBP and MAPKs.	
(AR-F8)	Discovery tool	Specific caspase inhibitors useful for the validation of pathways, for the discovery of unknown interactions, and for the identification of novel intracellular targets.	
PDZ domain	Peptide detector	For studying protease activities	(Ferrer et al., 2005)

1.3 Chaperone proteins

Molecular chaperones are proteins which facilitate the correct folding, assembly, transport and degradation of other proteins *in vivo* (Macario and de Macario, 2007, Ranson et al., 1998, Hartl and Hayer-Hartl, 2002). Chaperonin 10 (Cpn10), heat shock protein 10 (Hsp10), or GroES in *E. coli* is a member of the chaperone family. Cpn10 is an essential mitochondrial protein that binds with its cognate partner chaperonin 60 (Cpn60, Hsp60 or GroEL), and in the presence of Mg^{2+}/ATP, the complex forms a "barrel" type structure, in which unfolded or degraded proteins are refolded (Hartman et al., 1992, Hartl and Hayer-Hartl, 2002).

It has been generally accepted that proteins have fixed cellular localisations. However, recent studies have shown that proteins can serve a variety of functions, depending on which organelle these proteins are localised. The identification of a protein outside of its known functional zone in cellular preparations was once thought to be due to rupturing of cells and/or cell organelles. It is now known that proteins translocate between intracellular and extracellular compartments. For example, chaperone proteins such as Cpn10, Hsp70 and Hsp90, which are known to conventionally localise in various organelles for chaperone activity, have now been shown to exist in extracellular locations (Butler and Overall, 2009).

1.3.1 Human Chaperonin 10

Human Cpn10 (hCpn10) is located within the mitochondrial membrane and is functionally associated with the correct folding of proteins that enter the mitochondria from the cytosol (Hartl and Hayer-Hartl, 2002). hCpn10 expression is up-regulated in several diseases that are associated with protein-misfolding (Slavotinek and Biesecker, 2001, Ranford et al., 2000). In addition to its traditional chaperone activity, hCpn10 has been found to have anti-inflammatory properties. The initial identification of hCpn10 in the serum of pregnant women led to the hypothesis that hCpn10 possessed immunosuppressive properties (Cavanagh, 1996). More recently, recombinant hCpn10 (XToll) has been tested in clinical trials (completed Phase II) for the treatment of various diseases associated with inflammation including rheumatoid arthritis (RA), psoriasis (van Eden, 2008) and multiple sclerosis (MS) (Broadley et al., 2009).

1.3.2 Chaperonin 10 structure and features

The Cpn10 structure is highly conserved across all species despite the fact that hCpn10, for example, has only 30 % sequence homology with GroES (Taneja and Mande, 1999, Ranford et al., 2000, Perham and Wittung-Stafshede, 2007) (Chapter 3). As shown in Figure 1.3, the quaternary structure of hCpn10 is a heptameric fold (70 kDa) with seven identical monomers, where each monomer (10 kDa) consists of a rigid hydrophobic β-barrel core, a β-hairpin roof loop, and a mobile loop extending below the plane of the molecule (Higurashi et al., 2003, Lin et al., 2004, Guidry et al., 2000). hCpn10 monomers, like most of the members of the Cpn10 family, assemble into a heptameric structure through interactions between the β-barrel interfaces of each monomers, whereby the *C*-terminus of one monomer binds to the *N*-terminus of an adjacent monomer via strong hydrogen bonds (Sakane et al., 2007, Sakane et al., 2004). Heptameric hCpn10 subsequently forms a complex with Cpn60 for full chaperone functionality (Figure 1.3). The oligomeric Cpn10 protein is one of a few proteins that can fold reversibly *in vitro* (Perham and Wittung-Stafshede, 2007), which is an important aspect in molecular folding and protein assembly.

The flexible "mobile" loop (*Ala21-Gly38*) of hCpn10 consisting of 18 residues (*AAETVTKGGIMLPEKSQG),* is an unstructured region that is critical for binding to Cpn60 (Richardson et al., 2001, Lin et al., 2004) (Figure 1.3). The mobile loop within Cpn10 is highly flexible but its movements are restricted when Cpn10 is bound to Cpn60 (Higurashi et al., 2003). The structure of the hCpn10 *delta* mobile loop (ΔL-Cpn10) was solved to a resolution of 2 angstrom units (unpublished data). Although mutations within the loop can substantially reduce the binding affinity of Cpn10 towards Cpn60 (Lin et al., 2004), it does not affect the assembly of Cpn10 into a heptamer (Sakane et al., 2007) (Chapters 3 to 5).

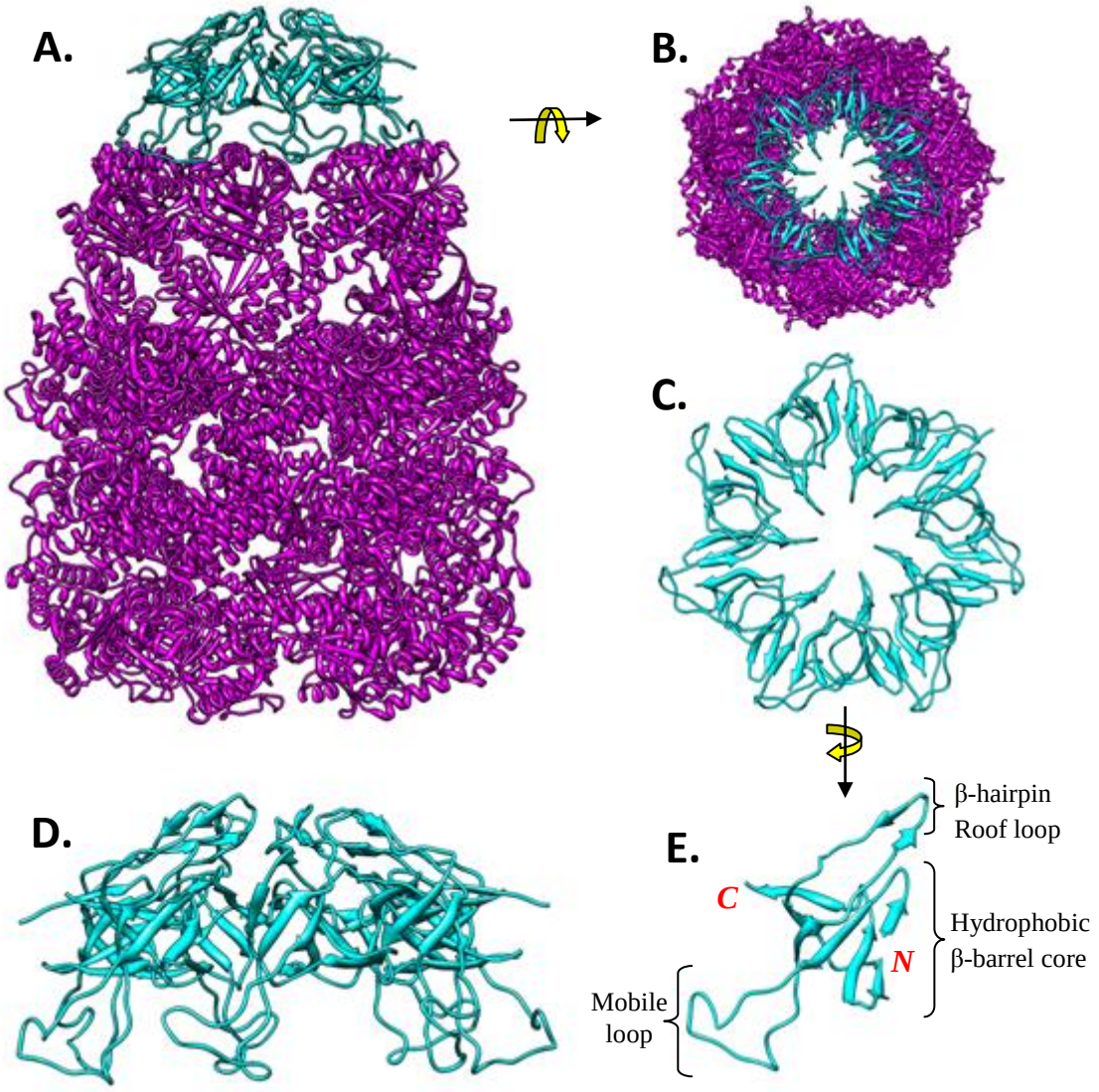


Figure 1.3 Three-dimensional model of human Chaperonin complex (Cpn10/Cpn60)

Schematic representation of protein backbones of human Chaperonin complex (Cpn10/Cpn60) taken from various angles. **(A)** Side view of chaperonin complex, showing Cpn10 (cyan) in complex with Cpn60 (magenta); **(B)** Top view of chaperonin complex, showing the dome-shaped Cpn10; **(C)** Top view of heptamer hCpn10 (70 kDa); **(D)** Side view of heptamer hCpn10; and **(E)** Side view of monomer hCpn10, shows the roof β-hairpins and the Cpn60-binding mobile loop extending outward from the lower edge of the hydrophobic β-barrel core. The hCpn10 models were generated by homology modelling using the chaperonin complex from *Thermus Thermophilus* (PDB ID. 1WE3) as a template (detailed in Chapter 3). All molecular structures are shown in ribbon representations and prepared using PyMOL software.

1.4 Scope of this thesis – hCpn10 as a novel molecular scaffold

Provided that introducing mutations within the mobile loop does not compromise the assembly of hCpn10 into a heptamer, the question arises as to whether hCpn10 can be utilised as an alternative molecular scaffold to display various peptides (Table 1.1). The tertiary structure of hCpn10 (monomer) is a rigid unit (anti-parallel β-barrel) consisting of a hydrophobic core (Figure 1.3). As with other molecular scaffolds, the loops form an extended interface that facilitates the binding of the molecular scaffold to the target molecule. In such loops, residues can be substituted or extended with a functional peptide domain that may vary in peptide sequence and length (Chapter 3).

In this study, hCpn10 is proposed as a novel molecular scaffold since it bears many ideal features of a scaffold wherein the mobile loop can be substituted with variant and non-native peptide sequences. As shown in Figure 1.3, the mobile loop lies below the plane of the ring of the heptameric structure and removal of this region does not compromise the assembly of hCpn10 monomers into the heptameric structure (Chapters 3 to 5). Thus one of the criteria for a molecular scaffold has been satisfied, whereby the mobile loop is an independent domain of hCpn10 which appears to not play a role in the assembly of the β-barrel monomer and the three dimensional conformation of heptameric hCpn10. In addition to the mobile loop, hCpn10 also possesses a roof loop at the apex of the molecule. The β-hairpin roof loop also presents a potential domain for amino acid randomization. The *N*- and *C*- termini also present opportunities for further fusion of protein domains.

The research described in this thesis aims to develop hCpn10 as a new molecular scaffold by demonstrating its utility through the creation of hCpn10 variants with specific functionality. This will be achieved by engineering protein sequences into the mobile loop region of native hCpn10, as similarly performed for the NCS scaffold (Section 1.2.7.). Importantly, a versatile molecular scaffold protein should constitute a stable conformation fold that is capable of displaying a multitude of loop structures or amino-acid sequences in a particular surface region (Skerra, 2000a). By definition, the features and properties of the hCpn10 molecule have shown that it is an ideal molecular scaffold candidate due to the rigid structure of the hydrophobic β-barrel core and the spatially separated mobile loop domain (Chapters 3 to 5). Furthermore, hCpn10 does not require post-translation modification (e.g. glycosylation) and can therefore be expressed as a soluble protein at high level using low cost prokaryotic

expression systems. Importantly, the quaternary structure of hCpn10 provides for higher binding avidity, with seven loops extending below the plane of the heptameric ring structure, which is an important aspect for the development of highly avid therapeutic agents (Chapter 4). Furthermore, the development of hCpn10 variants with new functionality offers an opportunity to create a platform for the development of a new class of NMEs based on hCpn10.

1.5 Potential applications for a hCpn10-based molecular scaffold

Several properties of hCpn10 meet the requirements of an ideal molecular scaffold, namely, small size (70 kDa), thermal stability and high expression level (30-40 % total cell protein) of soluble hCpn10 in *E. coli* production systems. Most importantly, removal of the mobile loop does not compromise assembly of the hCpn10 heptamer. Several potential applications for the hCpn10 molecular scaffold are summarised below:

- **Biologics:** NMEs that can bind to validated and new, potential disease targets.

- **Vaccines:** The hCpn10 scaffold could be loaded with major antigenic determinants of viral origin to elicit a broad polyclonal antibody response. For example, the display of DENV1, DENV2, DENV3 or DENV4 epitopes on scaffold-based hCpn10 for neutralizing antibodies and vaccine development against dengue virus (Mahler et al. Unpublished).

- **Multispecific Binding Entities:** Analogous to bispecific antibodies, peptides with specificity towards two (or more) targets could be expressed within the hCpn10 molecular scaffold. These peptides could be genetically engineered into the roof loop and mobile loop creating a bi (or tri etc)-specific-hCpn10-based NMEs. Such approaches could, for example, cross link receptors on cancer cells with cytotoxic T-cells.

- **Phage Display Systems:** Expression of the hCpn10 monomer in a phagemid vector, fused to the gene 3 could allow for the display of peptides within a hCpn10 monomer on filamentous phage library. By randomising the ~18 amino acids within the mobile loop, a library of diverse peptides could be screened for binders against specific targets

using a method known as “biopanning” in phage display technology (Mahler et al. unpublished data).

- ***In Vivo* Decoys:** Expressing multiple peptides (derived from receptors or soluble factors) on the hCpn10 scaffold may act as decoys *in vivo*, and have potential application in autoimmune diseases.

- **Recombinant Peptide Production:** hCpn10 can be expressed at high levels in low cost prokaryotic systems such as *E. coli* (~30-40 % of total cell protein). Under such circumstances, cleavage sites can be engineered at the *N*- and *C*-termini of the mobile loop whereby the recombinant protein can be extracted via protease digestion. This could provide a more cost effective method for peptide production for recombinant protein production.

1.6 Monovalent versus multivalent scaffolds

Binding affinity is one of the most crucial protein attributes for its successful engineering for therapeutic or diagnostic purposes. Non-antibody scaffolds, in some instances, have monovalent display of peptides, whereas antibodies are naturally bivalent and therefore, have a higher binding avidity towards a target antigen. In order to address this issue, various approaches have been tested to develop non-antibody scaffolds with multiple binding sites. For example, the Affibody (Z_{HER}) has been engineered to contain two binding sites and thus create avidity that is comparable to that of a typical antibody. Z_{HER} has been developed against HER2, which is overexpressed in breast cancer and studies have shown that the bivalent nature of Z_{HER} has contributed to a ten-fold decrease in the dissociation constant (K_D) and thus demonstrates a higher binding avidity relative to its monovalent counterpart (Steffen et al., 2005, Wikman et al., 2006).

Another approach is to create an NME, where a known scaffold is fused to a multidomain molecule. In this way, fusing $\alpha_V\beta_3$, which binds to FN3, to a pentamerisation domain, significantly increased the avidity towards the target and reduced the dissociation rate of integrin binding (Duan et al., 2007, Zhang et al., 2004).

Further strategies have been adopted to increase the avidity of non-antibody scaffolds. A multiple binding domain was developed by linking a number of monomeric avimers that bind to different IL-6 epitopes (Silverman et al., 2005). This approach was determined via protein modelling and studies have demonstrated that the avidity of this molecule had increased significantly. Non antibody-scaffolds have been developed principally in the context of their required applications; for example, Monobody (^{10}FN3) was used as an anti-angiogenic therapy, which only required a ratio of 1:1 monobody-antigen binding (Getmanova et al., 2006).

However, in the field of oncology, polyvalent biologic molecules are needed in order to effectively destroy cancer tissues. This is important because antibodies that typically bind to cancer cells can initiate the effector function and destroy the tumour cells (Vasudev and Reynolds, 2014, Carter, 2001). For this reason many approaches have been taken in order to create multivalent scaffolds, either by protein fusion (Zhang et al., 2004), conjugating certain molecules to antibodies or creating polyvalent antibody-fragments (Carter and Senter, 2008). Such strategies would superimpose and complement the molecular scaffold functions.

Unlike previous examples, because of the molecular structure nature of the heptameric hCpn10 neither fusions nor conjugations are required. The structural features of hCpn10 are capable of forming bispecific properties through simultaneously grafting functional peptides in both roof and mobile loops, as previously described. Thus, the scaffold based on hCpn10 has no limitation and is relevant to be used in any medical fields upon request

In addition, hCpn10 is an endogenous protein with natural multivalent properties and therefore, engineering additional bindings sites is not required (see section 1.5). It is practical to graft high affinity binding peptide sequences within the mobile or roof loop in hCpn10 and still retain its overall tertiary and quaternary structure (Chapter 3). This in turn will lead to the development of new biologics with a higher affinity towards a particular target.

1.7 Limitations of molecular scaffolds

A major disadvantage of many protein scaffolds, particularly those that are proposed as therapeutic or *in vivo* diagnostic agents, is their capacity to elicit an immune response. For example, Affibodies which are derived from the *S. aureus* protein A have the potential to

generate an immune response after being administered to patients multiple times. Nevertheless, Affibodies such as Z_{HER2} have been successfully used to image HER2 positive breast tumours in patients, as a single administration of Z_{HER2} for imaging purposes s would not stimulate a strong immune response.

Ideally a molecular scaffold should be of human origin to ensure a lower probability of eliciting an immune response (Li and Zhu, 2010, Samaranayake et al., 2009). However, there will be some limitations in the overall efficacy of these molecular scaffold proteins as therapeutic or diagnostic agents due to the host immune response, which ultimately results in neutralising the biologics. Nevertheless, the successful use of recombinant hCpn10 (XToll) in phase II clinical trials (i.e. completion of immunogenicity and toxicity studies of phase I) (van Eden, 2008), demonstrates the high tolerance of host immune systems to rhCpn10 (with $\leq$ 20 % mutations).

An additional limitation of molecular scaffolds, beyond their capability as neutralising agents, is the lack of a capacity to initiate the effector function of the innate immune system, mediated through the Fc portion of endogenous antibodies. Molecular scaffolds could be fused with an Fc fragment or other effector molecules, in order to facilitate the required therapeutic functions (Wu and Sun, 2014, Liu et al., 2012).

Another limiting factor in the use of alternative molecular scaffolds is the rapid metabolic clearance rates of these biologics, due to their size, by the renal system. Normally, small proteins ($\leq$ 60 kDa) are rapidly cleared from the serum and thus have shorter half-lives (< 1 hr) (Vugmeyster et al., 2012, Wu and Sun, 2014, Nord et al., 1997). However, these limitations can be overcome by increasing the molecular size of the biologics, which subsequently increases the serum half-life. Alternatively, the protein is modified through PEGylation, fusion to serum albumin, attachment to an antibody or antibody fragments, or other such methodologies (Nilsson and Tolmachev, 2007, Bernstein and Moellman, 2012, Xu et al., 2002). However, Cpn10 (70 kDa) has a half-life of up to 7 hrs (van Eden, 2008), which benefits short-term treatments and minimises the requirement of multiple dose administrations. Strategies such as PEGylation could be used to increase serum half-life if required.

In summary, hCpn10 is an ideal molecule for development as a molecular scaffold. NMEs based on hCpn10 molecular scaffold have the potential to exhibit ideal binding kinetics and parameters (k_a, k_d and K_D) to the target, display high potency, elicit minimal immunogenicity, have stability within a formulation, display ideal pharmacokinetics and show favourable bioprocess economics. Such attributes and qualities can be either inherent within the properties of a molecular scaffold or can be incorporated throughout the drug development process. hCpn10 inherently possesses ideal properties that fulfil many of the requirements for a molecular scaffold, and through appropriate production bioprocesses and formulation, has the potential to create a new class of NMEs for both diagnostic and therapeutic applications.

1.8 Hypothesis on which the research reported in this thesis is based

As previously stated, the hypothesis put forward in this thesis poses a number of leading questions regarding the suitability of hCpn10 as a molecular scaffold. These are as follows:

- Can the mobile loop of hCpn10 be substituted with non-hCpn10 peptide sequence, and can NMEs be created with binding specificity to specific targets?

- Will peptide substitution hinder the assembly of hCpn10 into heptamer from monomer?

1.9 Structure of the thesis

The general materials and methodologies used in these studies are detailed in Chapter 2. The establishment of fundamental basis for choosing hCpn10 as a molecular scaffold and optimisation of the molecular scaffold based-hCpn10 using dual computational and experimental approaches are detailed in Chapter 3. Application of the concept of using scaffold based-hCpn10 for therapeutic purposes is described in Chapter 4 with an example, supported with *in vitro* assays. A second proof of concept example NME based on hCpn10, as a diagnostic is provided in Chapter 5 with detailed analytical analyses. Chapter 6 discusses preliminary results of two hCpn10-based candidates with anti-angiogenic properties for cancer treatment. Finaly, Chapter 7 summarizes the outcomes of the study followed by recommendations for future directions of using scaffold based-hCpn10 for many applications.

Chapter 2. General Materials and Methods

2.1 Materials

Table 2.1 details the materials used in this research and their manufacturers/suppliers.

Table 2.1 List of materials for the research study

Item(s)	From
DNA Oligonucleotides:	
Target constructs (Table 2.2)	GeneArt® Gene Synthesis
DNA Primers (Table 2.8)	Sigma-Aldrich®
***E.coli* strains:**	
OneShot® Top10 cells	Invitrogen™
α-select cells	Bioline™
XL1-Blue cells	Stratagene
BL21(DE3) T7 Expression *LysY/I*^{q} cells	New England *BioLabs*® (NEB)
Media:	
Peptone from Casein	Becton Dickinson Co. (BD™)
Yeast Extract	BD™
LB Agar	Bio-Rad™ *Laboratories*
Antibiotics and chemical reagent:	
Ampicillin (sodium salt)	Fisher BioReagents®
Kanamycin (sulphate)	Fisher BioReagents®
Isopropyl-1-thio-β-D-galactopyranoside (IPTG)	Bioline™
Enzymes, Enzyme Buffers, and PCR reagents:	
EcoRI, BamHI	Roche™
NcoI, NotI, NdeI, XbaI	NEB
NEB Enzyme Buffer 1, 2, 3, 4	NEB
DNase I	Invitrogen™
cOmplete ULTRA multiple proteases Tablets	Roche™
Calf intestinal alkaline phophatase (CIAP)	Invitrogen™
AccuPrime™ Taq DNA Polymerase High Fidelity	*life* technology™
PCR Master Mix (2x)	Promega™
Antibodies:	
Monoclonal Anti-6xHis HRP-Conjugated	Abcam®
Mouse Monoclonal Anti-Human Cpn10	Abcam®
Molecular weight standards and dye:	
HyperLadder™ 100 pb	Bioline™
GeneRuler™ 1 kb DNA Ladder	Fisher Scientific™
SeeBule® Plus2 prestained protein standard	Invitrogen™
BioRad™ Gel filtration standard	Bio-Rad™ *Laboratories*
SimplyBlue™ SafeStain (Coomassie® G-250)	Invitrogen™

Item(s)	From
Kits:	
PureLink® Quick Plasmid Miniprep Kit	Invitrogen™
QIAEX-II® Gel Extraction Kit	QIAGEN™
Rapid DNA Ligation Kit	Roche™
Pierce™ Bicinchoninic acid (BCA) Assay Kit	Thermo™ Scientific
Columns and gel:	
HisTrap™FF crude, HisTrap™FF, HisTrap™HP 1-5 mL	GE Healthcare™
HiTrap™SP FF, HiTrap Capto™S 1-5 mL	GE Healthcare™
Superdex™75 GF columns	GE Healthcare™
Superdex™200 GF columns	GE Healthcare™
HiTrap™ Desalting, HiPrep™ 26/10 Desalting columns	GE Healthcare™
SnakeSkin® Dialysis Tubing	Thermo™ Scientific
TSK G3000SWXL SE-HPLC column	Tosoh™ Bioscience
Guard column (7.8mm)	Tosoh™ Bioscience
4-12% NuPAGE® Bis-Tris gel 12, 15 and 17-wells	Invitrogen™

2.2 Wild-type hCpn10 and designed variants

The hCpn10 and its variants are listed in Table 2.2 with their corresponding binding targets and amino acid (AA) sizes. AA sequences are described in more detail in Appendix 1.

Table 2.2 Wild-type hCpn10 protein and its variants and their amino acids size

No.	Construct term	Binding Target(s)	AA Size
1.	**hCpn10**	Cpn60	102
2.	**ΔL-Cpn10**	--	86
3.	**ΔRoof-Cpn10**	--	90
4.	**β-barrel (ΔL-ΔRoof-Cpn10)**	--	78
5.	**FH-Cpn10**	Cpn60	726
6.	**CE76**	FVII/FVIIa	102
7.	**$CE76_{P1}$**	FVII/FVIIa	108
8.	**$CE76_{P2}$**	FVII/FVIIa	106
9.	**$CE76_{E5}$**	FVII/FVIIa	112
10.	**FH-CE76**	FVII/FVIIa	726
11.	**CPMyc**	9E10	96
12.	**$CPMyc_{P1}$**	9E10	102
13.	**$CP7_{CD44\text{-}P1}$**	CD44	104
14.	**$CP11_{bFGF\text{-}P1}$**	bFGF	102
15.	**$CR2_{VEGF\text{-}P2}$**	VEGF1	111

2.3 Cells culture medium

The various culture media used for certain applications are described in Table 2.3.

Table 2.3 Media preparation applied for bacterial growth

Media	Compositions	Application(s)
Luria broth (LB)	1 % Peptone from Casein; 0.5 % Yeast Extract; and 1 % NaCl) per L of MiliQ water, autoclaved at 121 °C (250 °F), for 20-60 min, at 15 psi.	For plasmid growth amplification.
Super Optimal Broth with Catabolite repression (SOC)	10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2 % Peptone from Casein; and 0.5 % Yeast Extract) per L of MiliQ water, autoclaved at 121 °C (250 °F), for 20-60 min, at 15 psi. A 20 mM glucose filtered sterilizes by (0.22 µm) and added post-autoclaving.	To obtain maximal transformation efficiency of *E*. coli.
Terrific broth (TB)	1.2 % Peptone from Casein; 2.4 % Yeast Extract; 72 mM K_2HPO_4, 17 mM KH_2PO_4 and 0.4 % glycerol) per L of MiliQ water, autoclaved at 121 °C (250 °F), for 20-60 min, at 15 psi.	For small to lab-scale protein expressions (from 5 - 5000 mL).
Defined media (DM)	Solution 1: KH_2PO_4 (16.6 g/L), $(NH_4)_2HPO_4$ (4 g/L), Citric Acid (2.1 g/L). Solution 2: Glycerol (25 g/L), $MgSO_4.H_2O$ (1.5 g/L). Solution 3: Iron(III) citrate hydrate (75 mg/mL), Boric acid (3.8 mg/mL), $MnCl_2.4H_2O$ (18.8 mg/mL), EDTA (10.5 mg/mL), $CuCl_2.6H_2O$ (1.9 mg/mL), $Na_2MoO_4.2H_2O$ (3.1 mg/mL), $CoCl_2.6H_2O$ (3.1 mg/mL), Zinc acetate dihydrate (10 mg/mL). Solution 4: Thiamine HCL (2 mg/mL). All solutions sterilized by autoclaving (as above) or 0.22 µm filtering.	For large and single-batch cultivation (from 4 - 7 L). Note: Only mixed (solution 1 to 4) prior to expression and under sanitizes conditions with final pH of 6.7, adjust with NH_3 (27 %).

2.4 Buffer system

2.4.1 Buffer preparation

The choice of buffers (Table 2.4) was based on the nature and the pH suitability for target protein. This depends on pre-screening test for determining the stability of each target protein with respect to pH and the buffering compound. It also depends on the suitability of purification procedures. This was crucial to avoid loss of time and protein caused by an

additional step of buffer exchange. In addition, the lysis buffer should be compatible with the first chromatography step. It was necessary to improve the stability of the target protein and to keep the protein active in solution by including additives (e.g. salts, stabilizer, solubiliser, anti-oxidant agent, etc). Also, a cocktail of protease inhibitors was added to the lysis buffer to control the undesirable proteolysis effect during cells disruptions.

Table 2.4 Commonly applied buffers with stabilizing components and optimum pH value

Buffer Name	Components	pH
Lysis (+)	50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT or 0.05 % BME, 5 % glycerol, 1 mM EDTA, 1 tablet/10mL cOmplete ULTRA (proteases cocktail), 1 µg/mL DNaseI, 0.1 % Triton X-100.	7.6
Lysis (-)	50 mM Tris-HCl, 150 mM NaCl, 10 mM Imidazole.	7.6
PBS	8.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, 137 mM NaCl,	7.4
PBST	PBS, 0.1 % (v/v) Tween 20.	7.4
TBS	25 mM Tris-HCl, 2.7 mM KCl, 137 mM NaCl.	7.6
TBST	TBS, 0.1 % (v/v) Tween 20.	7.6
IB Washing	100 mM Phosphate, 1 M Urea, 1 % Trion X-100	7.4
IB Denaturation	100 mM Phosphate, 150 mM NaCl, 8 M Urea, 2 mM DTT or 0.5 % BME	7.4
IMAC Biding buffer I	20 mM Phosphate, 0.5 M NaCl, 20-40 mM imidazole, 0.01 % BME.	7.6
IMAC Biding buffer II	50 mM Tris, 150 mM NaCl, 20 mM Imidazole	8.0
IMAC Biding buffer III	100 mM Phosphate, 150 mM NaCl, 0.05 % BME, 8 M Urea, 10 mM Imidazole	7.8
IMAC Elusion buffer I	20 mM Phosphate, 0.5 M NaCl, 0.5 M imidazole, 0.01 % BME	7.6
IMAC Elusion buffer II	50 mM Tris-HCl, 150 mM NaCl, 0.5 M Imidazole	8.0
IMAC Elusion buffer III	100 mM Phosphate, 150 mM NaCl, 0.05 % BME, 8 M Urea, 0.5 M Imidazole	7.8
Exchange Buffer I	20mM Phosphate, 150mM NaCl	7.8
Exchange Buffer II	50 mM Tris, 150 mM NaCl	8.0
Exchange Buffer III	100mM Sodium Acetate, 150 mM NaCl	5.5
Refolding (4-0)	➢ 100 mM Phosphate, 150 mM NaCl, 4 M Urea, 0.05 % BME. ➢ 100 mM Phosphate, 150 mM NaCl, 2 M Urea, 0.05 % BME. ➢ 100 mM Phosphate, 150 mM NaCl, 1 M Urea, 0.05 % BME. ➢ 100 mM Phosphate, 150 mM NaCl, 0 M Urea, 0.05 % BME.	7.4
IEX Biding buffer	50 mM Sodium Acetate	5.5
IEX Elution buffer	50 mM Sodium Acetate, 1 M NaCl	5.5

Buffer Name	Components	pH
GF Running Buffer I	100 mM Phosphate, 150 mM NaCl, 1 mM DTT	7.6
GF Running Buffer II	50 mM Tris, 150 mM NaCl, 1 mM DTT	8.0
GF Running Buffer III	100 mM Sodium Acetate, 150 mM NaCl, 1 mM DTT	5.5
SE- Mobile Phase I	0.1 M Phosphate, 0.2 M NaCl	7.8
SE- Mobile Phase II	100 mM Sodium Acetate, 150 mM NaCl	5.5

* All buffers were sterilised by filteration (0.45 µm or 0.22 µm filter), prior to applications.
** PBS – Phosphate buffered saline, TBS – Tris buffered saline, IB – inclusion bodies, IMAC – Immobilized metal ion affinity chromatography, IEX – Ion exchange chromatography, GF – Gel filtration chromatography, SE – Size-exclusion HPLC.

2.4.2 Antibiotics

Table 2.5 Applied antibiotics in stock and working concentrations.

Antibiotic	Stock Concentration*	Working Concentration
Ampicillin (Amp)	100 mg/mL	100 µg/mL
Kanamycin (Kan)	30 mg/mL	30 µg/mL

*Filter sterilized (0.22 µm), aliquots stored in 1 mL (-20 °C).

2.5 Standard molecular techniques

2.5.1 Preparation of plasmids: growth, extraction and purification

2.5.1.1 Transformation into competent *E. coli* cells

DNA sequences encoding target proteins (Table 2.2) were designed *in silico* and synthesised with codon optimisation for expression in *E. coli*. The lyophilised DNA received from GeneArt® Gene Synthesis was transformed into *E. coli* OneShot® Top10 competent cells (Invitrogen™) for plasmid DNA transmission. Approximately 2 µL of each plasmid DNA (pGA-Script/Protein X) was added to a 50 µL of Top10 cells. Cells and DNA were incubated on ice for 30 min before heat-shocked at 42 °C for 30 sec. The heat-shocked cells containing plasmid DNA were placed back onto ice for 5 min before addition of 950 µL of SOC medium and allowed to recover at 37 °C with shaking for 1 hr. Approximately 50 µL (of 1:100 dilution) of recovered culture were plated onto an LB-Agar/(Amp 100 µg/mL) plate for selection. The plate was incubated overnight at 37 °C.

2.5.1.2 Small-scale plasmid preparation and DNA purification (MiniPrep)

A single colony from LB-Agar/Amp plate was used to inoculate 5 mL LB/Amp (37 °C, 220 rpm). A PureLink® Quick Plasmid Miniprep Kits were used for plasmid DNA extractions per the manufacturer's instructions. In brief, a 1.5 mL from overnight LB culture was pelleted at 14000xg for 5min. Pellets were resuspended in 250 µL PureLink®Resuspension buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA) with RNase A (20 mg/mL) and denatured with 250 µL PureLink®Lysis buffer (200 mM NaOH, 1 % w/v SDS), mixed gently and incubated for 5 min at room temperature (RT). A 350 µL PureLink®Precipitation buffer was added for precipitating protein contaminants by vortex mixing and pelleted by 14000xg for 10 min. After plasmid DNA was bound to spin the column, the supernatant was discarded. A 500 µL Wash buffer, containing 100 % ethanol was added, incubated for 1 min, then discarded after centrifugation at 12000xg for 1 min. DNA plasmid was recovered from adding 75 µL TE buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA), incubated for 1 min, 12000xg for 2 min at room temperature. Plasmid DNA concentrations were determined and stored in aliquots at -20 °C.

2.5.1.3 Quantification of nucleic acid concentration and purity determination

Quantification of DNA recovered from miniprep was determined by using NanoDrop1000 spectrophotometer (Thermo™ Scientific) and calculated from Equation 2.1.

$$DNA\ (\mu g/\mu L) = \frac{OD_{260} \times 50\ \mu g/mL}{1000} \times DF \qquad \textbf{Equation 2.1}$$

Based on the standard DNA formula a 50 µg/mL of DNA (at OD_{260}) is 1 unit and DF is the dilution factor of sample. The purity of DNA sample was measured relative to ration of OD_{280} to OD_{260}, whereby pure DNA has ratio of 1.8-2.0, and a spectrophotometer path length of 1 cm. The approximate concentrations of the DNA samples were 0.1-0.5 µg/µL.

The size and purity of plasmid DNA was visualized by agarose gel electrophoresis. A 1-5 µL of undigested and/or restrictions enzyme digested DNA was run in parallel with molecular weight standard, as detailed in Section 2.5.4.

2.5.2 Primer designs

Polymerase Chain Reaction (PCR) and sequencing of oligonucleotide primers was designed with enrichment of GC content (>50 %) and ended at the 3' termini with a G or C. Appropriate restriction enzyme sites were included at the 5' termini for PCR cloning primers, and extra nucleotides were incorporated for optimal cleavage according to restriction enzymes (NdeI/NdeI/NotI and EcoRI/BamHI/XbaI) and followed manufacturer's instructions. The primer oligonucleotides resuspended in MilliQ water to concentration of 1 µg/µL, and diluted into a working concentration of 100 ng/µL.

Table 2.6 Oligonucleotide primers used for PCR and DNA sequencing

Primer	bp Sequence (5'- 3')	bp Size
Cpn10_NcoI_For	GGTACCCC**ATG**GCGGGCCAG	**20**
Nde1_Cpn10_For	GAGATATACAT**ATG**GCGGGCCAGGCGTTTCG	**31**
Cpn10_His6_ EcoRI_Rev	CTGGGCAAATATGTGGATCATCATCACCATCA CCAT**TAA**GAATTCGAGCTC Reverse Complement: GAGCTCGAATTCTTAATGGTGATGGTGATGAT GATCCACATATTTGCCCAG	**51**
Cpn10_G4His6_EcoRI_ Rev	CTGGGCAAATATGTGGATGGTGGAGGTGGACA TCATCACCATCACCAT**TAA**GAATTCGAGCTC Reverse Complement: GAGCTCGAATTCTTAATGGTGATGGTGATGAT GTCCACCTCCACCATCCACATATTTGCCCAG	**63**
T7_pET30a_For	AGTTATTGCTCAGCGGTGGCAGC	**23**
T7_pET30a_Rev	CCACGCCGAAACAAGCGCTC Reverse Complement: GAGCGCTTGTTTCGGCGTGG	**20**

* **Bold** blue: Start codon. **Bold** red: Stop codon.

2.5.3 Restriction enzymes digestions

Appropriate restriction enzymes were applied for digestion of cloning and/or expression plasmids following manufacturer's instructions. Approximately 1 µg DNA (plasmid and/or PCR product) was mixed with appropriate concentration of restriction enzymes 0.5 µL NdeI/NdeI/NotI, 0.5 µL EcoRI/BamHI/XbaI, 2.5 µL NEB Buffer-4, MilliQ water was added to reach a final volume of 25 µL, and incubated at 37 °C for 2 hrs (incubation time varied based on restriction enzyme efficiency and buffer components). DNA digestions were determined by visualizing DNA bands with 1 µL loading dye, loaded into agarose gel electrophoresis.

2.5.4 Agarose gel electrophoresis

All nucleic acids in this study were visualised using a standard ethidium bromide stained agarose gel electrophoresis. For visualisation and/or separation of DNA: a fresh preparation of 1-2 % (w/v) agarose gel in 1x Tris-Acetate-EDTA (TAE) buffer was mixed with 2 µL UltraPure™ Ethidium bromide (10mg/mL, Invitrogen™). Then approximately 1-4 µL DNA samples loaded with 1 µL DNA-loading dye, electrophoresed at 90 V for 35 min or until it achieved an appropriate DNA separation size determined by molecular weight marker HyperLadder™ 100 bp (Bioline™) or GeneRuler™ 1 kb DNA Ladder (Fisher Scientific™), were run in parallel. Better separation and visualization occurred by using 1.5 % (w/v) agarose gels.

2.5.5 General DNA amplification, insertion and sequencing

2.5.5.1 Polymerase chain reaction

PCR reactions consisted of purified DNA (from Miniprep, used as template) ~10 ng (0.1 pg – 200 ng); 5 µL AccuPrime™ 10x reaction buffer of: (600 mM Tris-SO_4 (pH 8.9), 180 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 2 mM dGTP, 2 mM dATP, 2 mM dTTP, 2 mM dCTP, thermostable AccuPrime™ protein, 10 % glycerol); Cpn10_For_Primer (0.1 µM); Cpn10_Revr_Primer (0.1 µM); 0.4 µL *Taq* Polymerase, Accuprime™ Taq of: (r*Taq* DNA polymerase, *Pyrococcus* species *GB-D* polymerase, and Platinum® *Taq* Antibody); and MilliQ water to 50 µL. A negative control with template DNA replaced with MilliQ water was always included. DNA samples were amplified using Veriti Thermal Cycler (*life*

technologies™), following optimized PCR conditions protocol: initial denaturation at 94 °C for 5 min; denaturation at 94 °C for 30 sec, annealing at 55 °C for 1 min, extension at 72 °C for 1 min, for 30 cycles; and a final extension at 72 °C for 10 min. PCR products were visualised by agarose gel electrophoreses and stored at 4 °C (Appendix 2).

2.5.5.2 Enzymes digestions of DNA/PCR product

Digestion of plasmid and PCR products were performed simultaneously in separate tubes following manufacturer standard protocol. Restriction enzymes in Table 2.9 were also applied for such digestions. Approximately 1 µg DNA (plasmid and/or PCR product) mixed with appropriate concentration of restriction enzymes 0.5 µL NcoI, 0.5µL EcoRI, 2.5 µL NEB Buffer and MilliQ water was added to final volume of 25 µL and incubated at 37 °C for 2 hrs (incubation time may vary based on restriction enzyme efficiency and buffer components). Vector plasmid was dephosphorylated by 1 µL of calf intestinal alkaline phophatase (CIAP) incubated for further 1 hr as per recommendation (Invitrogen™). Digested and undigested products (of plasmid and PCR) were visualized as DNA bands with 1 µL loading dye, separated by agarose gel electrophoresis (see Section 2.5.4).

2.5.5.3 Ligation of DNA fragments

Digested DNA was extracted from agarose gels using QIAEX-II® Gel Extraction Kit (QIAGEN™). In brief, a 60 µL Buffer QG (3:1 ratio to sample) of DNA digestion mixtures or fragments from agarose gel sliced was incubated at 50 °C for 10 min, following all manufacturer instructions but excluding optional steps. Final clean DNA fragment was eluted from QIAquick column by 50 µL EB® buffer (10 mM Tris·Cl, pH 8.5) or TE® buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA). The sample were then stored at 4 °C.

Rapid DNA Ligation Kit (Roche™) was applied for insert/vector DNA ligations following standard protocol. A molar ratio of vector DNA (Figure 2.1) to insert DNA is (1:5 or 1:2), a 2 µL digested vector, mixed with 10 µL insert, 5 µL T4 DNA ligation buffer (2x), 1 µL Dilution buffer (5x), 0.4 µL T4 DNA ligase was incubated for 30 min, at RT. Controls of ligations were performed simultaneously, with excluding insert DNA fragment or vector.

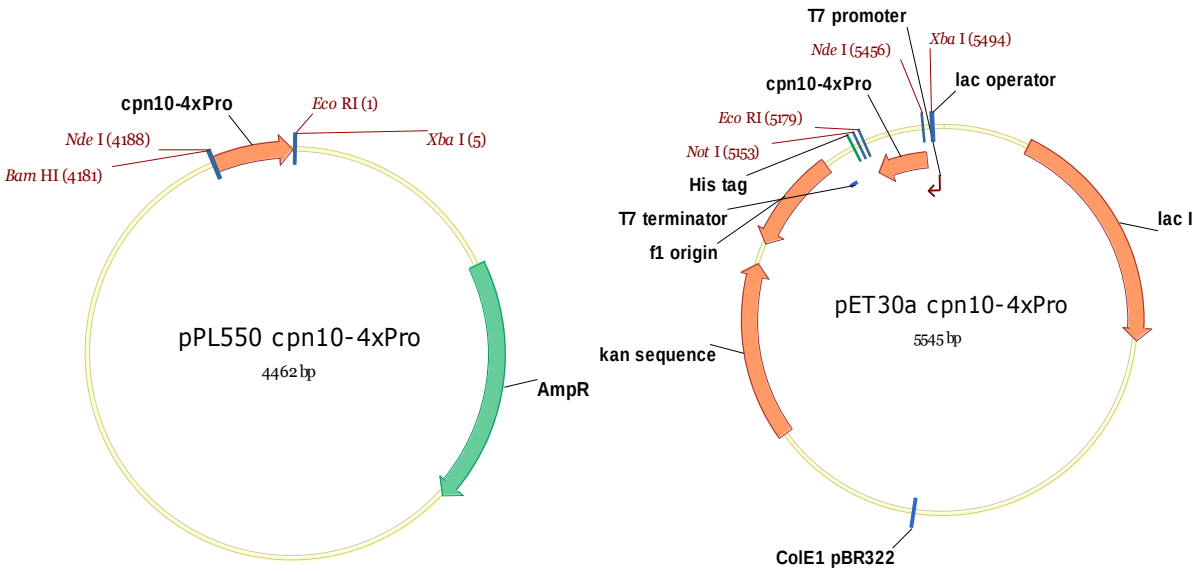


Figure 2.1 Expression vector maps of pPL550 and pET30a(+) used in this study

2.5.6 Transformations and positive colony identifications

Ligated products (i.e. expression plasmid with DNA fragments - encoding targeted proteins) were transformed into competent cells as previously described in Section 2.5.1.1 with minor modifications. Approximately 5 µL of ligated product was added to a 50 µL of *E. coli* competent Top10 cells (Invitrogen™) or α-select cells (Bioline™), incubated on ice for 30 min before heat-shock at 42 °C for 15-45 sec (based on recommendations). The cells were heat-shocked and were placed back onto ice for 5 min. Final cells were mixed with 950 µL of SOC medium and allowed to recover at 37 °C with shaking for 1 hr. Approximately 50-80 µL (of 5:100 or 1:100 diluted, respectively) of recovered culture were plated onto an LB-Agar/(Kanamycin 30 µg/mL) or LB-Agar/(Ampicillin 100 µg/mL) plate for selection. Plate was incubated overnight at 37 °C.

Random choosing of colonies for DNA sequence confirmation was done using colony PCR and visualised by gel electrophoresis. Following plasmid preparation procedures in Section 2.5.1.2, positive colonies were grown individually in 5 mL LB/Kan or LB/Amp media culture, shaking overnight at 37 °C, whereby, DNA sequencing were performed on pure plasmid DNA extracted from miniprep.

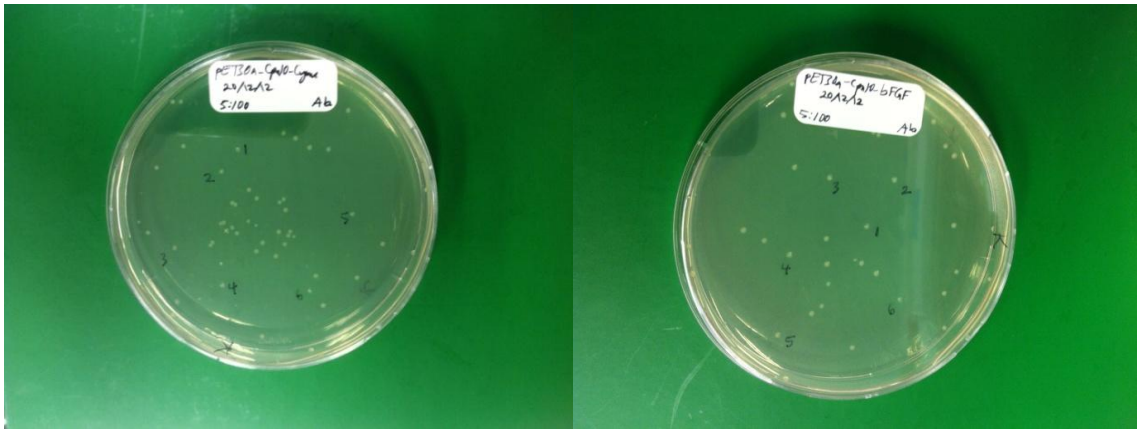


Figure 2.2 *E. coli* α-select cells with expression vector (pET30a-X) encoding target proteins

Examples of LB-Agar/Kan with colonies grown from successfully transformed *E. coli* α-select cells with expression vector (pET30a-X) encoding target proteins. Randomly selected colonies were numbered, picked and applied for cPCR and DNA sequencing.

2.5.6.1 Colony PCR (cPCR)

Individual colonies (Figure 2.2) picked onto a sterile pipette tip were added into PCR reaction mixtures of 6.5 µL PCR Master Mix (2x) (Promega™, containing 50 units/mL of *Taq* DNA polymerase, reaction buffer (pH 8.5), 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dTTP, 3 mM $MgCl_2$), 0.5 µL T7_For_Primer (1 µM); 0.5 µL T7_Revr_Primer (1 µM); and MilliQ water to 12 µL. A negative control with no colony added was always included. Samples were amplified using Veriti Thermal Cycler (*life* technologies™), following optimized cPCR conditions protocol: initial denaturation at 94 °C for 10 min; denaturation at 94 °C for 30 sec, annealing at 60 °C for 30 sec, extension at 72 °C for 1 min, for 25 cycles; and a final extension at 72 °C for 8 min. cPCR products were visualised by agarose gel electrophoreses and was stored at 4 °C (Appendix 2).

2.5.6.2 Nucleotide sequencing

Nucleotide sequencing was performed for purified plasmid DNA (miniprep) using ABI Prism® BigDye™ terminator (PE Applied Biosystems) as per manufacturer recommendations. Sample was prepared: as 1-5 µL (equivalent to 50-100 ng of >2000 bp) of pure plasmid DNA, was added to 1 µL (0.1 µM) primer (F or R), mixed with 8 µL ABI Prism®, and MilliQ water to 20 µL. Samples were electrophoresed and scanned by Australian Genome Research Facility (*agrf*), University of Queensland.

Software tool VectorNTI was used for plasmid sequence analysis and mapping. Multiple sequence alignments were performed using online BLAST tool (NCBI, 2014). DNA/Protein sequence translation were performed online using ExPASy Bioinformatics Resource Portal (ExPASy, 2014a).

2.5.6.3 Transformation into E. coli BL21(DE3)/E. coli XL1-Blue expression strains

E. coli BL21 (DE3) T7 Expression LysY/I^q competent cells (New England *BioLabs*®) were transformed with plasmid (pET30a-Insert-X-ligated) following a recommended transformations protocol. Similarly, *E. coli* XL1-Blue competent cells (Stratagene) was transformed with expression vector (pPL550-Insert-X-ligated). In brief, a 50 µL aliquot of competent cells was thawed in ice and incubated with 1-5 µL plasmid DNA (approximate 100-500 ng) for 30 min in ice. Normally, 1 µL of supercoiled plasmids and 5 µL of ligated products. Mixtures were heat-shocked at 42 °C for 10 sec (or more upon requirement but with maximum of 45 sec), and re-incubated in ice for further 5 min. A 950 µL of SOC (sterilised) broth was added to transformed cells, incubated with shaking for 1 hr at 37 °C. 50-80 µL of mixture (5:100 dilutions, respectively, allow individual colonies to grow) were spread onto LB-agar plate with appropriate antibiotic (Kan or Amp), incubated for overnight at 37 °C. Freshly grown colonies were applied for inoculation expression cultures and/or preparation of 15-25 % (v/v) glycerol stock (-80 °C) for future expression.

2.6 Protein expression

2.6.1 Expression from *E. coli* XL1-Blue-pPL550

2.6.1.1 *E. coli* culture growth and protein expression (≤ 1000 mL)

Following schematic procedures below (Figure 2.3), a single colony of *E. coli* XL1-Blue-pPL550 encoding recombinant proteins (Table 2.2) of hCpn10 or CE76, was used to inoculate 5 mL LB media with 100 µg/mL Amp, and incubated overnight at 37 °C, shaking at 200 rpm. A dilution (1:100) of overnight culture was used to inoculate 50 or 100 mL LB or TB media flask containing 100 µg/mL Amp, and continued to be grown at 37 °C, 220 rpm shaking, until OD_{600} reached an approximate 0.5-0.7. Protein expression was induced by increasing the temperature to 42 °C and continued for further 3-4 hrs. A negative induction

was performed for each protein by keeping temperature at 37 °C. The cell density at OD_{600} was measured at hourly intervals for pre- and post-inductions. *E. coli* cells were harvested by centrifugation at 5000xg at 4 °C for 30 min. Then, cell pellets were frozen and stored at -20 °C prior to cell lysis.

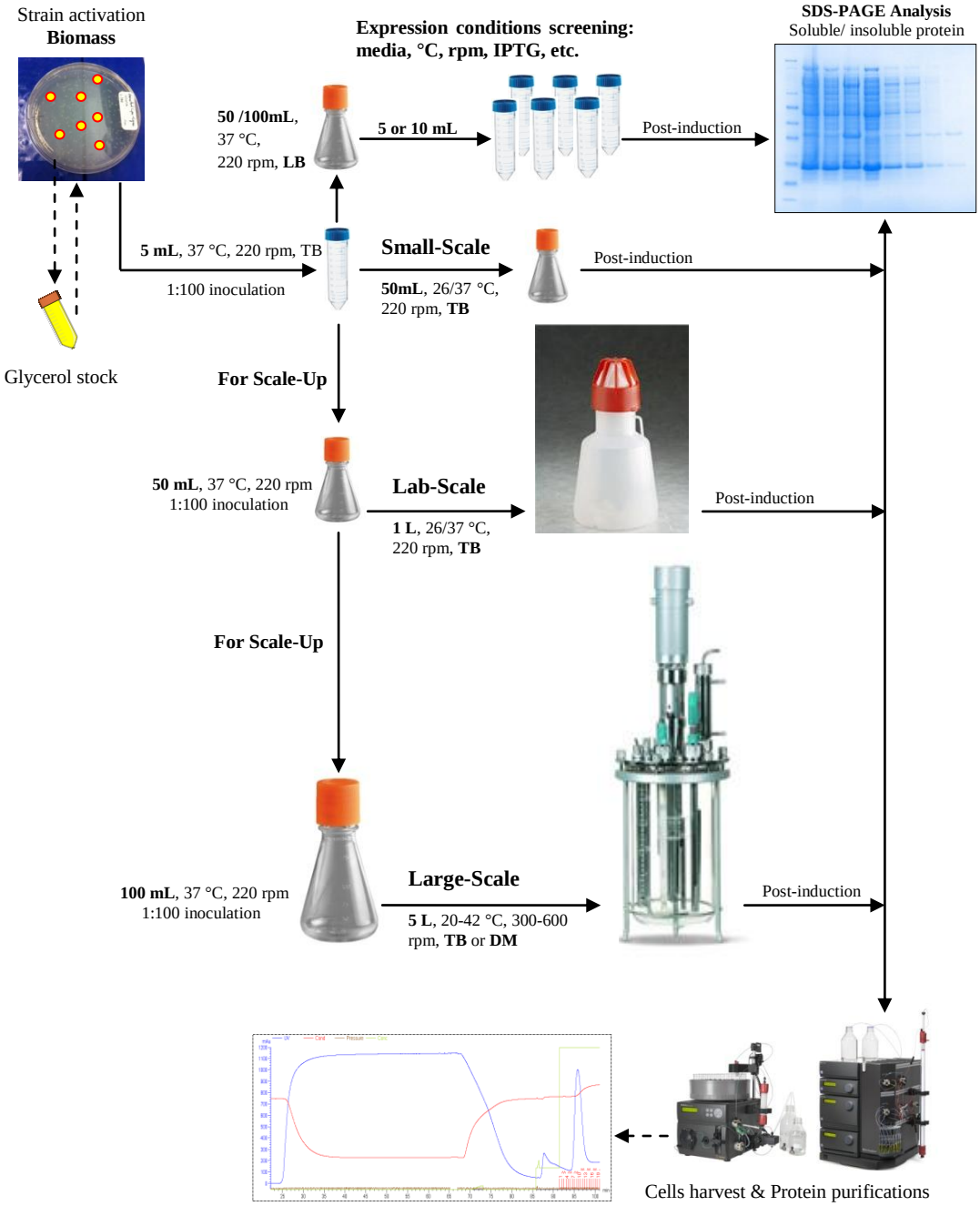


Figure 2.3 Scheme for recombinant protein expressions

The scheme applied for quick expression screening trials and for scaling up processes for recombanint proteins examined in this study. Different expression media were used: LB; Luria broth, TB; Terrific broth, and DM; Defined media (detailed in Table 2.3).

2.6.1.2 Scale-up *E. coli* culture growth and protein expression (≥ 4 L)

A single colony of *E. coli* XL1-Blue-pPL550 containing recombinant proteins of hCpn10 or its variant CE76 (Table 2.2), was used to inoculate a 5 mL LB media with Amp 100 µg/mL for overnight at 37 °C and then shaking at 200 rpm. A dilution (1:100) of overnight culture was used to inoculate 100 mL TB media containing 100 µg/mL Amp, and continued to grow at 37 °C, 220 rpm shaking, until OD_{600} reached an approximate 1. Following schematic procedures in Figure 2.3, a second dilution (1:100) of 100 mL culture was used for inoculating 5 L culture medium of DM or TB media in a 7 L Applikon bioreactor (Applikon® Biotechnology). DM medium was the best for use due to the high enrichments of media and ability to control cultural conditions (Temperature, DO_2 %, stirrer speed, and pH). However, TB broth was also used as a cheaper option. 5 L of freshly made DM was pre-warmed to 37 °C, and then inoculated. The inoculated cell cultures continued to grow with appropriate antibiotic at 37 °C, 400-600 rpm, DO_2 (40-60 %), and pH 6.7. At 0.5-0.7 OD_{600}, targeted protein was expressed by increasing temperature to 42 °C, and continued for 4 hrs. The cell density at OD_{600} was measured at hourly intervals, and for pre- and post-induction. Due to the cost of overall large-scale production, the negative induction was performed simultaneously only on 100 mL cell cultures (shaking flask) by keeping temperature at 37 °C. Finally, *E. coli* cells were harvested by centrifugation at 5000xg at 4 °C for 30 min. Then, cell pellets were stored at -20 °C prior to cell lysis, as detailed in Section 2.7.1.

2.6.1.3 Measurement of Cell Density (OD_{600})

Samples from cell culture were collected and celldensities measured at 600 nm (OD_{600}) using UV-Vis SpectraMax Plus^{384} Spectrophotometer. Samples were collected and measured at hourly intervals of cell culture growth as well as samples of pre- and post-inductions. In brief, a 1 mL of cell culture was diluted 1:1 to 1:100 in culture medium, and measured at 600 nm (blanked by same dilution media). The optical densities for all samples were analysed by SoftMax®Pro statistical software, and OD_{600} plotted against time in hour.

2.6.2 Protein expression from *E. coli* BL21(DE3)-pET30a

2.6.2.1 *E. coli* culture growth screening (Expression trials of recombinant proteins)

A single colony of *E. coli* BL21(DE3)-pET30a plate or glycerol stock (stored at -80 °C) containing recombinant proteins of hCpn10 and/or CE76 was used to inoculate 5 mL LB or TB media, containing 30 µg/mL Kan, shaken at 200 rpm overnight at 37 °C. A dilution (1:100) of overnight culture was used to inoculate 500 mL LB or TB media containing Kan and was incubated at 37 °C, 220 rpm shaking, until OD_{600} reached an approximate 0.6-0.8. The culture flask was divided into 10 mL cultures of non-induced culture (control) and multiple induced cultures at 20, 26, 30, and 37 °C. Protein expression was induced by adding IPTG to final concentrations of 0.1, 0.4, 1.0 mM, and continued for 3 hrs, 4 hrs, or overnight (approximately 14 hrs). To monitor protein expressions over time, a 100 µL (2x) of cell culture were collected at initiation, then at hourly intervals up to 4 hrs post-induction, then after overnight incubation and OD_{600} measured, as described in Section 2.6.1.3. Samples were centrifuged at top speed 16000xg for 5 min, supernatants discarded and cell pellets stored at -20 °C for SDS-PAGE analysis.

The above experiments were performed over different time frames based on whether the culture medium used was LB or TB as well as depending on induction conditions. These experiments were thought to be necessary for determining the optimum conditions for scalable expressions.

2.6.2.2 Small-scale protein expressions (≥ 100 mL)

A single colony of *E. coli* BL21(DE3)-pET30a encoding recombinant proteins from Table 2.2, was inoculated into 10 mL TB, containing 30 µg/mL Kan, shaking at 200 rpm for overnight at 37 °C. A dilution (1:100) of overnight culture was used to inoculate 100 mL TB media containing 30 µg/mL Kan, and was incubated at 37 °C, 220 rpm shaking, until OD_{600} reached 0.6-0.8. The culture flask was divided equally into two, as induced culture and non-induced culture (control). Protein expression was induced by adding IPTG to final concentrations of 0.4 mM, and continued for further 4 hrs, at reduced temperature of 26 °C. 1 mL samples were collected for cell density (OD_{600}) pre- and post-induction, centrifuged at 16000xg for 5 min, and cell pellets stored at -20 °C for SDS-PAGE analysis. Final culture

cells were harvested by centrifugation at 5000xg at 4 °C for 30 min. Then, cell pellets were stored at -20 °C prior to cell lysis (Appendix 2).

2.6.2.3 Large scale protein expressions (≥ 1L)

Large scale (≥ 1L) protein expression was carried out as described in Section 2.6.1.2 above, except that 100 mL TB-culture was used to inoculate 1 L of TB medium, containing appropriate antibiotic. The 1 L TB culture continued to grow in a well ventilated 2.5 L baffled flask (from Tunair™) at 37 °C, and shaking 220 rpm, until OD_{600} reached 0.6-0.8. Cultures were set to the optimal conditions and induced with final concentration of 0.4 mM IPTG, at reduced temperature of 26 °C. After incubation for 3-4 hrs, the cells were harvested by centrifugation at 5000xg, at 4 °C, for 30 min. Cells pellets were resuspended in Lysis buffer (Table 2.4), and stored frozen at -20 °C.

Table 2.7 Recombinant protein constructs, plasmid, *E. coli* stains and applied expression conditions

Construct	Plasmid	*E.coli* strain	Induction conditions*
hCpn10 **ΔL-Cpn10** **ΔRoof-Cpn10** **β-barrel** **CE76** **6x*His*-CE76**	pPL550	XL1-Blue	Inducer: increase temp to **42** °C Time: 3, **4** hrs Culture medium: LB, TB, **DM** Culture size: 5, 50, 100, 1000, **5000** mL
CE76 **$CE76_{P1}$** **$CE76_{P2}$** **$CE76_{E5}$** **CPMyc** **$CPMyc_{P1}$** **$CP7_{CD44\text{-}P1}$** **$CP11_{bFGF\text{-}P1}$** **$CR2_{\text{-}VEGF1\text{-}P2}$**	pET30a	BL21 (DE3)	Inducer: 0.1, **0.4**, 1.0 (mM) **IPTG** Temperature: 20, **26**, 30 or 37 °C Time: **3**, 4, overnight (14) hrs Culture medium: LB, **TB**, DM Culture size: 10, 50, 100, **400**, 1000 mL
FH-Cpn10 **FH-CE76**	pET30a	BL21 (DE3)	Inducer: 0.4 (mM) IPTG Temperature: 26 °C Time: 4 hrs Culture medium: TB Culture size: 100, 400 mL

* Bold highlight the best conditions applied.

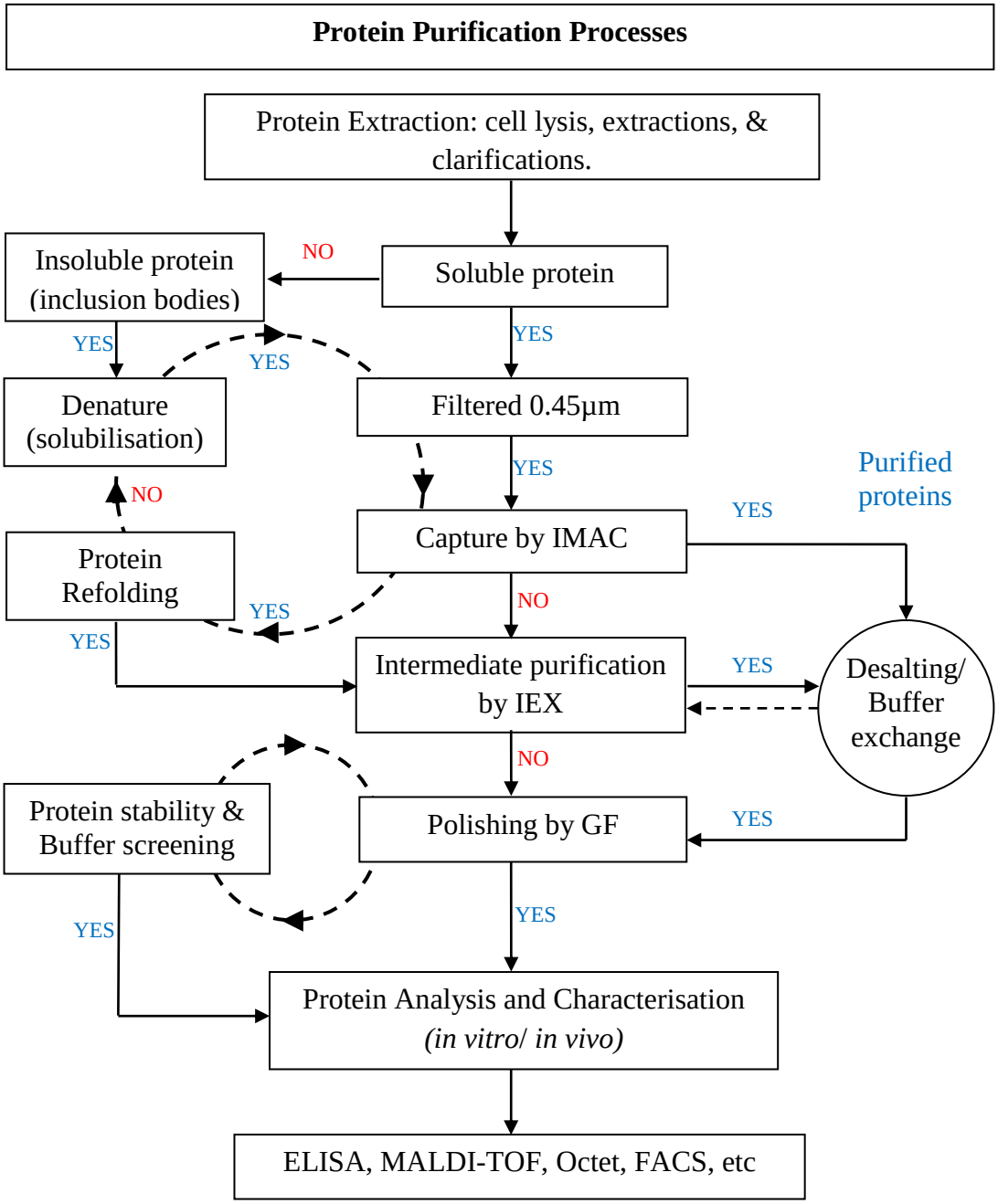


Figure 2.4 Steps in the sample protein preparations and purification

Scheme illustrates the processes applied for purification of each recombinant protein (Table 2.2). For a moderate purity of < 95 %, recombinant proteins were captured by affinity chromatography (IMAC) for removing bulk impurities, and/or followed by intermediate purification by ion exchange chromatography (IEX) for further isolation, removing impurities and concentrating of target proteins. For higher level of purity ≥ 95 %, the recombinant proteins were polished by gel-filtration chromatography (GF).

2.7 Recombinant proteins extraction and purification

2.7.1 Cell lysates for protein in native conditions

Harvested cell pellets following Figure 2.4, were cell pellets thawed at ≤ 4 °C and resuspended in Lysis buffer (Table 2.4) of 5 mL buffer per 1 mg of wet cell pellets. For small scale cell culture (≤ 50 mL), 1 mg/mL freshly made lysozyme and 5 µg/mL DNaseI were added to Lysis buffer and incubated for 1-2 hrs at ≤ 4 °C. For lab scale cell culture (≤ 100 mL), cells were resuspended in Lysis buffer (5 mL/1mg), with 5 µg/mL DNaseI, and sonicated on ice by 10 sec (5x) bursts at 200 W and cooling period of 10 sec between each round. For large scale cell cultures (> 100 mL), cells were resuspended in a pre-chilled Lysis buffer (5 mL/mg), and cells ruptured using a high pressure homogenizer at single pass of 15000 psi or 3x passes of 5000 psi (350 bar) with a cooling period of 5 min between each round. The approach of cell lysis and disruptions were based on culturel size and detailed are in Table 2.8.

Table 2.8 Common processes applied for E. coli intracellular recombinant protein extractions

Extraction process	Potency	Usage and comment(s)
Enzymatic digestions - Cell lysis-	Gentle	Small scale only (≤ 50 mL). Cell wall digested by osmotic disruption. Disadvantage: • Low product yield. • May Combined with mechanical disruption.
Ultra sonication	Moderate	Small to Lab-scale (50-100 mL). Microscale high-pressure sound waves cause disruption by shear forces and cavitations. Disadvantages: • Release of nucleic acid causing viscosity problem. • Required to be pre- and post- chill.
Homogeniser (EmulsiFlex-C5)	Vigorous	Lab to Large-scale (100-1000 mL), capacity 1-5 L/hr. Cells forced through small orifice at very high pressure; and shear force disrupts cells 500-30000psi/3.5-207MPa. Disadvantage: • Proteases release.
Homogeniser (AVP-1000)	Vigorous	Large-scale (≥ 1 L), capacity 22 L/hr. Cells forced through small orifice at very high pressure; shear force disrupts cells. Up to 1000 bar/14,500 psi. Disadvantage: • Proteases release.

Cell lysates were then centrifuged at a maximum speed (16000xg) at 4 °C for ≥ 30 min to separate cell debris and produce clear supernatants which contained soluble proteins including target recombinants. Protease inhibitor cocktails tablet (1/10 mL) was added to protect target proteins from any proteolytic activities, and it was then stored at 4 °C.

2.7.2 General recombinant proteins purification under native conditions

Chromatography purification were performed following cell lysis. Different purification schemes are illustrated in Figure 2.5. The purification schemes were based the on proteins state i.e. in native or in denatured form. Protein samples and buffers were pre-filtered with 0.45 µm and 0.22 µm, respectively. In every purification step, samples were retained for SDS-PAGE analysis.

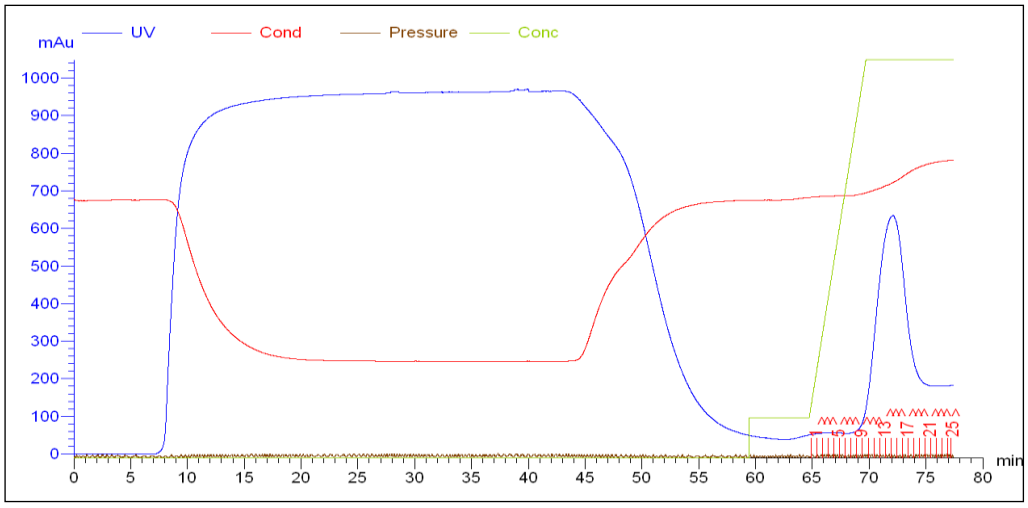


Figure 2.5 Affinity chromatography of CE76$_{E5}$

Representation chromatogm illustrating procedure of *His*-tagged CE76$_{E5}$ capturing and stepwise elusion using IMAC-Elution buffer: (20 mM Phosphate, 0.5 M NaCl, 0.5 M imidazole, 0.01 % BME).

2.7.2.1 *His*-tagged protein purification -affinity chromatography (Capture)

Following the purification scheme (Figure 2.4), the affinity purification method was first used for capturing *His*-tagged protein. Clarified lysate was loaded onto HisTrap™FF crude or HisTrap™HP (5 mL column, GE Healthcare) which was pre-equilibrated with 5 column volumes (CV) of IMAC Binding buffer (Table 2.4), using an ÄKTA prime plus FPLC System (GE Healthcare), with a flow rate of 5 mL/min at RT. After loading, the column was

washed with 10 CV Binding Buffer to remove contaminants and unwanted proteins (Figure 2.5). Finally, *His*-tagged proteins were eluted gradually or stepwise to 100 % using IMAC Elution buffer (Table 2.4). Eluted fractions were collected, pooled, and stored at 4 °C. 1-5 µL samples of pre-load, wash, and elution fractions were added to 5 µL (4x) Sample Buffer of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), containing 2 µL (10x) Reducing agent (1 M DTT), and loaded in 4-12 % SDS-PAGE for gel analysis. The eluted fractions were pooled, concentrated, and buffer exchanged with suitable functional/storage buffer. A further purification of recombinant proteins by ion exchange chromatography and/or gel-filtration chromatography was performed as required and is detailed below.

2.7.2.2 Ion exchange chromatography (intermediate purification)

A pool of target protein (with pI > 7) was further purified by ion exchange chromatography (IEX) as an intermediate purification step by using 1 mL cation-IEX column of HiTrap™SP FF or HiTrap Capto™S (GE Healthcare), using an ÄKTA prime plus FPLC System (GE Healthcare). The column was equilibrated with IEX Binding buffer: 50 mM Sodium Acetate (pH 5.5) (Table 2.4). The pooled sample (from Section 2.7.2.1) was pre-diluted (1:2) or buffer exchanged in no-salt buffer before loading onto equilibrated column. Samples were loaded at a flow rate of 1 mL/min at RT and eluted using a linear gradient with IEX Elusion buffer: 50 mM Sodium Acetate (pH 5.5), 1.0 M NaCl, at 0–1 M NaCl run over 10 CV (10 mL). Purified protein factions were retained for SDS-PAGE analysis, and then were pooled and buffer exchanged with suitable storage buffer and stored at 4 °C.

2.7.2.3 Gel filtration chromatography (polishing)

A pool of protein fragments were further purified (as required) by gel filtration (GF) chromatography using Superdex™ 75 or Superdex™ 200 prep grade gel filtration column (GE Healthcare) using FPLC Systems: ÄKTA prime plus (GE Healthcare) or ÄKTA Explorer (Amersham Pharmacia). Protein samples and GF Running buffers (Table 2.4) were filtered prior to loading with 0.45 µm and 0.22 µm, respectively. The proteins obtained from either IMAC or IEX were first buffer exchanged using HiTrap™ Desalting (5 mL column, GE Healthcare) or HiPrep™ 26/10 Desalting (53 mL column, GE Healthcare) with same running

buffer and concentrated prior to loading. The SEC column was pre-equilibrated with 5 CV of the same buffer.

For hCpn10 and other variants, the column was pre-equilibrated and run in GF Running buffer I (100 mM Phosphate (pH 7.6), 150mM NaCl, 1 mM DTT) or GF Running buffer II (50 mM Tris (pH 8.0), 150mM NaCl, 1 mM DTT) at a flow rate of 1 mL/min at RT. For CE76 construct and its variants ($CE76_{P1}$, $CE76_{P2}$, and $CE76_{E5}$) the column was pre-equilibrated and run in GF Running buffer III (100 mM Sodium Acetate (pH 5.5), 150 mM NaCl, 1 mM DTT) at the same flow rate and condition. The purified protein fractions were collected separately and samples were retained for SDS-PAGE analysis. Protein fractions and/or pool were used for further investigations.

2.7.3 Recombinant proteins extraction, purification, and refolding (non-native)

2.7.3.1 Cell lysate and inclusion bodies preparations (under denaturing conditions)

For dissolving and extracting proteins from insoluble fractions, i.e. inclusion bodies (IB) (Figure 2.6), the cells were harvested by centrifugation at 5000xg and stored at -20 °C. The harvested cell pellets were thawed overnight on ice at 4 °C, and resuspended in Lysis buffer: 50 mM Tris (pH 7.6), 150 mM NaCl, 0.1 % Triton X-100, 1 µg/mL DNaseI, 1 mM DTT, 1 tablet/10mL cOmplete ULTRA (Table 2.4) in a ratio of 5:1 (5 mL/gm) as described in Section 2.7.1. Cell pellets were vortexed until they were completely dissolved or placed on platform rocker for 1-2 hrs. A combination with mechanical disruption methods (Table 2.8) was also applied in some instances. Soluble fractions were separated from insoluble factions by centrifugation at 5000xg for 10 min at 4 °C. Pellets were then washed with IB Washing buffer: 100 mM phosphate (pH 7.4), 1 M Urea, 1 % Triton X-100 (Table 2.4), and centrifuged at maximum speed of 16000xg for 30 min at 4 °C. Pellets were then resuspended in IB Denaturation buffer: 100 mM phosphate (pH 7.4), 150 mM NaCl, 8 M Urea, 2 mM DTT or 0.5 % BME (Table 2.4), with stirring or shaking for 1-2 hrs at RT, to dissolve the inclusion bodies, and were then clarified by centrifugation at maximum speed of 16000xg for 30 min at 4 °C. The clarified supernatants containing the denatured proteins were then ready to be loaded onto a column compatible with chaotropic agent (e.g. 8M Urea) for protein purifications. SDS-PAGE analysis was performed by mixing 1-5 µL of denatured lysate,

supernatant, and purified denatured protein fractions, with 5 µL Sample Buffer (4x), diluted with MilliQ water to 15 µL, heated at 70 °C for 10 min, and then loaded onto SDS-PAGE gels.

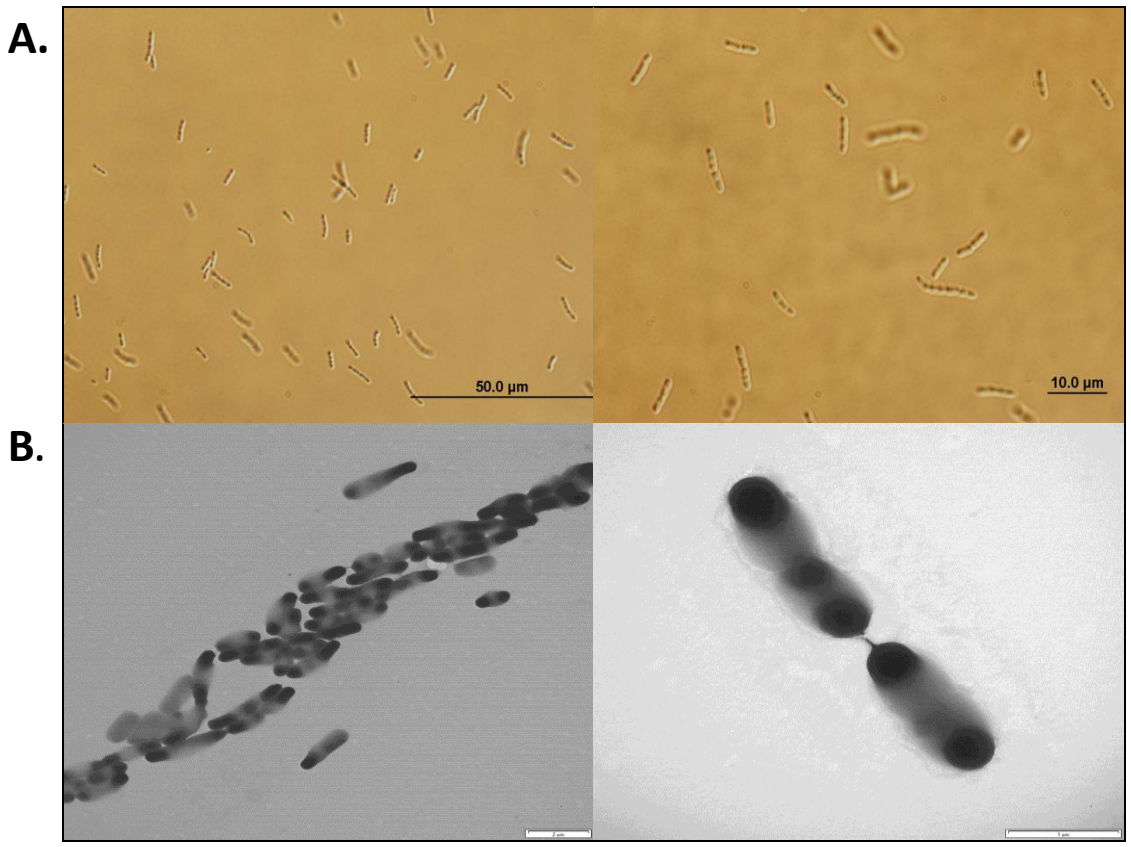


Figure 2.6 Inclusion bodies (IB) developed during protein expression of FH-CE76 in BL21 *E.coli.*

Images of the *E. coli* cells with IBs (dark aggregates) post-expression of FH-CE76. **(A)** *E. coli* cells examined under light microscopy, black bar shows scale of 50 and 10 µm, respectively. **(B)** *E. coli* cells examined under electronic microscopy, while bar shows scale of 2 and 1 µm, respectively.

2.7.3.2 *His*-tagged protein purification under denatured conditions (capture)

This affinity purification method is similar to Section 2.7.2.1 above with modification for *His*-tagged proteins in denatured conditions. The dissolved inclusion bodies containing the *His*-tagged FH-CE76 protein were loaded onto HisTrap™ FF crude or HisTrap™ FF (5 mL column, GE Healthcare), that was pre-equilibrated with 5 CV IMAC Binding buffer III: 100 mM Phosphate (pH 7.8), 150 mM NaCl, 0.05 % BME, 8 M Urea, 10 mM Imidazole (Table 2.4), with a flow rate of 1-5 mL/min, at RT. The column was washed with 5 CV of IMAC binding buffer III for removal of contaminants and unwanted proteins, followed by a stepwise

wash of 5 % of IMAC Elution Buffer III: 100 mM Phosphate, 150 mM NaCl, 0.05 % BME, 8 M Urea, 0.5 M Imidazole (Table 2.4). The *His*-tagged FH-CE76 fragments were eluted from the column (in 5 CV) by 100 % stepwise application of IMAC Elution buffer III (Table 2.4). Prior to protein refolding processes, the eluted fractions were collected, pooled, and stored at 4 °C. Samples (1-5 µL) of pre-load, wash, and elution fractions were mixed with 5 µL Sample buffer (4x), and then loaded into 4-12 % SDS-PAGE gels for analysis.

2.7.3.3 Refolding of denatured proteins

A quick screening test was performed to find the optimum buffering components and conditions. A set of multiple buffering systems with different salt concentrations, pH, reducing agents, non-ionic detergents, stabilizers, in a total of 190 µL/well were loaded into a 96-well flat plate. A 10 µL sample of dissolved protein (concentrated ≥1 mg/ml) was added, and set to shake in a platform rocker for 1-2 hrs at 4 °C. Protein aggregation was determined following the methods of Burgess (2009) and Tresaugues et al. (2004) by measuring the absorbance at 320 and 390 nm, respectively (Burgess, 2009; Tresaugues et al., 2004). Low absorbances suggested low levels of protein aggregation and determined the buffer of choice used for refolding.

The refolding of denatured proteins was performed in multiple dialysis steps in the optimal buffer (from above) at 4 °C. Proteins (from Section 2.7.3.2) were loaded into a SnakeSkin® Dialysis Tubing (Thermo™ Scientific), with 7 kDa or 10 kDa molecular weight cut-off (MWCO) and dialysed against 100 mM Phosphate (pH 7.6) or 50 mM Tris (pH 8.0), 100/150 mM NaCl, 5/10 % glycerol, 1 mM DTT, and 4M urea, for 3-4 hrs at 4 °C. The concentration of Urea was gradually reduced from dialysis buffer during each buffer exchange steps to final concentration of 0 M (from: 4, 2, 1 to 0 M Urea). Protein was kept in a fresh dialysis buffer (of 0 M Urea) overnight at 4 °C prior to polishing by GF purification (if required), as described in Section 2.7.2.3.

2.8 Protein analysis and characterisation

2.8.1 Detection of recombinant proteins by SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a standard method for detecting protein expression, purifications, degradations and the formation of multimers. 1-5 µL protein samples, mixed with 5 µL LDS Sample buffer (4x) NuPAGE®, containing 2 µL Reducing agent (10x) 1 M dithiothreitol (DTT) for reduced samples and/or 2 µL MilliQ water for nonreduced samples and MilliQ water were added to 15 µL. Samples were mixed and denatured by heating at 70 °C for 10 min or 90 °C for 5 min. For analysis, the denatured samples were loaded onto precast 4-12 % NuPAGE® Bis-Tris mini gels, electrophoresed for 35 min (or longer based on proteins size and required separation) at 200 V, run in NuPAGE® MES running buffer system form Invitrogen™. A 5 µL SeeBlue®Plus2 prestained standard (Invitrogen™) was run in parallel with samples for molecular weight determination. Following electrophoresis, the gel was washed 3x in MilliQ water for 5 min. Proteins were visualised by staining the gel in 50 mL SimplyBlue™ SafeStain (Coomassie®G-250, Invitrogen™) and placed on a platform rocker with gentle shaking for 1-2 hrs. Since SimplyBlue™ Safe Stain is a non-hazardous stain and does not require methanol or acetic acid fixatives, therefore, gels were only destained in MilliQ water for ≥ 2hrs or until protein bands were visible. The gel images were captured using a Bio-Rad Molecular Imager®Gel Doc™ XR^+ System and analysed by Image Lab™ ChemiDoc MP Software.

2.8.2 Detection of purified *His*-tag proteins by western blot

Post-electrophoresis SDS-PAGE gels were placed with a polyvinylidene fluoride (PVDF) transfer membrane, and covered in each side with three pieces of pre-soaked filter in Transfer buffer: (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS, 20 % (v/v) methanol). Protein bands were transferred to PVDF membrane, using a mini-cell apparatus (Invitrogen™) for 60 min, at 30 V at RT. The PVDF membrane was then incubated in 10 mL Blocking buffer PBST-M: (PBS; 137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4 (pH 7.4), 0.1 % (v/v) Tween-20, 5 % (w/v) skim milk) in orbital shaker for 1 hr at 4 °C. The membrane was washed (3x for 10 min) in PBST buffer: (PBS, 0.1 % (v/v) Tween-20), at RT. The washed membrane was incubated with 5 µL (1:5000 dilutions in 25 mL PBST) HRP-Mouse Anti-6xHis (50 µg/mL), for 1 hr at 25 °C in orbital shaker. The membrane was washed with PBST

(2x for 15 min) at RT, and 1x wash with PBS for 5 min. The membrane was then drained and target proteins were visualized using SuperSignal West Pico Chemiluminescent Substrate (Thermo™ Scientific) by incubating with 2 mL detection substrate for 3-5 min and then excess reagent was drained. Membrane was exposed to Kodak BioMax X-ray film (Kodak™) in the dark room and processed using Konica SRX-101A film processor (KonicaMinolta™ Healthcare) for 1, 2, 5, 10, 30, and 60 sec.

2.8.3 Protein quantification by spectrophotometric and BCA protein assays

Sample protein concentration was measured using a UV-visible NanoDrop1000 spectrophotometer (Thermo™ Scientific) with appropriate sample dilutions with same buffer used as blank. The concentration was measured at 280 nm absorbance according to the Beer-Lambert Equation (2.2) using extinction coefficients (ε) listed in Table 2.10. A is absorbance at 280 nm, ε is molar absorptivity (L. mol^{-1} cm^{-1}), l is path length (1 cm), C is concentration (mol. L^{-1}) or 100 mg/mL. The ε of proteins were calculated from amino acid sequences using ProtParam (online program) by ExPASy proteomics server (ExPASy, 2014b). The protein concentrations were then determined from rearranged Equations 2.3 and Equation 2.4.

$$A = \varepsilon\, l\, C \qquad \textbf{Equation 2.2}$$

$$\varepsilon\ 1\% = \frac{\varepsilon_{molar} \times 10}{Mwt\ (Da)} \qquad \textbf{Equation 2.3}$$

$$C\ (mg/mL) = \frac{A_{280}}{\varepsilon\ 1\%} \times 10 \qquad \textbf{Equation 2.4}$$

Table 2.9 Proteins molecular weight and molar extinction coefficients

Protein	Mwt (Da)	ε (M^{-1} cm^{-1})	Alternative ε (M^{-1} cm^{-1})
hCpn10	75626.3	31290	
ΔL-Cpn10	63930.0	31290	
ΔRoof-Cpn10	67238.2	31290	
ΔL-ΔRoof-Cpn10	57924.2	31290	
FH-Cpn10	76653.5	31290	
CE76	78901.0	80220	81095 for non-reduced *Cys*
$CE76_{P1}$	82250.9	80220	81095 for non-reduced *Cys*
$CE76_{P2}$	81620.3	80220	81095 for non-reduced *Cys*
$CE76_{E5}$	86399.1	80220	81095 for non-reduced *Cys*
FH- CE76	79140.8	80220	81095 for non-reduced *Cys*

Protein	Mwt (Da)	ε (M^{-1} cm^{-1})	Alternative ε (M^{-1} cm^{-1})
CPMyc	73014.2	31290	
$CPMyc_{P1}$	76364.0	31290	
$CP11_{-bFGF-P1}$	74535.4	31290	
$CP7_{-CD44-P1}$	77900.0	31290	
$CR2_{-VFGF-P2}$	84691.0	31290	

Protein concentrations were also measured at 562 nm for samples (from cell lysate, purification, refolding, concentration, etc), using the Pierce™ Bicinchoninic acid (BCA) assay (Pierce™ BCA Assay Kit, Thermo™ Scientific), in flat-bottom 96-well plate, following the manufacturers protocol. 20 µL samples diluted by a factor of 4 or in diluted net of 1:10 to 1:100 (to give absorbance values within assay range) in duplicate or triplicate was mixed with 180 µL BCA reagent mixture per well. 20 µL of bovine serum albumin (BSA) protein standard of known concentration (1 mg/mL) made in serial dilution by factor of 4 (in duplicate) was also mixed with 180 µL BCA reagent mixture per well, incubated at 37 °C for 30 min. Samples were read by UV-Vis SpectraMax Plus^{384} Spectrophotometer plate reader at A_{562}. Sample concentrations were then interpolated from BSA standard curve (Figure 2.7).

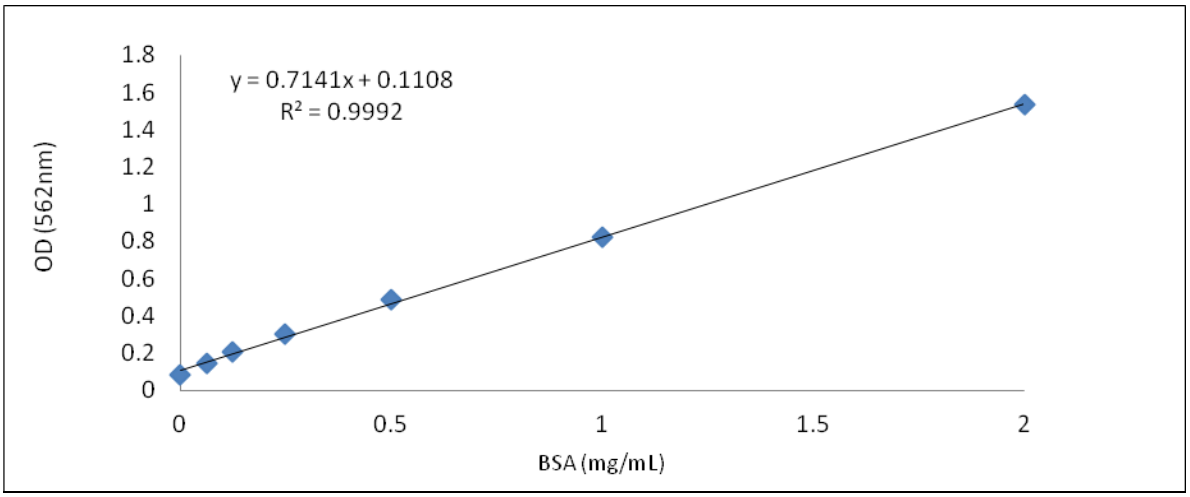


Figure 2.7 BSA standard curve

The standard curve of BSA used to determine unknown protein concentrations using Pierce™ BCA assay in 96-well plate after incubation for 30 min in 37 °C, and then measured at 562 nm.

2.8.4 Identification of protein by MALDI-TOF MS

Verification of correct protein expressions and purifications, sample purity and protein primary molecular weights in dalton (Da) were determined by using matrix assisted laser desorption ionization (MALDI) coupled to time of flight (TOF) mass spectrometer (MS) for

accurate mass measurement analyses. The peptide fingerprints used to determine structural information and protein maps were generated by injecting samples onto a Prominence Nano-LC system (Shimadzu™ Scientific) using a flow rate of 30 µL/min and desalted on an Agilent C18 trap (0.3 x 5 mm, 5 um, Agilent™ Technologies) for 3 min, followed by separation on a Vydac® Everest® C18 (300 A, 5um, 150 mm x 150 um) column at a flow rate of 1 µL/min. A gradient of 10-60 % of mobile-phase B (80 % acetonitrile (AcN)/ 0.1 % Formic acid (FA)) for 30 min over mobile-phase A (1 % AcN/ 0.1 % FA) was used to separate peptides. Eluted peptides were directly analysed on a Triple TOF® 5600 system (ABSciex™) using a NanoSpray® III interface (ABSciex™). Gas and voltage settings were adjusted as required. MS TOF scan across m/z 350-1800 was performed for 0.5 sec followed by information dependent acquisition of the top 20 peptides across m/z 50-1800 per second.

Data was converted to mgf format and searched in MASCOT accessed via the Australian Proteomics Computational Facility (APCF). Data was searched in SwissProt against all species using trypsin as enzyme, mis-cleavages, MS tolerance of 0.5 Da, MS/MS tolerance of 0.2 Da. Carbamidomethylation (cys, fixed) and oxidation (met, variable) modifications were also included.

Molecular weight analyses of protein samples were also characterized and determined by MALDI-TOF MS. The intact protein mass was fitted exactly with the theoretical Mwt calculated from the amino acid sequence using ExPASy proteomics server (ExPASy, 2014b), or within ±1-2 Da variations due to the intact protein molecule modifications and/or heterogeneity.

2.8.5 Characterisation of protein molecular mass and native configuration by SE-HPLC

Sample molecular weights in their native state (in full homogeneity), and soluble aggregates were further investigated using analytical size exclusion (SE) high performance liquid chromatography (HPLC) with a TSK column (TSK G3000SWXL, Tosoh™ Bioscience) and Agilent 1200 Series HPLC System (Agilent™ Technologies). Protein samples and all SE Running buffers (Table 2.6) were filtered (0.22 µm) prior to loading onto TSK column. A guard column (Guard column (7.8 mm), Tosoh™ Bioscience) was placed ahead of the SE-HPLC TSK column and the whole system was pre-equilibrated with SE mobile phase I: 0.1

M Phosphate buffer (pH 7.8), 0.2 M sodium chloride, or by SE mobile phase II: 100 mM Sodium Acetate (pH 5.5), 150 mM sodium chloride. A maximum of 100 µL sample/injection (total of 1 mg/mL) were analysed at a flow rate of 0.8 mL/min (≤ 70 bar), at 25 °C. A 10 µL BioRad™ Gel filtration standard (Bio-Rad™ Laboratories), was applied for determining apparent molecular weight of target recombinant proteins from a standard curve of proteins of known size (Figure 2.8).

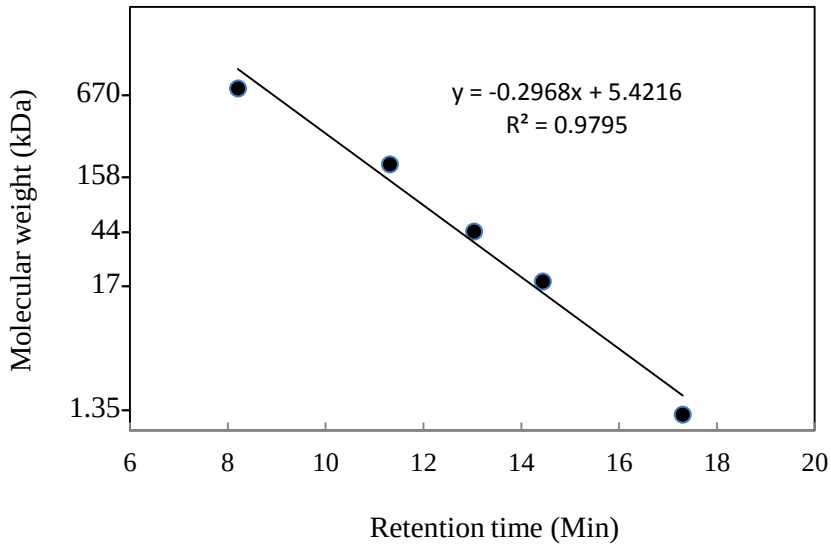


Figure 2.8 Determination of apparent molecular mass of unknown proteins

Determination of apparent molecular mass of known Mwt proteins by SE-HPLC using TSK-G3000SWXL column, used to calculated apparent molecular mass of native hCpn10 and its variants by comparing with the migration length of reference proteins. The column was calibrated with BioRad™ reference proteins: Thyroglobulin (670 kDa); Ovalbumin (44 kDa); γ-Globulin (158 kDa); Myoglobin (17 kDa); and Vitamin B12 (1.35 kDa); respectively.

2.8.6 Determination of storage buffer and protein stability by spectrophotometry

A screening test was performed to find optimum buffering components and conditions for each protein constructs (Table 2.2). This was done by adding a 10 µL sample of purified and dissolved protein (concentrated ≥ 1 mg/ml) to a set of 190 µL/well of multiple buffering systems (Tables 2.4 and 2.5) with different concentrations of salts (NaCl, KCl), pH, reducing agents (DTT, BME), non-ionic detergents (Trition X-100), stabilizers (glycerol), loaded into 96-well flat plate, and set to shake a platform rocker for 1-2 hrs, at 4 °C. Protein aggregates were then determined by following the methods of Burgess (2009) and Tresaugues et al.

(2004) by measuring the absorbance at 320 and 390 nm, respectively (Burgess, 2009; Tresaugues, et al., 2004), using UV-Vis SpectraMax Plus^{384} Spectrophotometer. Buffers with low protein aggregates (determined by low UV absorbance) were selected as the appropriate buffer system with additive concentrations used for buffer exchange and protein storage.

Protein solubilities, stabilities, and degradations were also illustrated by SE-HPLC (similar to Section 2.8.5), whereby protein samples were monitored in multiple storage conditions at (-20, 4, 25, and 37 °C) over different periods of time: (day, week, and month). In brief, a ≤ 100 µL sample (equivalent to 1 mg/mL) with the same SE mobile phase (Table 2.4) were injected onto a SE-TSK column (Tosoh™ Bioscience) at a flow rate of 0.8 mL/min (≤ 70 bar), was heated at 25 °C. A molecular weight standard was run in parallel by injecting 10 µL of BioRad™ GF standard, and absorbance measured at 280 nm using Agilent™ 1200 Series HPLC System. Stacked comparison of SE-HPLC chromatograms was used to determine changes for each sample.

Chapter 3. Designing a Novel Molecular Scaffold based on Chaperonin 10: Experimental and Computational Approaches

3.1 Introduction

Binding entities based on molecular scaffolds will have increasing application as therapeutic and diagnostic agents in the future. There are more than ten non-antibody-based products based on molecular scaffolds currently in clinical trials, and there is at least one that has been approved for therapeutic and/or diagnostic applications (Chapter 1). Non-antibody molecular scaffolds can be modified where a particular region of the molecule, typically a flexible loop, can be replaced with a polypeptide that will bind to a particular antigen. In essence the scaffold has been modified in a manner that mimics the binding functionality of conventional mAbs. However the antibody scaffold is the product of an evolutionary process that enables varying peptide sequence to be inserted within the CDRs without impacting the structural integrity and the correct folding of the antibody molecule. In the case where a peptide loop with specific functionality within a molecular scaffold is replaced with a non-native sequence, optimal display of the peptide loop and/or structural integrity (i.e. folding and three dimensional conformation) of the molecular scaffold cannot be guaranteed. The research described in this chapter is associated with experimental and computational approaches to assessing the impact of various hCpn10 mobile loop substitutions on the structural integrity of the hCpn10 molecular scaffold. Strategies to overcome an unfavourable impact on stability as a result of peptide substitution are described.

Human chaperonin 10 (hCpn10), otherwise known as Heat shock protein 10 (Hsp10), is an essential mitochondrial chaperone protein that assists in the folding of post-translational polypeptides (Hartman et al., 1992, Pilkington and Walker, 1993, Martin et al., 1993, Hartl and Hayer-Hartl, 2002). As shown in Figures 3.1 and 3.2, Cpn10 is structurally conserved across many species despite minor sequence homology (Hunt et al., 1996, Xu et al., 1997, Roberts et al., 1999, Numoto et al., 2005, Luke et al., 2005b, Luke et al., 2005a). It also has the capacity to form a heptameric ring under physiological conditions (Higurashi et al., 2003, Sakane et al., 2004, Perham and Wittung-Stafshede, 2007). The Cpn10 heptamer has seven identical monomers (10 kDa) that consist of a hydrophobic β-barrel core, a β-hairpin roof

loop and a mobile loop extending below the core of the molecule (Landry et al., 1997). Cpn10 is a very stable structure. A study by Sakane et al (2004) demonstrated that the introduction of *Ala*-mutations of *Ile3* and *Val95* residues (i.e. central residues within anti-parallel β-strand pairing) did not affect structural stability. The unfolding-refolding characteristics of hCpn10 are similar to GroES (i.e. Cpn10 derived from *E. coli*) after denaturation using various concentrations of guanidine hydrochloride (GdnHCl) at 25 °C (Sakane et al., 2004).

The mobile loop within Cpn10 monomers is highly flexible and exhibits a variety of structural conformations due to the lack of a defined secondary structure (Richardson et al., 2001, Lin et al., 2004). However this mobility is restricted when the Cpn10 heptamer interacts with its cognate partner Cpn60/GroEL (Richardson et al., 2001). Similar to other molecular scaffolds (as discussed in Chapter 1), the mobile loop forms an extended interface that can bind to the target molecule. The mobile loop of hCpn10 consists of 18 residues, as described in Table 3.1. In this study, the removal of the mobile loop (ΔL-Cpn10), the removal of β-hairpin roof loop (ΔRf-Cpn10) or the removal of both regions ΔRf-ΔL-Cpn10 (i.e. leaving an isolated β-barrel) did not affect the assembly of hCpn10 into a heptamer.

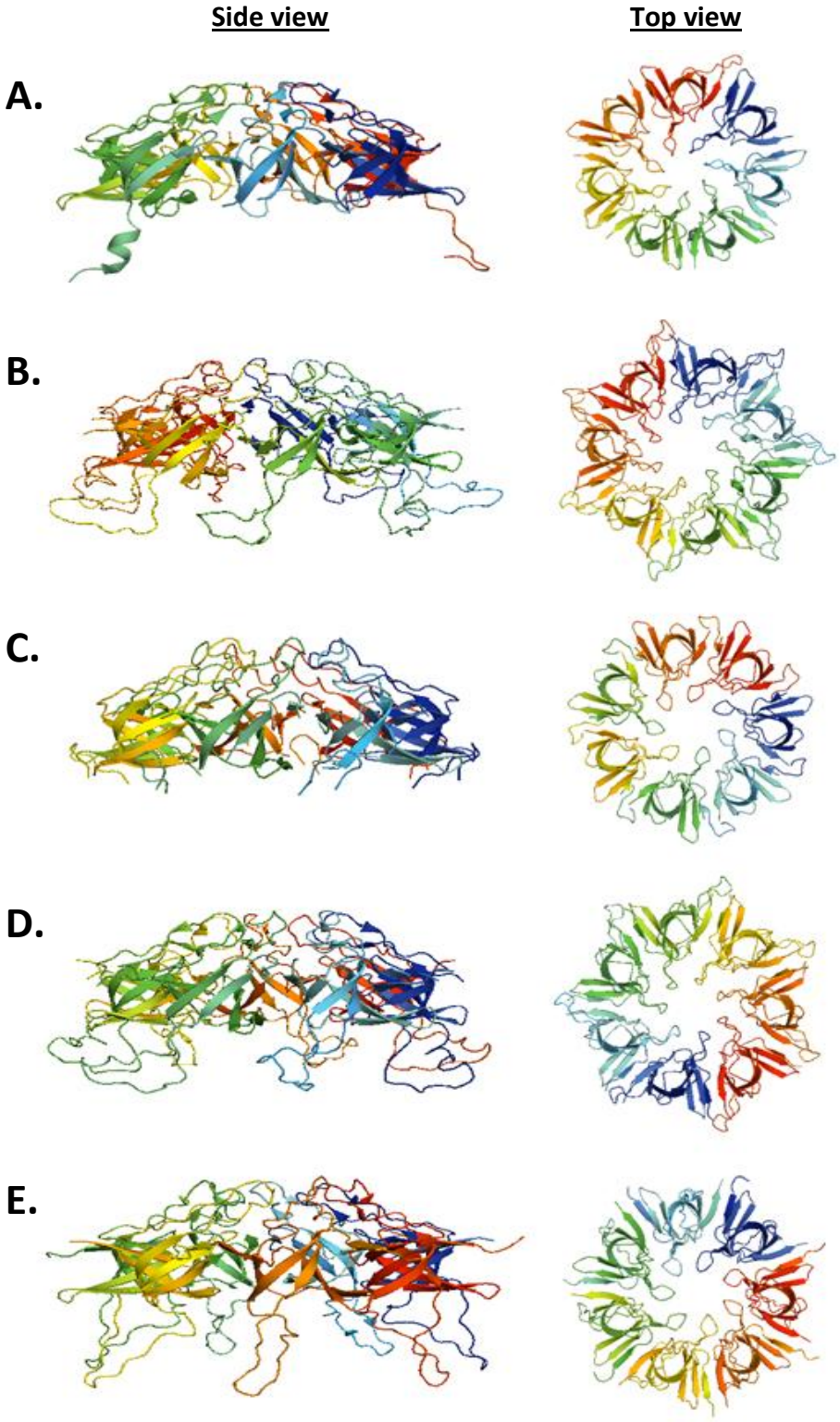
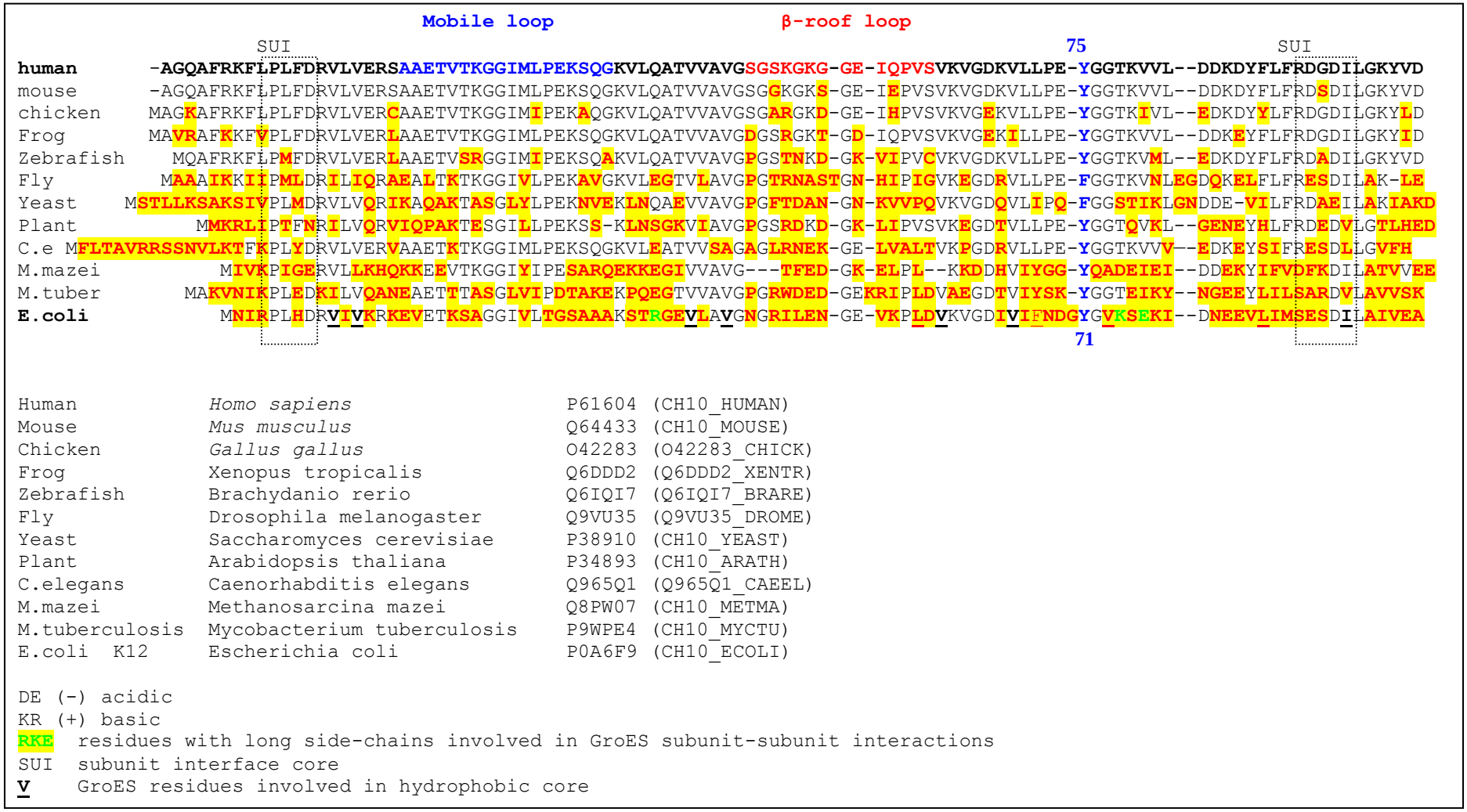


Figure 3.1 Three-dimensional representation of Cpn10 from different species

3D representations of protein backbones (side and top view) of Cpn10 from various species: **(A)** *Thermus thermophilus* (PDB ID. 1WNR); **(B)** *Escherichia coli* BL21 (PDB ID. 3ZPZ); **(C)** *Mycobacterium tuberculosis* (PDB ID. 1HX5); **(D)** *Escherichia coli* K-12 (PDB ID. 3WVL); and **(E)** *Mycobacterium tuberculosis* in tetradecamer (PDB ID 1P3H). All protein structures shown in ribbon representations and modelled by PyMOL using crystallographic solved structures available in RCSB Protein Data Bank (RCSB-PDB, 2010).



Human	*Homo sapiens*	P61604 (CH10_HUMAN)
Mouse	*Mus musculus*	Q64433 (CH10_MOUSE)
Chicken	*Gallus gallus*	O42283 (O42283_CHICK)
Frog	Xenopus tropicalis	Q6DDD2 (Q6DDD2_XENTR)
Zebrafish	Brachydanio rerio	Q6IQI7 (Q6IQI7_BRARE)
Fly	Drosophila melanogaster	Q9VU35 (Q9VU35_DROME)
Yeast	Saccharomyces cerevisiae	P38910 (CH10_YEAST)
Plant	Arabidopsis thaliana	P34893 (CH10_ARATH)
C.elegans	Caenorhabditis elegans	Q965Q1 (Q965Q1_CAEEL)
M.mazei	Methanosarcina mazei	Q8PW07 (CH10_METMA)
M.tuberculosis	Mycobacterium tuberculosis	P9WPE4 (CH10_MYCTU)
E.coli K12	Escherichia coli	P0A6F9 (CH10_ECOLI)

DE (-) acidic
KR (+) basic
RKE residues with long side-chains involved in GroES subunit-subunit interactions
SUI subunit interface core
V GroES residues involved in hydrophobic core

Figure 3.2 Sequence alignment of Cpn10 derived from various species

Sequence data illustrates an approximate 30 % homology of human Cpn10/Hsp10 and *E.* coli GroES, Cpn10 sequence alignment was obtained from UniRef/UniPotKB provided by UniPort online database (UniProt, 2010).

Table 3.1 shows the peptides used as substitutes for the native mobile loop of hCpn10 in this study. They are further characterised based on their sequence, length, isoelectric point (pI), target and biological activity.

Table 3.1 Peptides as substitutes for native mobile loop of human Cpn10

Peptide	Loop sequence	#AA	pI	Target	Biological activity	Reference(s)
hCpn10	*AAETVTKGGIMLPEKSQG*	18	8.91	**Cpn60**	- Protein folding	(Landry et al., 1997)
ΔL-Cpn10	*AG*	2	8.91	**--**	n/a	--
E76	*ALCDDPRVDRWYCQFVEG*	18	**7.21**	**FVII/ FVIIa**	- Anticoagulant	(Dennis et al., 2000)
Myc	*EQKLISEEDL*	10	**6.65**	**9E10**	- Protein localization, - ELISA, - Protein purification	(Hilpert et al., 2001)
P7	*PFNLPLPSRPLLR*	13	9.52	**CD44**	- Diagnostic probe	(Park et al., 2012)
P11	*PLLQATLGGGS*	11	8.98	**bFGF**	- Anti-Angiogenic	(Wu et al., 2010)
VEGF1	*ITMQIMRIKPHQGQHIGEMSF*	21	8.76	**EGFR/ HER1**	- Anti-Angiogenic	(Vicari et al., 2011)

Based on the flexibility of the mobile loop region and the observation that deletion of mobile loop does not affect hCpn10 assembly, this study investigated the use of hCpn10 as a molecular scaffold with the aim of developing NMEs capable of binding to targets through the insertion of various functional polypeptide domains at the mobile loop site (Table 3.1). Substitution with different peptides may compromise the assembly of hCpn10 into a heptamer through molecular interactions of amino acids within the mobile loop with those at the interface of the β-barrel that mediate heptamer assembly. For example there may be physicochemical interactions between loop and β-barrel charged amino acids (salt bridges) or hydrophobic residues. Therefore, molecular simulations were conducted in order to investigate the dynamics of assembly of hCpn10 heptamer with different peptides and to investigate how disruption of heptamer formation through loop peptide steric hindrance could occur. Protein engineering strategies that might avoid loop peptide, β-barrel interface interactions were also investigated to prevent aggregation and ensure that hCpn10 variant forms stable heptamers.

Traditional methods of assessing protein stability and 3D structure include X-ray crystallography (Roberts et al., 1999, Higurashi et al., 2003, Tereshko et al., 2008, Nandi et al., 2012), nuclear magnetic resonance (NMR) (Wu et al., 2012) and other biophysical methods (Buehler et al., 2010). However, there are limitations to these techniques since sample preparation (i.e. preparing refractable protein crystals for crystallography) is time consuming and needs to be of high purity (Tereshko et al., 2008). Molecular modelling is an emerging discipline where state-of-art computers are used to study the theoretical molecular dynamics of polypeptides (Leach, 2001) in a range of successive time frames between 10^{-15} and 10^{-6} seconds (Petrenko and Meller, 2010). Molecular dynamics (MD) simulations are expeditious approaches, and can yield valuable data on the supposed interactions between protein domains/amino acids within a protein structure.

3.2 Rationale for the project

Based on its known properties, hCpn10 has the potential to serve as a molecular scaffold where the mobile loop can be replaced with non-native peptides to mimic the functional binding properties of mAbs as shown in Table 3.1. MD simulations were conducted on hCpn10 in order to investigate, on a theoretical basis, the effect of native mobile loop substitution with unrelated peptides upon heptamer formation. This research focused on investigating molecular structure perturbations of hCpn10 due to mobile loop substitution via MD simulations and investigates strategies to prevent or minimise peptide loop molecular interactions with the β-barrel interface of hCpn10, so that heptamer formation is conserved.

3.3 Research aims

The aims of the research presented in this chapter are:

- To investigate the potential of hCpn10 and its variants to form heptamers, using computational and experimental approaches;
- To investigate the physicochemical interactions and interfaces on the molecular level, by using MD simulations;
- To apply the results of MD simulations to improve the molecular stability and efficiency of hCpn10 variants; and
- To establish a fundamental molecular design for hCpn10 as a molecular scaffold.

3.4 Materials and methods

3.4.1 Protein expression, purification and characterisation

All protein variants in this study were produced, purified (≥ 95 %), and characterised accordingly to the general materials and methods described in Chapter 2.

3.4.2 Homology modelling

Homology modelling of hCpn10 variants was performed in conjunction with Computist Bio-Nanotech (Melbourne, Australia). Structural homologues based on sequences defined in this study (Table 1.3) were created, based on available template X-ray crystal structure of the chaperonin complex from *Thermus thermophilus* (PDB ID. 1WE3). The overall folding topology of chaperonin 10 is highly conserved, therefore it was highly feasible that homology modelling of hCpn10 variants could be performed based on the known X-ray structure of the related *T. thermophilus* Cpn10 structure of 1WE3.

The conformations of the inserted sequences replacing the native mobile loop were modelled as close as possible to the Cpn10 conformation found in 1WE3. This conformation is a free energy minimum for heptameric hCpn10, while the conformation is likely not a global energy minimum of hCpn10 variants (i.e. mutants) structure, but only a local minimum that might not be stable. In order to find more stable and energetically favourable conformations, MD simulations to sample the accessible conformational space were performed. The *N*-terminus modelled as extended chains similarly to the template structure. The histidine residues in the *His*-tag (of the *C*-terminus) were simulated as uncharged side chains. The inserted E-76 peptide (Table 3.1) was simulated with and without an intact disulphide bond (*Cys23-Cys33)*.

All models showed heptameric structure of the proteins, however, for the structural investigation only one monomer was chosen, that is chain "U" of 1WE3 heptamer. This remainder was not part of the actual simulation and figures were prepared after the simulation by superimposing the simulated monomer "U" onto the structure heptamer.

3.4.3 Molecular dynamics (MD) simulations

MD simulations in this study were performed in conjunction with Computist Bio-Nanotech and were based on the NAMD2 program (Kale et al., 1999) combined with the CHARMM27 force field (MacKerell et al., 1998). Generalised born implicit solvent (GBIS) method (Tanner et al., 2011) was applied to increase the conformational sampling time. Approximate dielectric effect of water treated outside of the protein as dielectric continuum, solves the Poisson Boltzmann equation (Fogolari et al., 2002) in order to calculate additional forces from solvent to protein. The van der Waals packing interactions were disregarded in NAMD2's implementation of the method.

The monomeric form of hCpn10 protein has about 11,000 atoms and the protein is surrounded by a 13-15 Å thick shell of water molecules at a density of 1 g/cm^3. The simulations of the monomers were performed without explicit water. However, the solvent effects were taken into account by using an implicit water model of GBIS method in 1 nano-second (ns) (Tanner et al., 2011, Onufriev, 2010).

A 10 ns simulation of the hCpn10 sequence was performed to test the stability and optimise the *N*-terminus of each monomer and heptamer. However, the stability of the loop conformations was evaluated in 1 ns duration.

Primarily, all simulations were performed twice with different initial condition: (1) simulation with random initial velocities selected according to a Boltzman distribution at 250 K, (2) simulation with random initial velocities at higher temperature at 310 K. The temperature was kept constant using "temperature scaling" method, and the entire system is set to the desired temperature of all velocities were periodically rescaled every 1 pico-second (ps).

Subsequently, in addition to the methods described previously, three MD simulations for 1 ns were performed for each of modified sequences of CE76 and CPMyc variants (Table 3.1): (1) simulation in an aqueous environment (ε=78.5), (2) simulation with the *C*-terminus *His*-tag sequence removed, and (3) simulation in cellular-like environment (ε=40).

The conformational changes that occurred during the MD simulations of examined structures, were reflected as trajectory changes and were processed according to specified geometric descriptors of VMD program (Humphrey et al., 1996). The variations of these trajectories were visually examined.

3.5 Results

3.5.1 Mutation approaches

Cpn10 is structurally conserved across several species despite the differences in amino acid sequences (Figure 3.1). This study focussed on hCpn10 as a potential molecular scaffold of human origin (Figure 3.3). As shown in Figure 3.4, the preliminary studies conducted assessed the stability of hCpn10 involved in the development of three mutant hCpn10 constructs, namely: ΔL-Cpn10 (i.e. mobile loop deletion), ΔRf-Cpn10 (i.e. roof loop deletion) and β-barrel (i.e. ΔRf-ΔL-Cpn10). Analytical size-exclusion HPLC revealed that the heptameric assembly of these hCpn10 mutants is not compromised despite these deletions and the apparent change in molecular weights (Figure 3.5). It can therefore be concluded from this data that the amino-acid residues of the hCpn10 loop regions are not critical for the assembly of heptameric Cpn10.

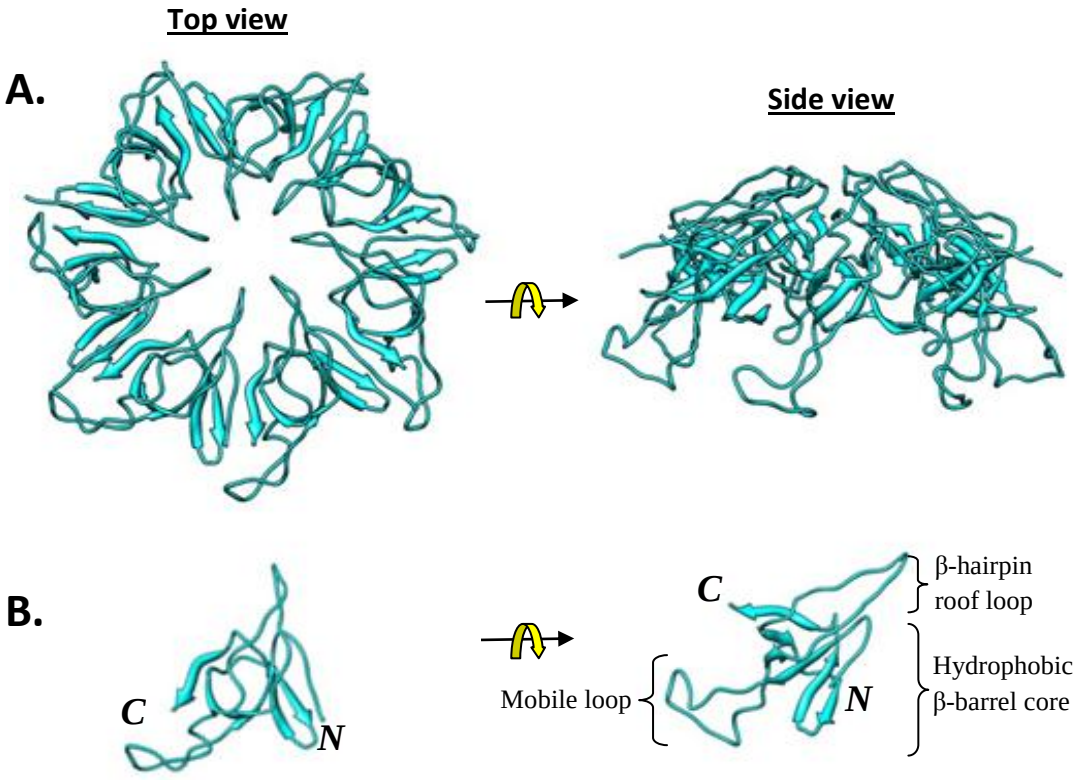


Figure 3.3 Three dimensional model of human Cpn10

hCpn10 structure modelled based on solved structure of *Thermus Thermophilus* (PDB ID. 1WE3). **(A)** Top vs. side view of heptamer hCpn10; and **(B)** Top vs. side view of monomer hCpn10. The structure is shown in ribbon representations and modelled by PyMOL.

Construct	AA Sequences	Size
wt Cpn10	AGQAFRKFLPLFDRVLVERSAAETVTKGGIMLPEKSQGKVLQATVVAVGSGSKGKGGEIQPVSVKVGDKVLLPEYGGTKVVLDDKDYFLFRDGDILGKYVD	101
ΔRf-Cpn10	AGQAFRKFLPLFDRVLVERSAAETVTKGGIMLPEKSQGKVLQATVVAVGGG----------VKVGDKVLLPEYGGTKVVLDDKDYFLFRDGDILGKYVD	89
β-barrel	AGQAFRKFLPLFDRVLVERSAGG----------KVLQATVVAVGGG---------PVSVKVGDKVLLPEYGGTKVVLDDKDYFLFRDGDILGKYVD	77
ΔL-Cpn10	AGQAFRKFLPLFDRVLVERSAG----------KVLQATVVAVGSGSKGKGGEIQPVSVKVGDKVLLPEYGGTKVVLDDKDYFLFRDGDILGKYVD	85
	********** ********** **********	

Figure 3.4 Sequence alignments of hCpn10 and its variants

Amino acid sequence and size of hCpn10 (wt) versus its variants; ΔRf-Cpn10, β-barrel and ΔL-Cpn10, illustrating mutation sites (i.e. mobile loop in blue and roof loop in red).

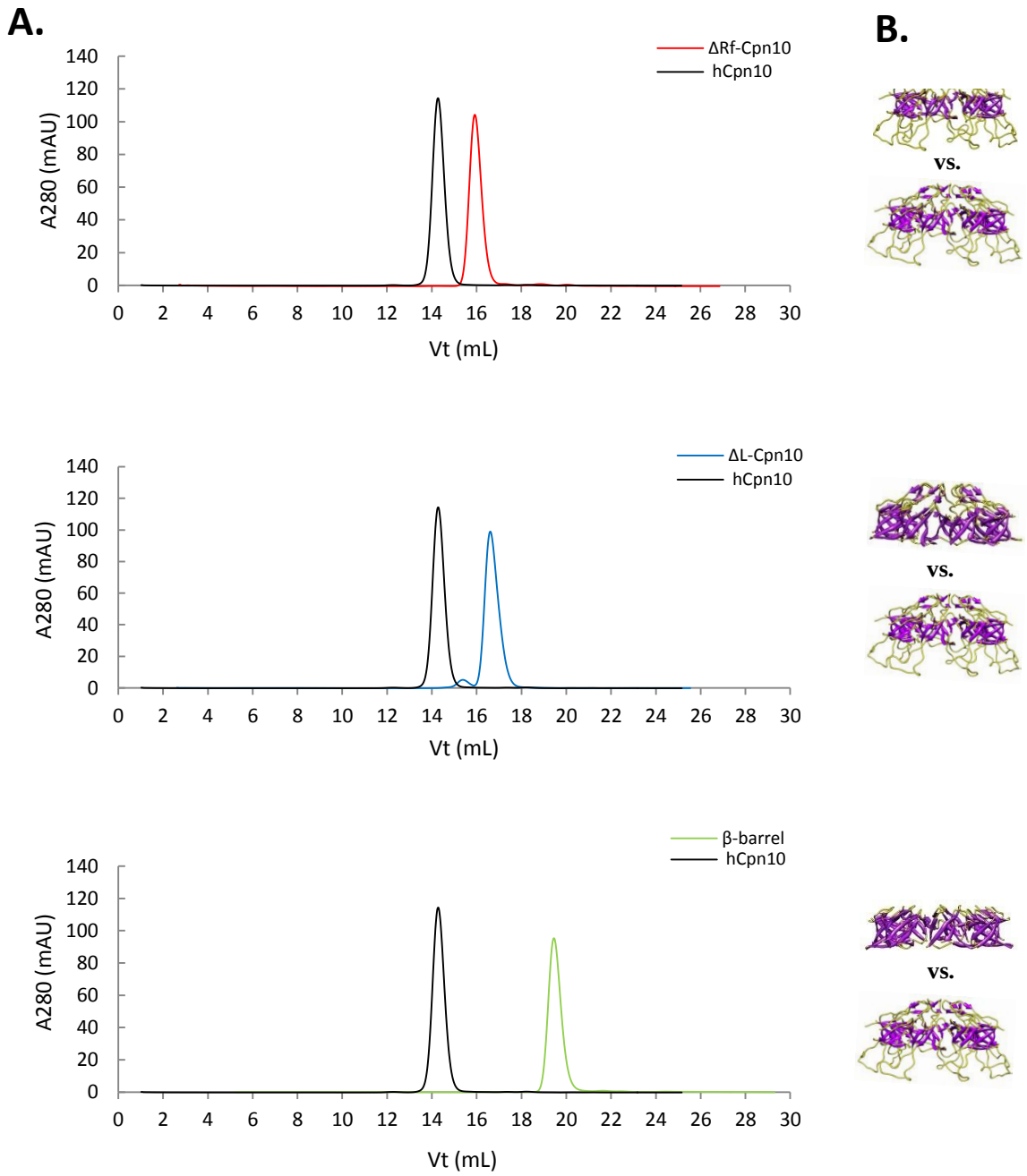


Figure 3.5 HPLC-size exclusion chromatograms of hCpn10 versus its mutants

Comparison of hCpn10 with its mutants using size-exclusion HPLC and UCSF Chimera modelling. **(A)** Absorbance at 280 nm of hCpn10 (black, theoretical molecular weight 70 kDa) versus ΔRf-Cpn10 (red, 62.3 kDa); ΔL-Cpn10 (blue, 59.5 kDa); and β-barrel only (green, 51.8k Da), respectively. All constructs apparently developed heptamers with various molecular sizes resulting from incorporation of molecular weight standards curve with known elusion retention times (detailed Chapter 2). **(B)** The side view of 3D model of the quaternary structures of each hCpn10 variants shown in ribbon representations (i.e. β-barrel in violet; & coil in yellow) modelled by UCSF Chimera 1.5.3.

3.5.2 Molecular modelling approaches

The schematic shown in Figure 3.6 outlines the integration of the experimental and computational (i.e. MD simulations) methodologies in order to design NMEs based on the hCpn10. MD simulations were conducted on multiple phases, or until the optimal model of hCpn10 variants could be developed and resembled into a native-like configuration. Although the polypeptide sequence of Cpn10 varies across all species, the protein is structurally conserved (Figure 3.1), therefore, three-dimensional modelling of hCpn10 in this study was based on the previous crystallographic solved structure of Cpn10 derived from the gram-negative *Thermus thermophilus* eubacterium species PDB ID. 1WE3 (Shimamura et al., 2004).

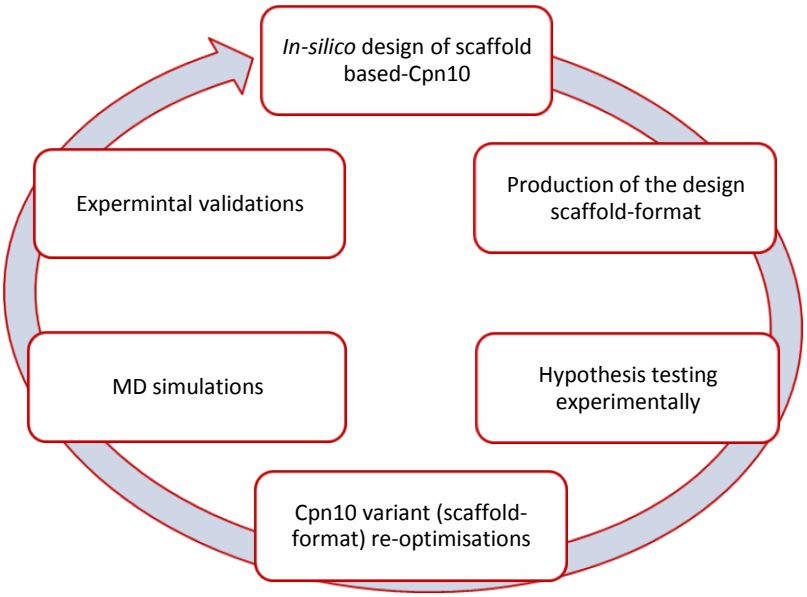


Figure 3.6 An iterative guide and experimental cycle for the design and creation of NMEs based on the hCpn10 molecular scaffold

3.5.2.1 MD simulation strategies

Homology models of all hCpn10 variants sequences defined in Table 3.1 were created based on the template crystal structure of the chaperonin-complex from 1WE3. Although the conformation for heptameric hCpn10 demonstrated a free energy minimum (Figure 3.3), this was not observed in mutant hCpn10 structures (Table 3.1), despite the insertion of sequences within the mobile loop that closely resembled to the X-ray structure of the chaperonin

complex of 1WE3. Therefore, in order to find more stable and energetically favourable conformations, molecular MD simulations were performed in order to sample the accessible conformational space of hCpn10 vairants.

Initially, a 10 ns MD simulation of the hCpn10 monomer was performed in order to test the stability of the protein followed by optimization of the *N*-terminus in both monomeric and heptameric conformations with the result shown in Figure 3.7. Monomeric and heptameric hCpn10 had an overall degree of stability although some flexibility was observed in the β-hairpin (roof-loop) and mobile loop (Figure 3.7, Appendix 3). In addition, the *N*-termini of each monomer, which is eleven residues in length (*AGQAFRKFLPL*), remained in an extended conformation. Since the extended *N*-terminus did not appear to affect the assembly of hCpn10 into an oligomer, it is possible to reduce the length of the *N*-terminus by up to nine residues.

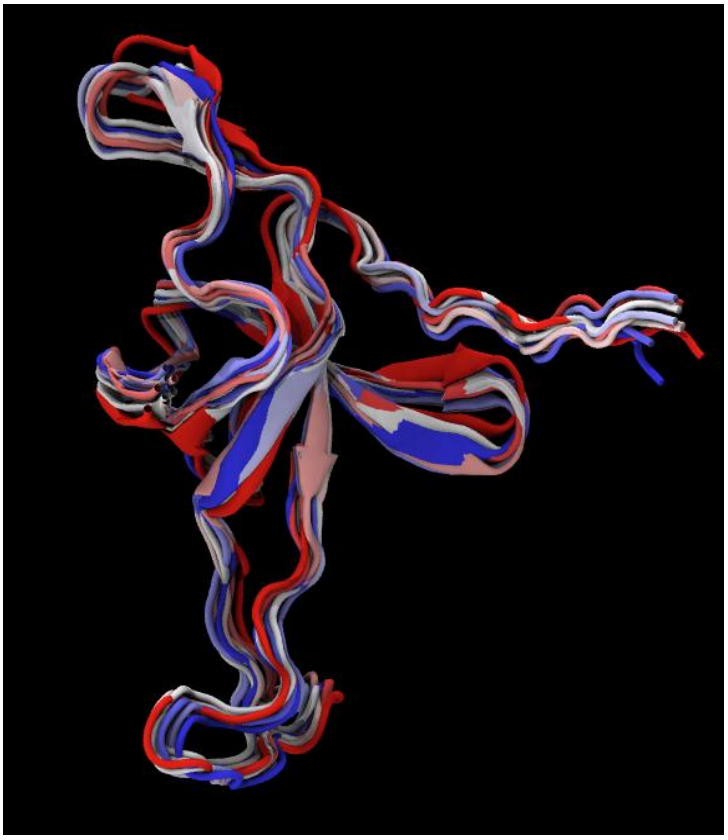


Figure 3.7 MD simulations of hCpn10 monomer over 1 ns duration

Snapshots of hCpn10 shown in ribbon representations of the last 700 ps (i.e. from initial: red to final: blue) at intervals of 50 ps and subsequently overlaid.

MD simulations for hCpn10 molecules were performed twice with different initial conditions. In the first simulation, random initial velocities were selected according to a Boltzman distribution at a temperature of 250 K. The first simulation was an enhanced energy minimisation in order to find a more stable conformation. The second simulation was conducted at a higher temperature of 310 K (i.e. above room temperature for escalated process) where the increased kinetic energy of the atomic structure of hCpn10 allowed for a better sampling of the various structural changes of the protein conformations. In both simulations, the temperatures were kept constant using a temperature scaling method where all of the atom velocities in the system were periodically rescaled every 1 ps so that the entire system was adjusted to the set temperature.

In essence, the heptameric formation of a variant hCpn10 is primarily dependent on the successful dimerization between adjacent monomers. However, in order for this to occur, the two adjacent monomers need to have a complementary surface shape and an electrostatic potential distribution that is conducive for dimerization. If the conditions are unfavourable, the binding strength of the dimer and its ability to form a self-assembling heptamer will be reduced (e.g. in the case of hCpn10 variant CE76, Appendix 3).

A general strategy was applied in order to evaluate whether or not hCpn10 and its variants could develop a stable heptamer-like conformation. Explicit treatment of water molecules helped to assess the effects of the solvent and various ions on the hCpn10 during MD simulations. In essence, the hCpn10 was surrounded by a 13-15 Å shell of water molecules. However, this led to a significant increase in the number of atoms in this simulation. The hCpn10 monomer is composed of roughly 11,000 atoms whereas the heptameric form of this protein has an estimated of 109,000 atoms. This made it difficult to simulate the hCpn10 heptamer as the amount of CPU power that is required to calculate the motion of bulk molecules is very high and subsequently time consuming. Furthermore, the assessment of each protein configuration such as the motion of the flexible loops was hindered as this requires a significant displacement of the adjacent water molecules. Nonetheless, the advantage of using water to assess the dynamics of a particular molecule yields more accurate and realistic results as a consequence of the electrostatic interactions due to the inherent polar nature of the water molecules as well as the van der Waals interactions between water and protein atoms.

In order to speed up the simulations and allow a faster conformational search, MD simulations of the monomers were also performed without using explicit water. In addition, the solvent effects were also taken into consideration by using an implicit water model of the GBIS method (Tanner et al., 2011, Onufriev, 2010, Chen et al., 2008). The GBIS is a method that does not use explicit water molecules or ions for molecular simulation, which subsequently increases the sampling speed. For hCpn10 and its variants, a simulation of approximately 1 ns with the GBIS method was performed which correspond to a much longer conformational sampling time for protein structures than in a simulation with explicit water. Furthermore, GBIS method utilises the Poisson Boltzmann equation as it approximates the dielectric (electrostatic) effect of water and treats the outside of the protein as a dielectric continuum (Fogolari et al., 2002, Moreira et al., 2005). Moreover, calculations of any additional forces from the solvent to the protein are resolved using the NAMD2 program (Kale et al., 1999, Phillips et al., 2005). However, the van der Waals packing interactions were disregarded in NAMD2's implementation of the method (Guvench and Alexander D. MacKerell, 2008). This approximation, on most occasions, is quite reliable as evident from several conducted experiments. Thus, the GBIS method was chosen, as more CPU time was required to see the effects on the mobile loops of the newly inserted functional peptides from Table 3.1. Consequently, it is important to look at the effect of replacing the mobile loop with a functional peptide and how this will affect the dimerisation interfaces on the monomer. This will be done by measuring the deviation of the surface geometry from the hCpn10 structure and investigating the significant changes to the electrostatic field distributions of the hCpn10 variants.

3.5.2.2 MD simulations I: Molecular designing of hCpn10, CE76 and CPMyc variants

3.5.2.2.1 Simulation of hCpn10

MD simulations were performed on the β-hairpin roof and mobile loops of hCpn10. Data acquired from these simulations indicated that the hCpn10 was quite stable as evident from the 1 ns simulation snapshots shown in Figure 3.8. These images further illustrate that both the β-hairpin and mobile loops were flexible and did not impede heptamer formation. Importantly, as indicated by the MD simulations shown in Figure 3.9A, the cluster of conserved amino acid sequences, namely, *Arg19 Ser20 Ala21* (*RSA*) and *Gly38 Lys39 Val40* (*GKV*) at the *N*- and the *C*-termini of mobile loop, respectively, was packed very tightly during the simulations. A second possible packing scheme was observed during the MD simulations where the lysines (*Lys27* and *Lys35*) side chains pointed away from the loop and the polar interactions stabilised the overall extended structure of two β-strands from the molecule β-core. The electrostatic potentials generated by two adjacent hCpn10 monomers were complementary and therefore provided a greater prospect for dimerization (Figure 3.10A).

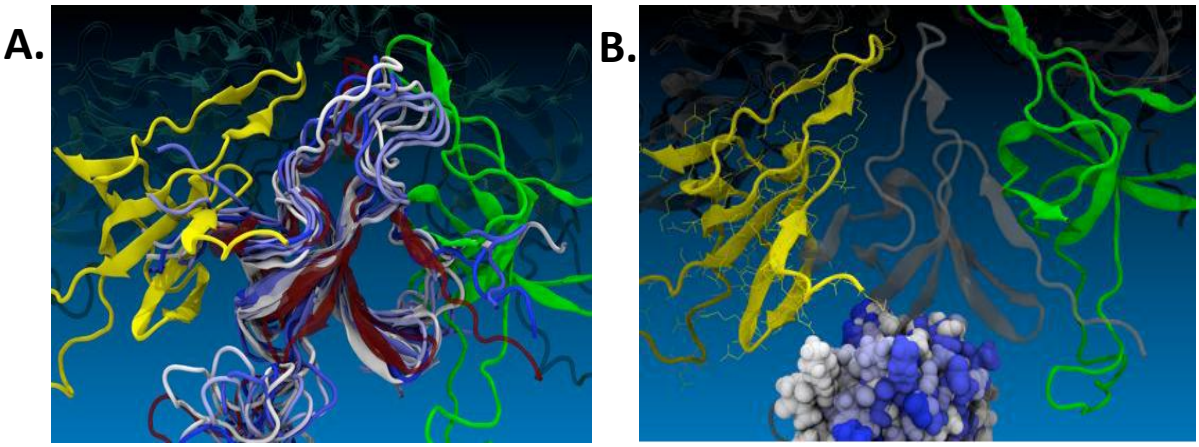


Figure 3.8 MD simulation snapshots of hCpn10 over duration of 1 ns at 310 K

Snapshots taken of hCpn10 MD simulation: **(A)** Overlaid snapshots of MD simulations of the hCpn10 monomer in ribbon representation coloured from white to dark blue. **(B)** Overlaid snapshots of MD simulations were atoms of hCpn10 mobile loop are shown as space-filling model. Both snapshots show monomers over 1 ns MD simulation at 310 K, in 50 ps intervals and coloured from white (350 ps) to dark blue (700 ps), then overlaid post-MD simulations on top of the static model of the hCpn10 heptamer.

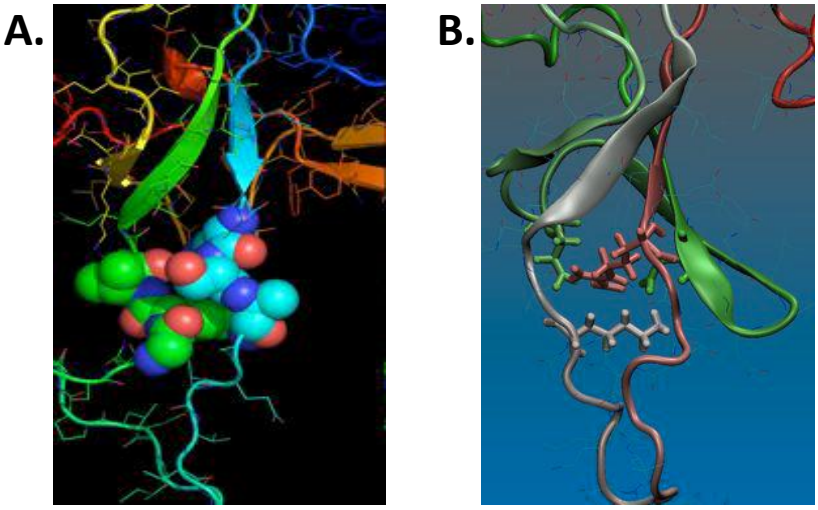


Figure 3.9 Cluster motif *RSA*/*GKV* required for β-barrel core stabilisition

Cluster motif to stabilise hCpn10 protein core. **(A)** hCpn10 clustering of conserved sequences of "*RSA*" and "*GKV*" shown as space-filled models and coloured blue and green, respectively, with strong and tight packing for β-core stabilisation. **(B)** CPMyc, showing the stereo model of the modified *RSA*/*GKV* cluster with weak packing and unstable β-barrel core.

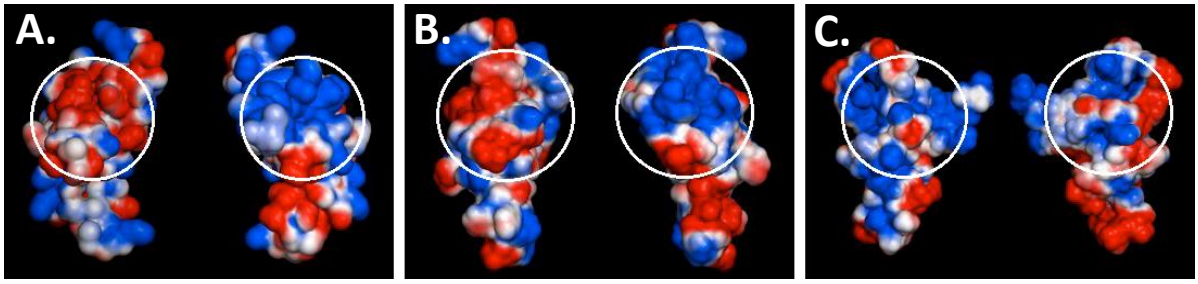


Figure 3.10 Figure 3.10 Electrostatic potential map

Representation of the electrostatic potential on the molecular surface of hCpn10 and its variant obtained from 1 ns MD simulations. **(A)** hCpn10 **(B)** CE76, and **(D)** CPMyc. The molecular surface representation and the electrostatic potential distribution is colour coded as follows: Blue-Positive, Red-Negative and White-Neutral. The dimerisation domains are indicated by a white circle and the mobile loop is located below the circle.

3.5.2.2.2 Simulation of CE76 variant

CE76 is a proposed, NME created by insertion of a peptide termed E-76 which binds Factor VIIa, for the hCpn10 mobile loop between *Ala21* and *Gly38*, (Table 3.1) (detailed in Chapter 4). The E-76 peptide is stabilised with a single disulphide bond between *Cys23* and *Cys33*. Molecular simulations conducted on CE76 revealed an overall stability in 1 ns MD

simulations. However, the E-76 peptide within the mobile loop region was continually changing conformation from the initial structure to a conformation where the E-76 peptide loop made contact with the surface of the protein core that formed the β-barrel like structure (Figure 3.11).

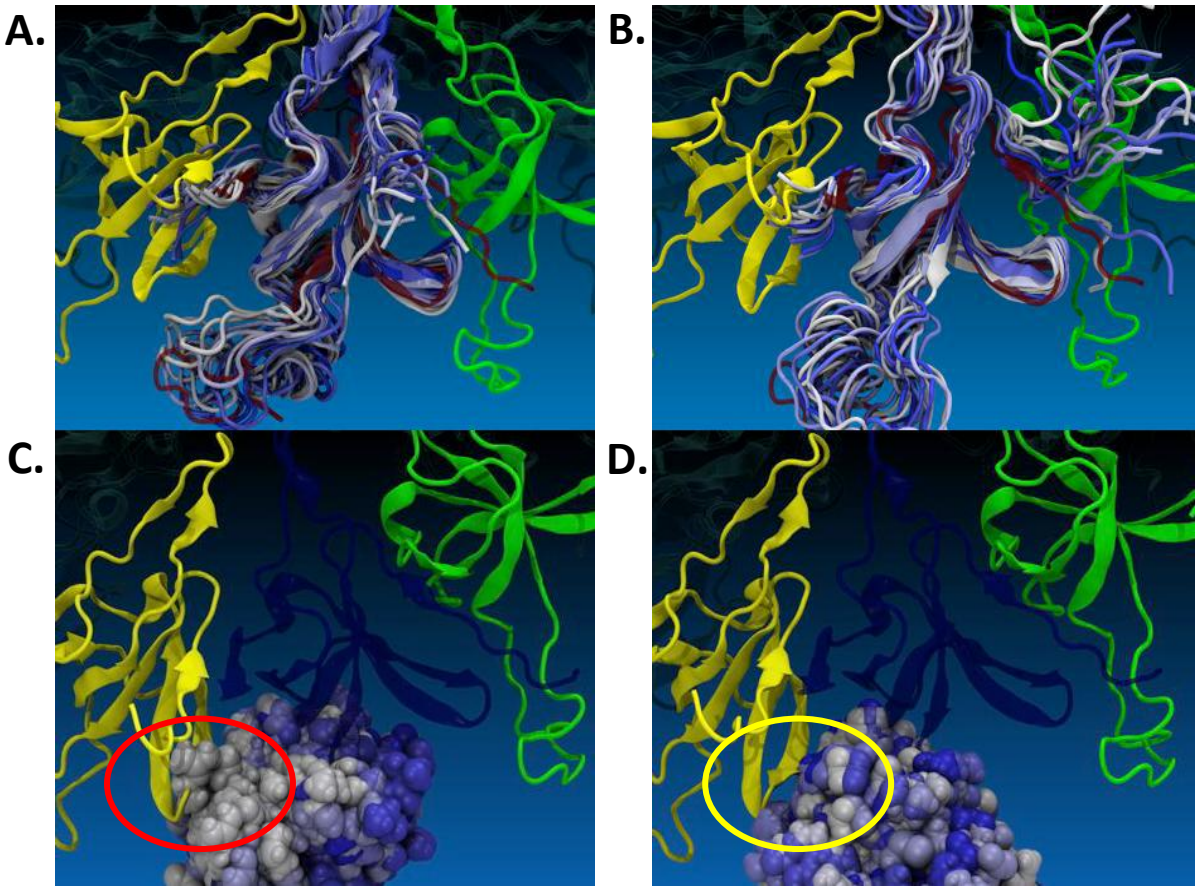


Figure 3.11 MD simulations of CE76 over duration of 1 ns at 310 K

Snapshots of MD simulations of CE76 shown in ribbon representations: **(A)** CE76 M1-Model, and **(B)** CE76 M2-Model with a disulphide bond between *Cys23* and *Cys33* in the mobile loop domain. Atoms of the E-76 mobile loop shown in space-filling representations: **(C)** M1-model showing a hydrophobic cluster formed (red ellipsoid) through interaction of E-76 loop with the protein core, and **(D)** M2-model showing a weaker cluster formed (yellow ellipsoid) through interactions between E-76 loop with the protein core. The detailed descriptions of colour codes are as previously described in Figure 3.8.

CE76 was modelled in two forms; model 1 (M1) was created as a homologous model based on hCpn10 (Figure 3.11A), while model 2 (M2) was created by continuing the simulation of model M1 but enforcing the formation of the disulphide bond between *Cys23* and *Cys33* (Figure 3.11B). This allowed for the proper study of larger conformational spaces for two structurally similar constructs even though both models had different specifications. This was thought to be a necessary step given that the disulphide bond reduces the flexibility of the

loop and makes sampling of the conformational space slower. In the CE76 M1 model, a helix was formed between residues *Leu32* and *Gly38* in the mobile loop. A similar yet shorter helix was formed between residues *Pro33* and *Ser36* in the CE76 M2 model. However, the remaining regions of the mobile loop from both models folded differently as shown in Figure 3.11. It seemed apparent that a transition between the two loop conformations could occur. This was, however, hindered by the disulphide bond, where *Cys33* is part of the helix. This in turn restricted the movement of the loop and therefore, the two conformations were analysed separately.

The newly formed helix observed in the M1 model triggered the formation of a more "hydrophobic cluster" that formed through the interaction of *Lys39* (of E-76 loop) with *Val100, Leu71* and *Glu42* of the core region and thereby potentially preventing the dimerisation of the monomers (Figure 3.11C). Also, the M1 model construct showed that the loop came in close proximity to the core which in turn interacted with *Leu82* to *Lys85* between the β-strands 4 and 5 of the protein core as part of the dimerisation interfaces (Figure 3.11C).

In the M2 model, residues *Val100, Leu71* and *Glu42* also formed a "weaker cluster". However, polar interactions were observed between *Arg30* and *Asp24* of the E-76 loop, *Glu74* of the β-barrel core; and *Glu37* of the mobile-loop and *Lys79* within the core of the protein. Due to more bending within the loop conformations, the mobile loop clashed even more with the turn between the β-strands 4 and 5 (Figure 3.11D). Furthermore, due to the distortions observed in the mutant CE76 structure relative to hCpn10, the electrostatic potential on both dimerisation surfaces was not completely complimentary as clearly illustrated in Figure 3.10B. There were two areas of negative potential that had appeared towards the bottom of the dimerisation site, indicating that there was a lack of electrostatic force in facilitating the dimerisation of the CE76 monomers.

Accordingly, the collisions of the substituted mobile loop domain (E-76 peptide) with the protein β-barrel core and the reductions of the electrostatic complementarities of the contiguous monomers surfaces essentially inhibited the dimerisation of the CE76 conformation. This in turn did not facilitate the development of a heptamer in a similar

structure to hCpn10. This might explain why SE-HPLC revealed the presence of numerous aggregates of varying sizes shown below in section 3.5.5.

3.5.2.2.3 Simulation of CPMyc variant

An MD simulation was conducted on another hCpn10 based NME that contained the Myc-tag (*EQKLISEEDLNG)* within the mobile loop, and called CPMyc (Table 3.1). The Myc-tag was investigated as it is commonly utilised in conjunction with the 9E10 monoclonal antibody (that binds a Myc linear epitope) in protein engineering studies. The Myc peptide contains both charged and hydrophobic residues, and as such could be useful in investigating how charged and hydrophobic residues interact with the β-barrel interface.

The MD simulation indicated that the CPMyc was generally stable. However, the mobile loop region which contained the Myc peptide was highly flexible with a continuous change in conformation during the 1 ns simulation time (Figure 3.12). Remarkably, the Myc-loop did not reach a stable conformation, which could indicate that it might be possible to form a heptamer if the mobile loop was able to adopt a more stable conformation. The MD simulations indicated that the Myc-loop was interacting with the core β-barrel protein at β-strands 4 and 5 thereby disrupting the dimerisation interface towards the adjacent monomer, (Figure 3.12). Since these two β-strands seem to be necessary for the two CPMyc monomers to merge and form a heptamer, the assembly of this construct into an oligomer does not seem very likely. For future study, it might be worth it to extend the MD simulation timeframes (over than 1 ns) to see whether a stable conformation of the mobile loop with Myc-peptide might be formed eventually.

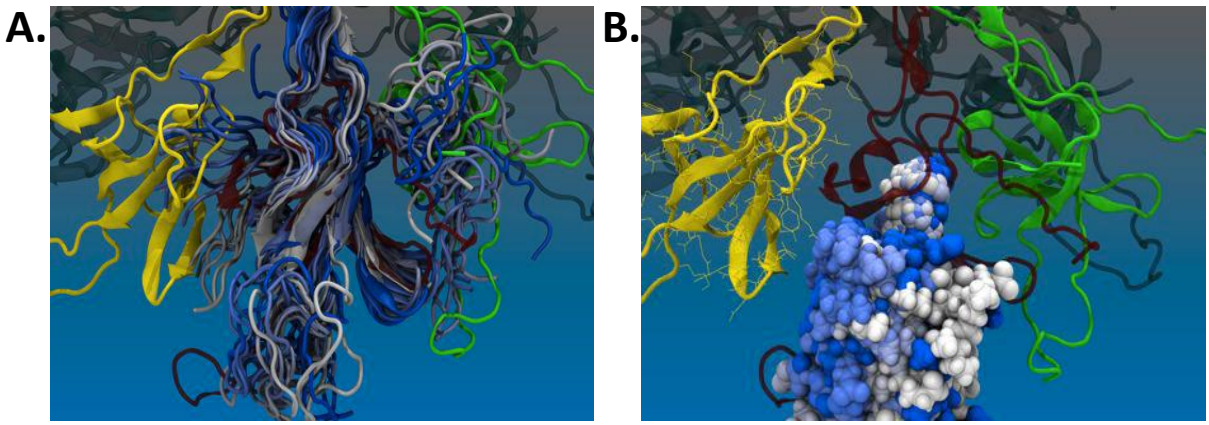


Figure 3.12 MD simulations of CPMyc over duration of 1 ns at 310 K

Snapshots of CPMyc MD simulations: **(A)** Snapshots of overlaid CPMyc in ribbon representations and shown loop spatial extent conformations. **(B)** Snapshots of overlaid CPMyc shown the inserted Myc-loop in space-filling model. The detailed descriptions of colour codes are as previously described in Figure 3.8.

The flexibility of the mobile Myc-loop was likely due to the high content of charged amino acids within the loop (***E*****Q*****K***LIS***EED***LNG*). Such flexibility and the close proximity of the charged side chain residues (of the loop) to the protein β-core caused a different packing scheme of *Arg19 Ser20 Ala21* and *Gly34 Lys35 Val36* motif, as previously discussed. Furthermore, the creation of the "polar cluster" due to the side chains of *Asp82*, *Arg19*, *Glu70*, and *Lys35* led to a close alignment of β-strands 4 and 5 with *RSA*-sequence of the Myc-mobile loop (Figure 3.9B). This new "cluster" which contained residues of the β-core *Asp82* and *Glu70* could be the reason for the observed modification of the binding interfaces between the contiguous monomers (Figure 3.12).

Due to the dramatic changes that were observed during the simulation of CPMyc, another MD simulation was performed. This simulation was shorter in duration (less than 1 ns) and conducted at high temperature at 310 K using a second CPMyc-model based on the final structure of the hCpn10 simulation with the "*RSA/GKV* cluster" as a template because of the observed stabilities of such template (Figure 3.9A). This simulation illustrated that β-strands 4 and 5 were also moved towards the mobile Myc-loop similar to the movement described previously despite the differences in motion details.

The final structure of the CPMyc simulations illustrated that the *C*-terminus also diverged from the initial configuration and interacted with the top β-hairpin roof loop which also

disrupted the second dimerisation interfaces. This distortion also caused a polarity change at the dimerisation sites and therefore a disparity in electrostatic potential (Figure 3.10C). Notably, the high flexibility of both the *N*- and *C*-termini at the polarity of the dimerisation sites might be easily reversible.

The model of CPMyc variant illustrated in this configuration with two β-strands (4 and 5) moving towards the mobile loop may hinder the formation of a heptamer in similar manner as seen in hCpn10. This however, is clearly explained in the experimental results through the unfolding of CPMyc molecule where analytical SE-HPLC of CPMyc showed unstable molecules and the development of soluble aggregates (data not shown).

3.5.2.3 MD simulations II: Molecular redesigning of CE76 and CPMyc variants for heptamer assembly

Previous MD simulations of the hCpn10 variants of CE76 and CPMyc containing E-76 peptide and Myc-peptide, respectively, revealed that the mobile region was highly flexible. This in turn hindered the proper formation of stable oligomers. Therefore, to stabilise the mobile loop region, several peptide-linkers were added in order to create more stabilisation and/or fixation within the mobile loop that could prevent any interference with the dimerisation region of the subunit-subunit hCpn10 monomers. The insertion positions of three Linker IDs including their characteristic AA sequence and various sizes are shown in Table 3.2.

Table 3.2 Summary of peptide linkers used to mimic mobile loop elasticity.

Linker ID	AA Sequence	AA Size	Insert Positions
P1	*AAP/PAA*	3/3	A21 – G38
P2	*PP/PP*	2/2	A21 – G38
E5	*AAETV/EKSQG**	5/5	A21 – G38

*Using native residues of *N*- & *C*-termini of hCpn10 mobile loop.

Following the strategy outlined in the schematic in Figure 3.6, another molecular simulation was conducted in order to determine the ability of the linker sequences to stabilise the mobile loop and therefore facilitate the proper assembly of the variants into a heptamer.

The computational methods utilised in this second phase (II) was based on the same strategies used in previous simulations (Section 3.5.2.2, MD Simulation I). In addition, three simulation models were also run for each modified variants (as described in Tables 3.1 and 3.2, in conjunction). The first MD simulation was to investigate the stability and interaction of the inserted-loop, at the *N*- and *C*-junctures of the loop domain, and the full length monomeric form within an aqueous environment (i.e. dielectric constant ε=78.5). The second MD simulation was conducted with a dielectric constant of ε=78.5, and with removal of the *His*-tag sequence at the *C*-terminus whereas *His*-tag was added in order to facilitate downstream processes for high purity protein yields. This simulation focused on assessing the stability of the inserted and modified loop on its own. Finally, the third simulation involved increasing the statistical relevance of the results by changing the dielectric condition and also to investigate the stability of the modified full length constructs in an environment where bulk water is less abundant (i.e. in an actual cellular surrounding environments) at dielectric constant of ε=40. Each of these MD simulations ran for about 1 ns at 310 K. The final strategies that were applied to this second phase were to stimulate the six histidine residues of the *C*-terminus as uncharged and the inserted E-76 (of CE76) with an intact disulphide bond.

Simulations of all modified constructs were conducted in their monomeric form. However, the snapshots illustrated in Figures 3.13 to 3.18 will often show the heptameric form of the crystallised wt Cpn10 of PDB. 1WE3. However, as discussed previously in section 3.2.2.2.1, simulation of heptameric hCpn10 (>109,000 atoms) was beyond the technical capabilities of the CPU and therefore, all simulation data on monomeric hCpn10 were superimposed on the crystal structure of 1WE3. This in turn would allow for a reasonable estimation on whether an inserted peptide such as E-76 or Myc- would impinge on the assembly of the hCpn10 variants monomer of the redesigned CE76 and CPMyc, respectively.

3.5.2.3.1 **Simulation of CE76$_{P1}$ variant**

The CE76$_{P1}$ variant model contains the E-76-insert within the mobile loop but is separated from the hCpn10 β-barrel core by an *Ala-Ala-Pro* (P1-linker) with a single proline residue in each *N*- and *C*-termini loop extensions (Table 3.2). Simulation of full length CE76$_{P1}$ in an aqueous environment (ε=78.5) revealed that the β-core and the mobile loop adopted a more stable conformation (Figure 3.13). However, it appeared that the mobile loop had some

degree of mobility as this region was bending in an upward direction and could possibly interfere with the heptamer formation. Conversely, it is also possible that the limited flexibility may prevent heptamer formation of the hCpn10 variant.

In the second simulation, $CE76_{P1}$ was subjected to an aqueous environment with a dielectric constant of ε=78.5. In addition, the *C*-terminal *His*-tag was removed. The additional P1-linker, however, failed to fully stabilise the mobile loop. In fact, the last two β-sheet residues from *Lys79* to *Phe90* of the $CE76_{P1}$ β-strands were lost during this simulation due to the strong interactions of the mobile loop residues with the protein core. This in turn could prevent the monomer from dimerising and then forming a heptamer.

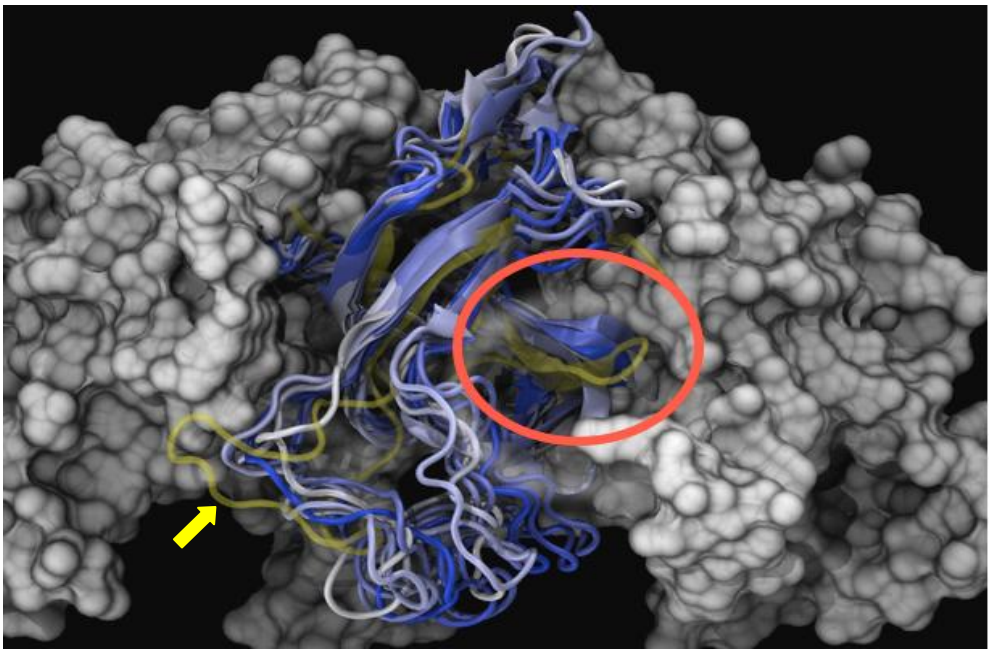


Figure 3.13 MD simulation of $CE76_{P1}$ over duration of 1 ns at 310 K

Snapshots from 1 ns MD simulations at 310 K shown $CE76_{P1}$ variant in ribbon representation and overlaid onto the hypothetical hCpn10 structure (yellow), the location in a cut-out of the heptamer to estimate the ability of $CE76_{P1}$ to form heptamer. Red ellipsoid illustrates a possible β-barrel core distortion by the end of the simulations. Series snapshots overlaid in 10 ps intervals from the last 100 ps of the calculation, and coloured from light to dark blue.

The E-76 peptide contains a number of polar residues which in turn are more than likely to influence its interaction with the β-barrel core. Therefore, a third simulation was conducted on full length $CE76_{P1}$ monomer at a dielectric constant of ε=40 in order to gain an understanding of the electrostatic interactions that may occur between the E-76 peptide and the β-core. This particular simulation will be far more realistic than the simulations that were

conducted in pure water as the conditions mimic the intracellular environment. Results from the third simulation indicated that the *C*-terminus *His*-tag had a slight interaction with the E-76 loop which in turn stabilised its conformation. The β-core was far more stable in the third relative to the first simulation (Appendix 3).

The electrostatic potentials of the $CE76_{P1}$ protein surface were calculated in a similar manner as described in Section 3.5.2.2.2. The electrostatic potential generated by the $CE76_{P1}$ monomer supported dimerization as shown Figure 3.14A and furthermore, the distribution of the positively and negatively charged characters at the molecular surface of monomer/monomer interfaces appeared to vary from the original design of CE76 (i.e. CE76 without P1-linker) and similar to the hCpn10, despite the extension and difference of the mobile loop conformation. It is quite possible that the P1-linker of "*AAP*/*PAA*" sequence could have an influence in rearranging the electrostatic configurations.

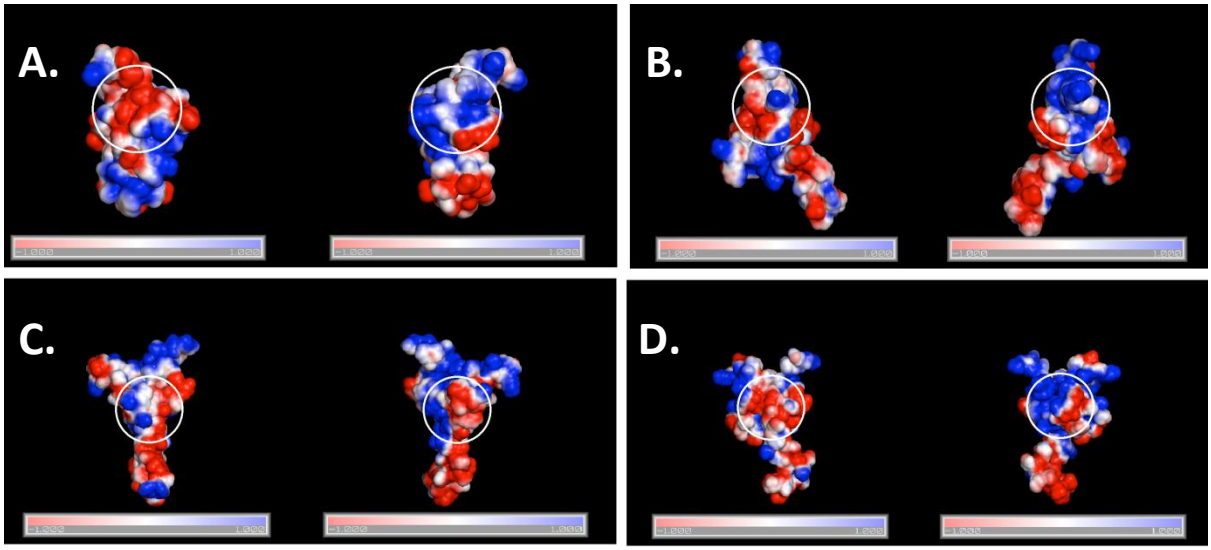


Figure 3.14 Electrostatic potential maps

Representation of the electrostatics potential at the molecular surface of hCpn10 variants obtained from 1 ns MD simulations of: **(A)** $CE76_{P1}$; **(B)** $CE76_{P2}$; **(C)** $CE76_{E5}$; and **(D)** $CPMyc_{P1}$. Detailed descriptions of colour codes are as described in Figure 3.11.

3.5.2.3.2 Simulation of $CE76_{P2}$ variant

$CE76_{P2}$ was a second CE76 variant that contained the E-76 peptide within the mobile region. This peptide was separated from the β-core by a *Pro-Pro* (P2-linker) (Table 3.2) with double proline residues in both *N*- and *C*-termini extensions. $CE76_{P2}$ underwent a simulation in an

aqueous environment at a dielectric constant of ε=78.5, similar to the first simulation as discussed in Section 3.2.2.3.1. Simulations for the CE76$_{P2}$ variant revealed that the *C*-terminus *His*-tag was highly flexible, similar in nature to the CE76$_{P1}$ variant. However, this tag did not appear to interact with the E-76 loop within the hCpn10 (Figure 3.15). On the other hand, because the *C*-terminus, including the *His*-tag, is quite extended (i.e. *KYVDGGGGHHHHHH*) it may subsequently hinder heptamer formation. This in turn will make computational predictions of the CE76$_{P2}$ dynamics more difficult. Nevertheless, given the high flexibility observed in the extended *C*-terminus, it is likely that this region could also adopt a stable conformation in the multimeric structure and allow for the interaction of adjacent monomers to stabilise the heptamer. However this will require wet-lab verifications. Another interesting observation was that the mobile loop region of CE76$_{P2}$ was flexible and yet could adopt a number of stable conformations, none of which interfered with heptamer formation (Figure 3.15). In essence, the protein β-barrel core of CE76$_{P2}$ was well conserved and showed many similarities to the hCpn10 protein core as exemplified in Figure 3.8.

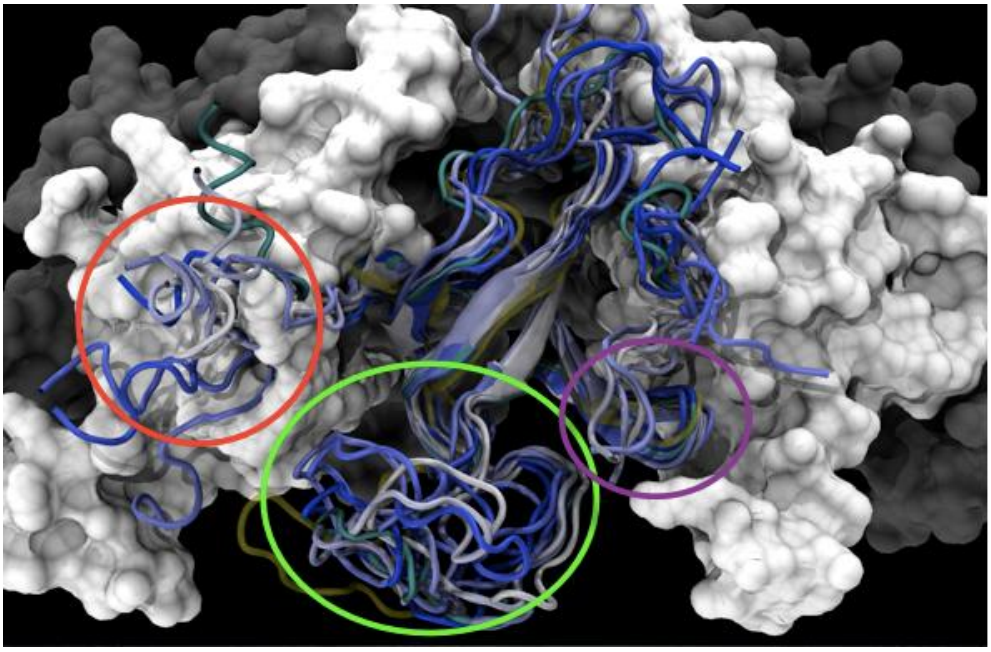


Figure 3.15 MD simulation of CE76$_{P2}$ over duration of 1 ns at 310 K

Snapshots from 1 ns MD simulations at 310 K shown CE76$_{P2}$ variant in ribbon representation. The snapshots show the conformation of the β-core barrel and the mobile loop of the redesigned CE76$_{P2}$. Where, red, green and purple ellipsoids show the locations of the *C*-terminus *His*-tag, the inserted E-76 mobile loop and the last two β-strands (4 and 5) of the central β-sheet of the protein core, respectively. Detailed descriptions of colour codes are as described in Figure 3.13.

A second simulation was conducted on the $CE76_{P2}$ variant whereby the *C*-terminal *His*-tag was removed. Simulations were conducted in an aqueous environment at a dielectric constant of ε=78.5, and simulation analysis showed that the P2-linker can independently stabilize the inserted E-76 at the mobile loop domain. Furthermore, like with the first simulation, the structures of the β-sheets were not compromised in the *N*- and *C*-termini. However, there was some degree of flexibility and interaction with the β-hairpin roof loop (Appendix 3). It is possible that this interaction is an artefact of the simulation. Importantly, the mobile loop did not appear to collide with the contiguous monomers which can prove to be critical in the assembly of the $CE76_{P2}$ heptamer as shown in Figure 3.15.

The third simulation was conducted on full length $CE76_{P2}$ at a dielectric medium of ε=40. This simulation further confirmed that the P2-linker was rigid enough to create stability in the E-76 loop and prevent interaction with the β-Core of hCpn10. Based on the data acquired from these simulations it is possible that the P2-linker may be conducive to heptameric formation. Similar to the previous simulation, both *N*- and *C*-termini were slightly unfolded and interacting with the β-hairpin roof loop. However, this could once again be regarded as an artefact of the simulations. Nevertheless, despite the small degree of flexibility seen in the *N*- and *C*-termini, the remainder of the $CE76_{P2}$ β-barrel core was very stable. The electrostatic potential that was generated by the modified $CE76_{P2}$ variant illustrated that the electrostatic allocations supported the development of the heptamer molecule (Figure 3.14B).

3.5.2.3.3 **Simulation of $CE76_{E5}$ variant**

The third variant was $CE76_{E5}$ whereby the E5-linker was positioned between the inserted E-76 loop and the protein β-barrel core (Table 3.2). Specifically, the E-76 peptide was inserted further downstream of the protein core by retaining five native amino acids of the native-mobile loop, namely *AAETV* and *EKSQG* of *N*- and *C*-termini, respectively. As seen in the previous simulations discussed, the *C*-terminus *His*-tag of $CE76_{E5}$ was highly flexible in the aqueous environment at a dielectric constant of ε=78.5 (Figure 3.16). The first simulation illustrated that the *His*-tag did not interact significantly with the mobile loop but it folded in the reverse direction and disrupted the first strand of the core β-sheet. Moreover, the extended sequence of the histidine-tag obstructed the heptamer formation. However, given that the *C*-terminus is highly flexible, it could adopt a more stable conformation at the dimerisation site, and subsequently allow for interaction between adjacent monomers in order to stabilise the

heptamer. It was apparent that the mobile loop of $CE76_{E5}$ was flexible and could adopt a number of conformations, although none of these appeared to interfere with heptamer formation. Finally, the protein β-barrel core of $CE76_{E5}$ was structurally well conserved as seen with hCpn10 despite the unfolding of both the *N*- and *C*- termini (Figure 3.16).

In the second simulation, the *C*-terminus *His*-tag of $CE76_{E5}$ was removed and the modelling was performed in an aqueous environment at 310 K (ε=78.5). Similar to the first simulation, the E-76 loop region showed some degree of flexibility but this peptide stabilised the stem region of hCpn10 and did not interact with the β-hydrophobic core (see Appendix 3 for more detail).

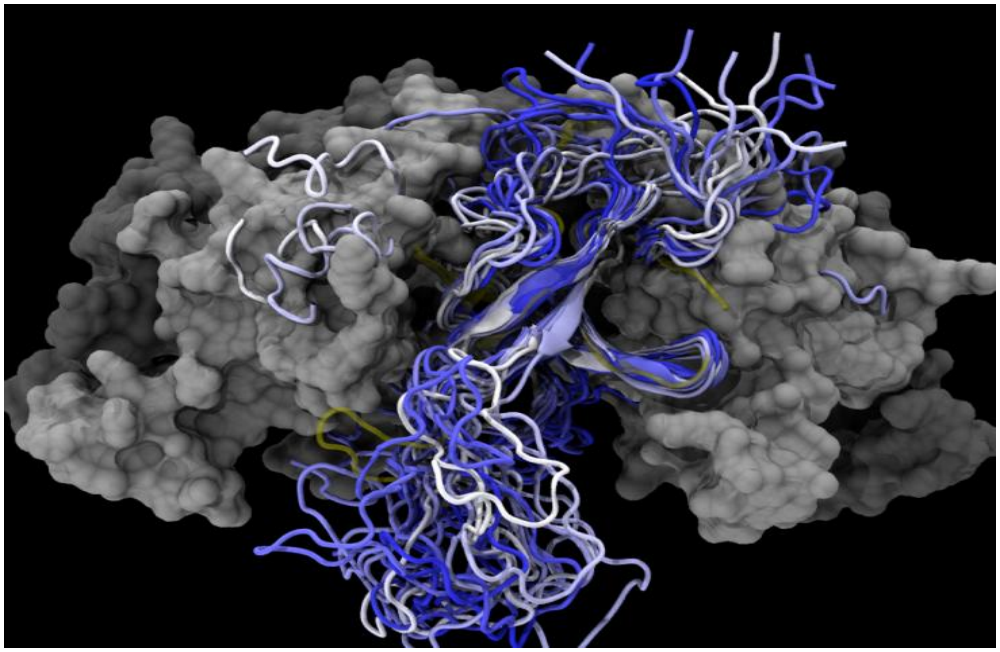


Figure 3.16 MD simulation of $CE76_{E5}$ over duration of 1 ns at 310 K

Snapshots from 1 ns simulations at 310 K shown $CE76_{E5}$ in ribbon representation shows the conformations of the β-core and the mobile loop, green ellipsoid illustrated a structural distortion appeared toward the end of MD simulations. Detailed descriptions of colour codes are as described in Figure 3.13.

In the third simulation, full length $CE76_{E5}$ was modelled in a medium dielectric environment at ε=40. In this simulation, the $CE76_{E5}$ variant appeared to be distorted where residues *Glu43* of the E-76 loop and *Glu45* of the E5-linker "*P**E**KSQ*" formed hydrogen bonds with *Lys90* and *Arg102* of the protein β-barrel core (second last β-strand) (Figure 3.17A). The E5-linker of $CE76_{E5}$ which contained polar and charged residues cannot constantly prevent the mobile loop from interacting with the protein core. In such circumstances, it is uncertain whether the

$CE76_{E5}$ structure can form dimers (i.e. a prerequisite for heptamer configuration) rather than a heptamer as observed in hCpn10 (as seen in Figure 3.8).

Electrostatic potentials were calculated for the $CE76_{E5}$ variant. Despite the protein being more distorted in this case as shown in Figure 3.17A, the presence of electrostatic charges at the dimerisation interfaces and topological complementarity may be the ideal conditions to support dimerisation and heptamer formation (Figure 3.14C).

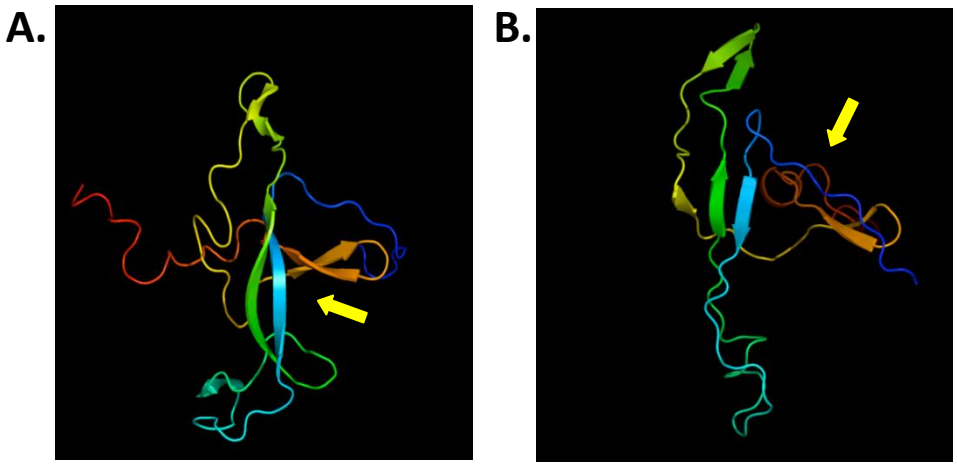


Figure 3.17 The final structures from MD simulations of hCpm10 variants

MD simulation final structure of hCpm10 variants. **(A)** $CE76_{E5}$ monomer construct from simulation-3 in medium dielectric environment (ε=40), shows monomer structure distortion and rearrangement of the core β-barrel, central β-strands extended, and β-strands 3 and 6 lost. **(B)** $CPMyc_{P1}$ monomer from simulation-3 in medium dielectric environment with ε=40, shows monomer structure distortion and the loss of the central topology of the β-stands.

3.5.2.3.4 Simulation of $CPMyc_{P1}$ variant

The modified $CPMyc_{P1}$ variant contained a Myc-loop domain that was stabilized with a P1-linker at *N*- and *C*-junctures of the mobile loop (Table 3.2). It was hypothesised that the use of the P1-linker might bring more stability to the Myc-loop despite the inherent properties of the peptide. As such, similar MD simulations were performed for the redesigned $CPMyc_{P1}$that also contained extended a *C*-terminus *His*-tag. The first MD simulation was performed in an aqueous environment (ε=78.5) where the *C*-terminus was highly flexible and the *His*-tag interacted loosely with the Myc-loop during certain parts of the simulation, (Figure 3.18). As a consequence, the slight extension of the *C*-terminal domain obstructed formation of the heptamer. However, given the high flexibility (i.e. visually of snapshots and numerical value of rmsf/rmsd) of this domain, interaction between adjacent monomers may

stabilise the *C*-terminus domain and facilitate the formation of heptamers (data not available). Furthermore, despite the high flexibility of the Myc-loop, none of the adopted conformations appeared to interfere with the heptamer configuration. Like previous simulation studies, the β-barrel core of $CPMyc_{P1}$ was structurally conserved in similar topology to hCpn10 (Figure 3.18). Moreover, a slight unfolding of the *N*- and *C*-termini domains was observed during the simulations and could potentially interfere with the ideal configuration of some β-strands but the β-strands still remained in an extended conformation.

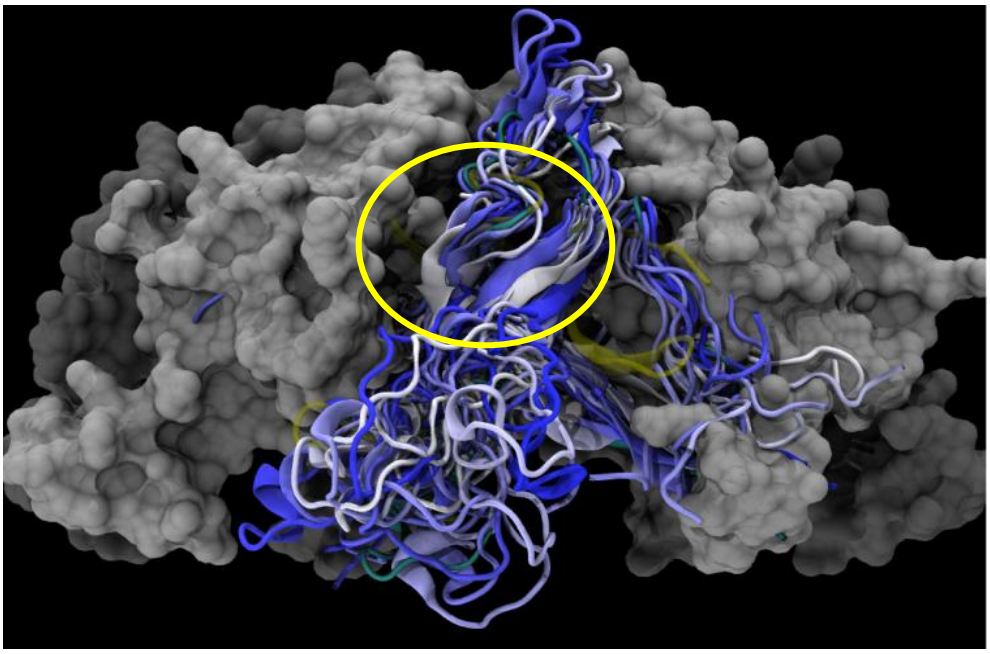


Figure 3.18 MD simulation of $CPMyc_{P1}$ over duration of 1 ns at 310 K

Snapshots from 1 ns MD simulation at 310 K shown $CPMyc_{P1}$ in ribbon representation showing the conformational changes of the β-core and the mobile Myc-loop; yellow ellipsoid illustrates β-strands 3 and 6 which were lost toward the end of MD simulations. Detailed descriptions of colour codes are as described in Figure 3.13.

In the second simulation of $CPMyc_{P1}$, the extended *C*-terminus *His*-tag was removed and simulated in an aqueous environment (ε=78.5). The model of $CPMyc_{P1}$ was distorted during simulation, similar to what was observed in the $CE76_{-E5}$ model whereby the central β-strands of the $CPMyc_{P1}$ were extended despite the loss of two β-strands 3 and 6. On the other hand, no interactions between the *C*-terminus and the β-core were observed during the MD simulations and thus would not interfere with heptamer formation. The P1-linker in conjunction with the inserted Myc-peptide might induce an extension of the two central β-strands, and therefore cause a distortion of the $CPMyc_{P1}$ variant and influence the

dimerisations and heptamer development which would presumably be different to the wild-type hCpn10.

The third MD simulation was on full length CPMyc$_{P1}$, in a biological dielectric medium of ε=40 that resembled the intracellular environment. In this simulation, the surrounding conditions forced a conformational change where the extended *His*-tag began to fold back towards the vicinity of the *C*-terminus. Remarkably, the Myc-loop was still in an extended conformation and did not interact with the protein core as indicated by the third simulation mode. Despite that, the CPMyc$_{P1}$ structure was distorted and had lost the central topology of the two β-sheets towards the end of the MD simulations (Figure 3.17B).

The distributions of the electrostatic potentials were calculated for CPMyc$_{P1}$ using the shortened molecular model without the *C*-terminus *His*-tag, despite the fact that the protein was distorted in this simulation (Figure 3.17B). As a consequence, this study illustrated that the electrostatic distributions at the potential dimerisation sites were not entirely complimentary and would more likely hamper dimerisation and thus heptamer formation (Figure 3.14D).

3.5.3 Force-Mode approach (Non-MD simulations)

An *in vitro/*wet-lab approach towards forcing heptamer formation of hCpn10-based NME was undertaken through a "Force-Mode" approach. In this particular technique, each hCpn10 monomer was linked at the genetic level, whereby three glycine residues (*GGG*) were introduced between the *C*- and *N*-termini of the monomers, thus fusing all seven monomers together in one gene. The first step involved utilising the CE76 gene as a template where three glycine (*Gly102 Gly103 Gly104*) residues were genetically inserted from *Asp101* of *C*-terminus to *Ala1* of adjacent monomer *N*-terminus (Figure 3.19). An artificial DNA encoding FH-CE76 (i.e. forced-heptamer design construct with 3x*Gly*) was then transformed and ligated to an expression pET30a vector (i.e. pET30a-FH-CE76) using the standard molecular biology described in Chapter 2. The FH-CE76 was then expressed and purified from low cost bacterial expression systems (i.e. *E. coli* BL21), followed by subsequent verifications using optimum protein characterisation techniques, such as SDS-PAGE and analytical SE-HPLC.

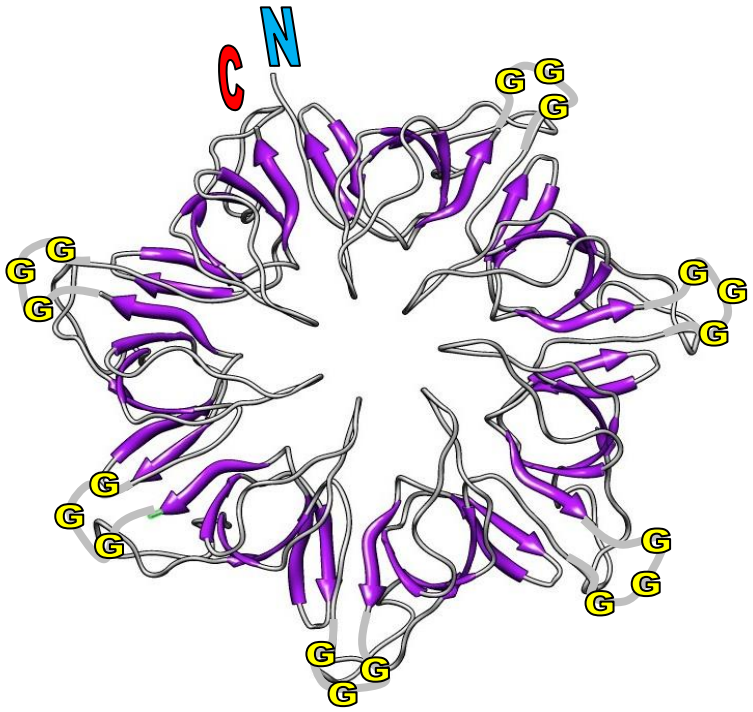


Figure 3.19 Three-dimensional representation of proposed Forced Heptamer-CE76

Representation of FH-CE76 (Top view) illustrating the location of glycine-linker in Force-Mode approach. The *C*-terminus of one monomer was linked to the *N*-terminus of the adjacent monomer by three glycine residues (*Gly-Gly-Gly*). FH-CE76 shown in ribbon representations (i.e. β-barrel: violet, coil: gray, G: glycine, ***N***: *N*-terminus, and ***C***: *C*-terminus) modelled by UCSF Chimera 1.5.3.

The estimated molecular weight of the FH-CE76 was roughly 84 kDa and therefore due to the protein size and other characteristics of inserted E-76 peptide (i.e. combination of charge and hydrophobic residues) in comparison to native hCpn10, the expression of this protein resulted in the formation of inclusion bodies (IB). This is typically the case when large recombinant proteins are rapidly expressed in *E. coli* (BL21) systems. Nevertheless, the process is reversible and the IBs were resolubilised in 8 M of Urea and dialysed into an alternative optimized buffer (see Chapter 2) that facilitated the refolding of the FH-CE76.

The Force-Mode approach indeed allowed for the formation of a heptamer as determined by SDS-PAGE (Figure 3.20). However, FH-CE76 proved to be relatively unstable during the purification process and resulted in a degraded product. Nevertheless, when the same method was applied to hCpn10 (named FH-Cpn10), this variant not only formed a stable heptamer,

but was also more stable during the purification process. Thus, there is potential in using the force-mode strategy as an alternative method for heptamer development.

3.5.4 Protein expression and purification

In order to test theoretical predictions by MD simulations of NME stability and heptamer formation, all designed hCpn10 variants that underwent MD simulations were subsequently synthesized, expressed in *E. coli* and purified. In order to assess the stability and oligomerisation of the hCpn10-based NME *in vitro*, the hCpn10 variants were designed *in silico* whereby the mobile loop was substituted with different functional peptides as described in Table 3.1. In addition, each hCpn10 variant contained a *His*-tag at the *C*-terminus in order to purify the protein at a high yield. Finally, based on the simulation-cycle schematic diagram in Figure 3.6, each hCpn10 variant underwent two phases of MD simulations in order to further optimise the construct that contained different non-native peptides and be able to develop heptamer configuration.

Following standard expression and purification procedures (detailed in Chapter. 2), hCpn10 and all designed variants were expressed in *E. coli* strains of BL21(D3) and extracted as soluble proteins. Following expression and *E. coli* cell lysis, proteins of supernatant were subsequently passed through Ni^{2+}Sepharose column of immobilized metal affinity chromatography (IMAC) in order to purify the *His*-tagged hCpn10 variants. Purity of the hCpn10 variant eluants was verified through gel electrophoresis by running the sample through a dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 3.20). Eluted hCpn10 and its engineered counterparts were further purified by running the samples though a gel filtration Superdex® 200 GL column and subsequently identified using matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Figure 3.21).

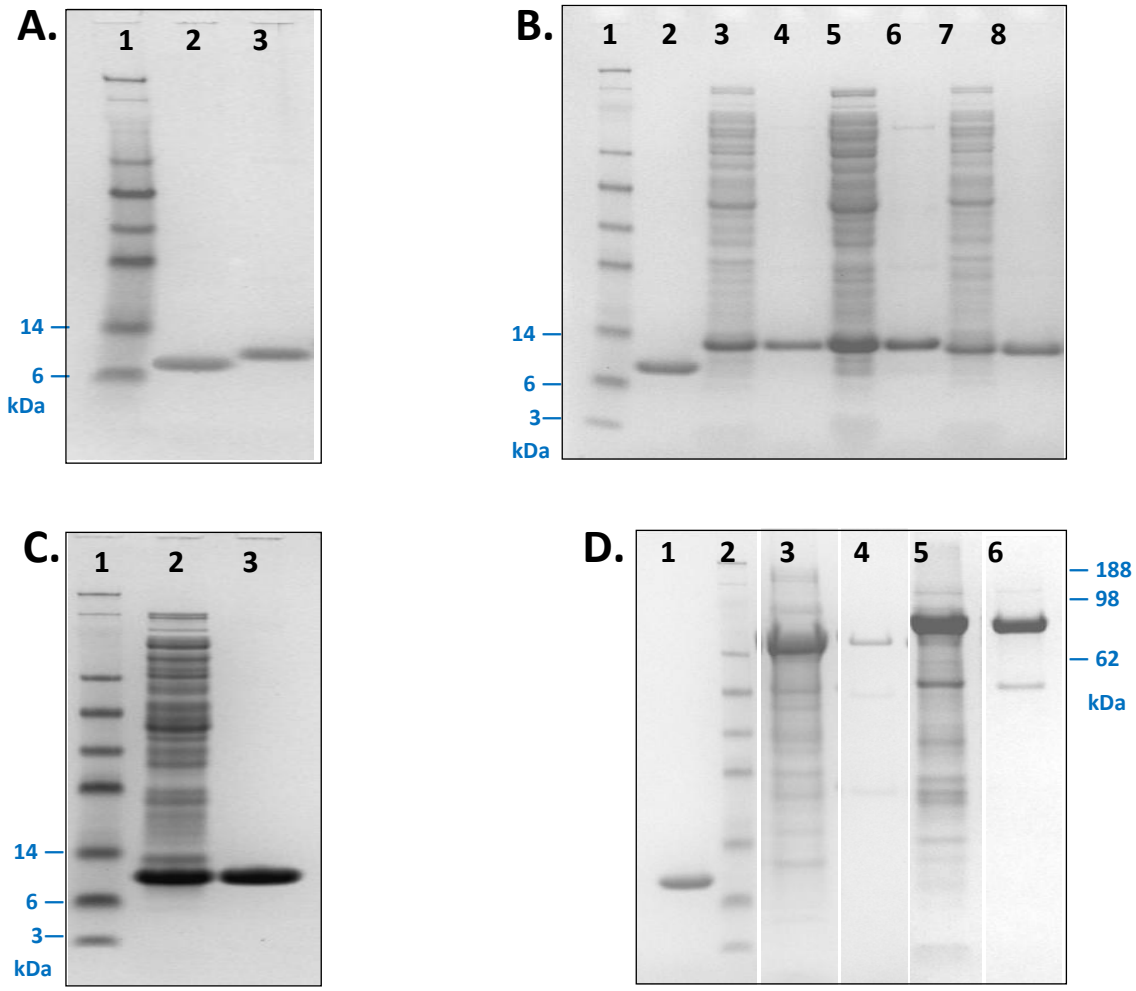


Figure 3.20 SDS-PAGE analyses of hCpn10 and its variants.

SDS PAGE analyses of hCpn10 and hCpn10 variants using a 4-12 % Bis-Tris acrylamide gel (NuPAGE™) and Coomassie stain (SimplyBlue™SafeStain). **(A)** Phase I of MD simulation (original construct): Lane1, Mwt marker (SeeBlue®Plus2, from Invitrogen™); Lane2, hCpn10 (control, 10 kDa); Lane3, CE76 (12.2 kDa) from HisTrap™ FF eluate; **(B)** Phase II of MD simulation -details as before- Lane1, SeeBlue® Plus2 Mwt marker; Lane2, hCpn10 (10 kDa, control); Lane3-4, $CE76_{E5}$ (13 kDa) of post-induction/purified elute; Lane5-6, $CE76_{P1}$ (12.6 kDa) of post-induction/purified elute; Lane7-8, $CE76_{P2}$ (12.5 kDa) of post-induction/purified elute; **(C)** Phase II of MD simulation (variants): Lane1, SeeBlue® Plus2 Mwt marker; Lane 2, $CPMyc_{P1}$ (3 hrs) of 0.4mM IPTG post-induction cell lysate extract of *E. coli* BL21(DE3); Lane 3, $CPMyc_{P1}$ (11.8 kDa) of HisTrap™ FF elute; **(D)** Force-mode Approach: Lane1, hCpn10 (10 kDa, control); Lane2, SeeBlue® Plus2 Mwt marker; Lane3, post-induction soluble factions of FH-Cpn10 (71 kDa); Lane4, pooled of purified fractions from IMAC; Lane5, post-induction soluble factions of FH-CE76 (84 kDa); Lane6, pooled of purified fractions from IMAC.

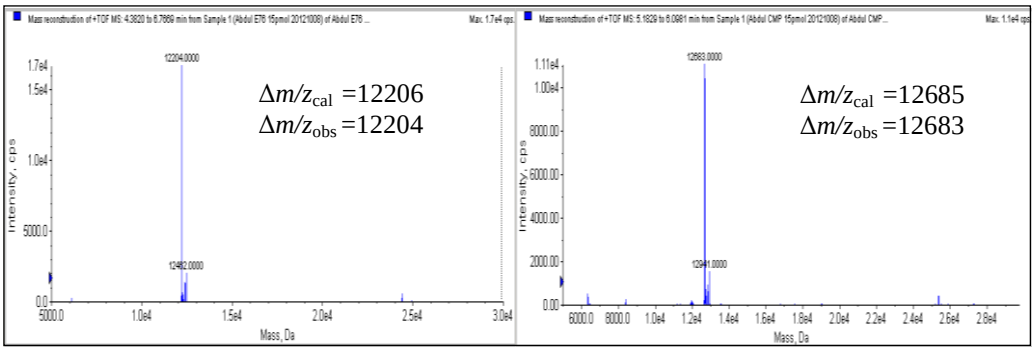


Figure 3.21 MALDI-TOF MS comparisons of CE76 and redesigned CE76$_{P1}$

Calculated molecular weight for CE76 (left frame) and CE76$_{P1}$ (right frame), as compared to observed molecular weight. The resolved samples were determined from SE-HPLC eluates.

3.5.5 Analytical chromatography

In order to verify the molecular modelling simulations of engineered Cpn10 *in vitro* (Table 3.1) all proteins were subjected to an analytical size exclusion-HPLC and chromatograms from each run were compared with hCpn10 (Figure 3.22).

The SE-HPLC chromatograms of mutant proteins from Phase-I MD simulations (i.e. unmodified constructs) revealed that a heptamer did not form indicating the protein to be unstable and subsequently degraded during the purification process. Conversely, unmodified CE76 only formed a tetramer (Figure 3.22). Of all the engineered constructs based on hCpn10, the CPMyc not only failed to form a stable heptamer but the protein also formed soluble aggregates that could not be analysed through SE-HPLC (data not shown).

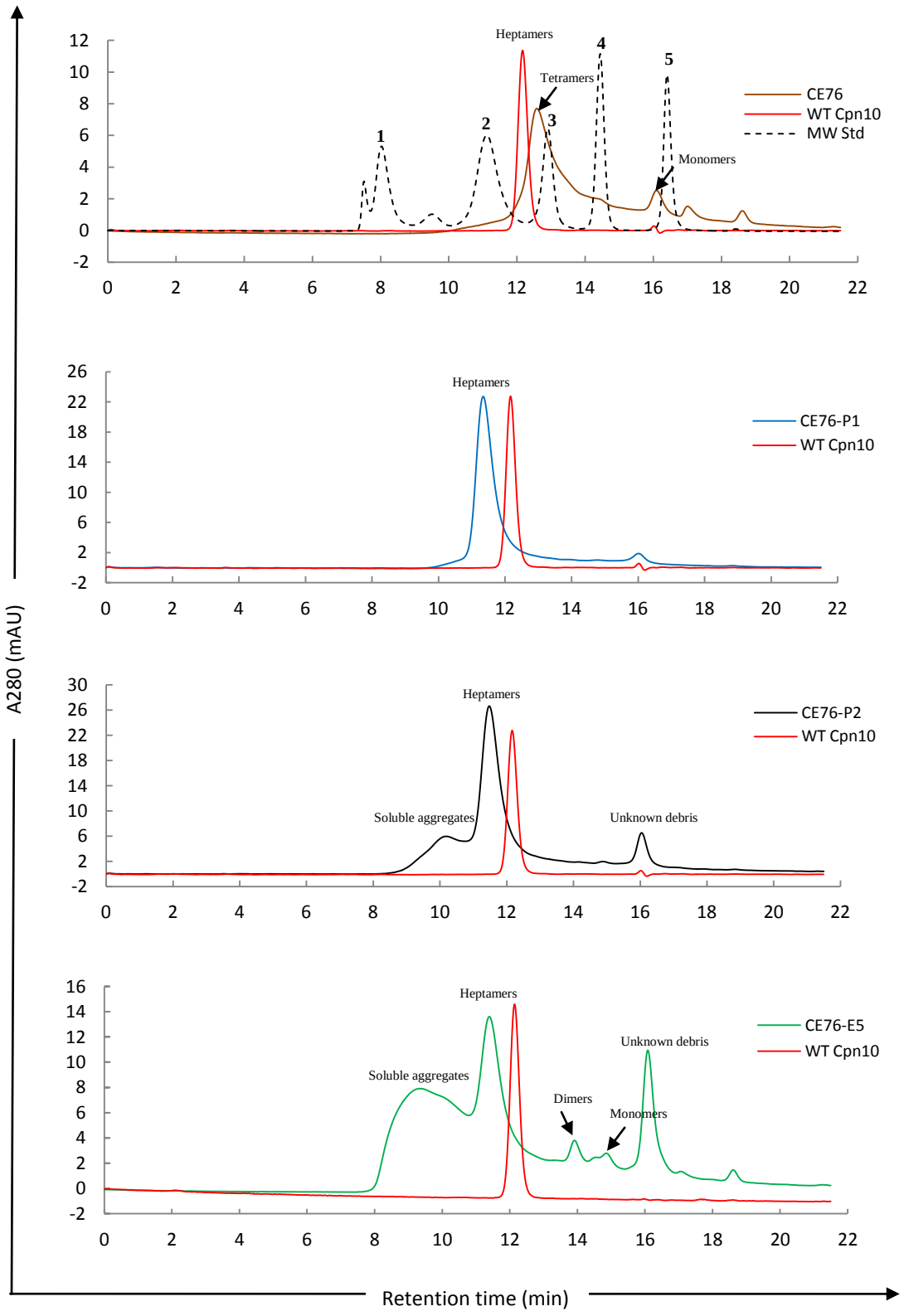


Figure 3.22 Size exclusion HPLC chromatograms of recombinant hCpn10 variants

SE-HPLC chromatograms of hCpn10 (red, 70 kDa, wt control) versus NME variants CE76 (brown); $CE76_{P1}$ (blue); $CE76_{P2}$ (black); and $CE76_{E5}$ (green), respectively. $CE76_{P2}$ and $CE76_{E5}$ variants developed soluble aggregates. $CE76_{E5}$ also developed dimers and monomers. BioRad™ (dash line) standard combined of; **1**/Thyroglobulin (670 kDa); **2**/γ-globulin (158 kDa); **3**/Ovalbumin (44 kDa); **4**/Myoglubin (17 kDa); **5**/Vitamin B_{12} (1.35 kDa), used for Mwt integrations of target proteins (detailed in Chapter 2).

Clearly, the modified constructs predicated from Phase-II MD simulations (i.e. stabilised by various linkers from Table 3.2) predicted that a stable heptamer could be formed, albeit under specific conditions for some mutants (e.g. lower pH level). As the MD simulations had revealed, the P1- and P2-linkers were effective in restricting and stabilizing the E-76 peptide within the mobile loop region. This in turn prevented the interaction of the E-76 peptide with the β-core complex of hCpn10. Thus, the P1- and P2-linkers conferred an overall stability in the protein that eventually facilitated the proper formation of a heptamer (Figure 3.22). Based on the promising data seen with the $CE76_{P1}$ and $CE76_{P2}$, further stability studies were conducted by storing these samples at -20 °C, 4 °C, 25 °C and 37 °C at a timeframe of 1 and 4 weeks. Analyses conducted through SDS-PAGE and SE-HPLC revealed that the proteins were stable under these varying storage conditions (Appendix 2).

The E5-linker within the $CE76_{E5}$ model was rather contrary to the observations made in the MD simulation which indicated that the $CE76_{E5}$ would not form a heptamer based on the flexibility of the E-76 loop and the distortions within $CE76_{E5}$. Nonetheless, the SE-HPLC chromatogram showed a presence of monomers, dimers and soluble aggregates, in addition to formation of heptamers (Figure 3.22). Despite forming heptamer, the overall stability of the $CE76_{E5}$ model was temporary and the protein quickly degraded. Further studies were conducted where the P1-linker was incorporated into the CPMyc model construct. The SE-HPLC chromatogram revealed that the $CPMyc_{P1}$, associated with control hCpn10 formed a heptamer under physiological conditions that was similar to hCpn10 (Figure 3.23).

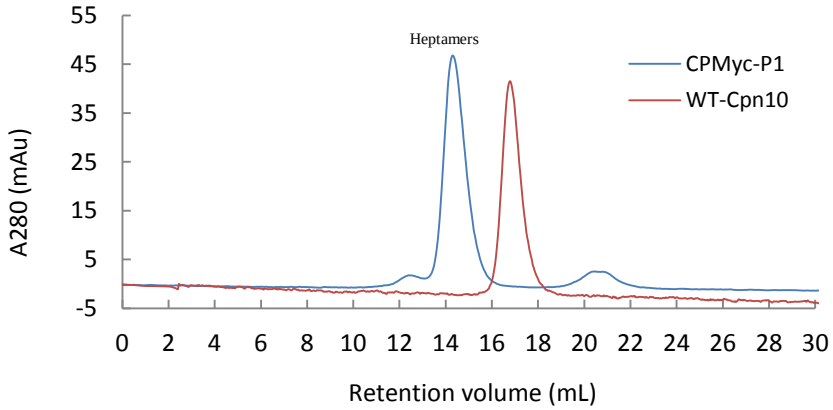


Figure 3.23 Size exclusion chromatogram of $CPMyc_{P1}$ versus hCpn10

SEC chromatogram determined molecular size of recombinant proteins by gel filtration Superdex® 75 column. The $CPMyc_{P1}$ (88 kDa) and hCpn10 (wt control, 70 kDa) showed in blue and red, respectively. Calculated Mwt resolved from Mwt standard curve of known proteins (detailed in Chapter 2).

3.6 Discussion

There are opportunities for the development of alternative molecular scaffolds with similar binding capabilities as mAbs, that may have niche applications and generate novel intellectual property. There are currently over fifty endogenous and non-endogenous proteins that have been developed as alternative scaffolds for a number of applications such as Affibodies, Anticalin, Monobodies, DARPins and others that have been extensively reviewed in Chapter 1 and elsewhere (Skerra, 2007, Koide and Koide, 2007, Stumpp et al., 2008, Lofblom et al., 2010).

The general principle behind the success of designing an effective scaffold is to introduce a non-native and functional peptide(s) that will not compromise the overall stability of the scaffold in use. Similar to other mAb structures, many non-antibody based scaffolds possess flexible loop structures that can be substituted for a non-native peptide sequence such as Monobodies derived from human ^{10}FN3, whereby random mutations of loop residues led to VEGFR2-antagonist as anti-angiogenesis for cancer treatment (Getmanova et al., 2006). Also, an immunoglobulin-CDR3 sequence was fully grafted into one of the Neocarzinostatin (NCS) loops for future therapeutic applications (Nicaise et al., 2004).

Many of the protein scaffolds discussed in Chapter 1 have varying secondary and tertiary structures and consequently not all oligomeric proteins have the desirable features for creating new binders. Therefore, proteins with weak secondary structures are not suitable for the development of new molecular scaffolds. The ideal scaffold must, therefore, have a rigid framework with several regions that have different secondary structural conformations that can be substituted for numerous peptides or utilised in the development of a peptide library where the substituted regions are intentionally diversified.

Cpn10 is an essential oligomeric mitochondrial protein that assists in folding of translated polypeptides or refolding of denatured proteins (Hartman et al., 1992, Martin et al., 1993, Pilkington and Walker, 1993, Taneja and Mande, 1999, Hartl and Hayer-Hartl, 2002, Perham and Wittung-Stafshede, 2007). Cpn10 is structurally conserved across many species (Hunt et al., 1996, Xu et al., 1997, Roberts et al., 1999, Luke et al., 2005b, Luke et al., 2005a, Numoto et al., 2005) and the quaternary structure (70 kDa) is composed of seven identical monomers as shown in Figure 3.1. Each monomer consists of a β-hairpin roof loop and a longer flexible

mobile loop that extends below the core of the protein. Importantly, these loops are interconnected by a structurally rigid hydrophobic β-barrel core (Guidry et al., 2000, Higurashi et al., 2003, Lin et al., 2004). Prior to forming a complex with its cognate Cpn60, i.e. for assisting protein folding (Shimamura et al., 2004), Cpn10 is assembled into a heptameric structure through a strong hydrogen bond between the β-barrels of contiguous monomers (Sakane et al., 2004, Sakane et al., 2007) (Figure 3.3). The mobile loop is highly flexible in structure and is known to play a crucial role in interacting with Cpn60 (Richardson et al., 2001, Lin et al., 2004).

The Cpn10 core is composed of hydrophobic amino acids while the residues on the surface are mostly hydrophilic. In the Cpn10 protein core, the folding process is driven mainly through the collapse of hydrophobic residues while heptamer formation is dependent on hydrogen bonding between the adjacent monomer interfaces. Therefore, a well-structured hydrophobic core is crucial for the stability of the protein (Lazar et al., 1997, Street and Mayo, 1999). In addition, the protein surface plays a crucial role in its stability and for that reason it is important to ensure that insertion of various functional peptides does not compromise the core structure of the protein. In Cpn10, the main source of β-sheet stability is via hydrogen bonding and a subsequent alignment with adjacent antiparallel β-strands. The interaction between the antiparallel pairings of the β-strand interface of each monomer, and the inter-protein interactions govern over 75 % of the Cpn10 heptamer configuration stability (Guidry et al., 2003, Luke et al., 2006, Aguilar et al., 2011). Studies conducted by Sakane et al. (2004) determined that mutations in various residues such as *Ile3* and *Val95* of *N*- and *C*-terminus, respectively, did not affect Cpn10 refolding. Interestingly, oligomeric Cpn10 is one of few proteins that can fold reversibly *in vitro* despite the introduction of various mutations (Perham and Wittung-Stafshede, 2007) and this is an important aspect in molecular folding and protein assembly for an advanced scaffold protein design.

The mobile loop within Cpn10 exhibits a large variation in structural conformations as this region lacks a defined secondary structure. However, mobility is restricted only when the Cpn10 heptamer is in contact with Cpn60 (Richardson et al., 2001, Higurashi et al., 2003). Human Cpn10 mobile loop consists of 18 residues (*AAETVTKGGIMLPEKSQG)* while the β-hairpin roof loop is composed of 12 amino acids (*GSKGKGGEIQPV)* (Figure 3.3). In general, the two loop turns (i.e. roof & mobile) that join the β-strands might play a crucial role, whereby modification, substitution, or mutation of any of these residues may seriously

affect the overall protein stability. However, in practice, the substitution or exclusion of the mobile loop (i.e. ΔL-Cpn10), the β-hairpin loop (i.e. ΔRf-Cpn10), or both loops ΔL-ΔRf-Cpn10 (i.e. β-barrel) had no significant impacts on assembling of mutant-hCpn10 into heptamer configurations (Figures 3.4 and 3.5).

As with other molecular scaffolds, loops form an extended interface that facilitates binding towards a specific target molecule (Chapter 1). Therefore, in these particular regions, the loops can be substituted or extended with a functional peptide domain which may vary in sequence and length. In this study, hCpn10 has been identified and proposed as an ideal molecular scaffold, whereby peptides and/or antibody fragments may be substituted into the mobile loop for various medical and biotechnological purposes (Table 3.1).

Normally, information on the structure of various proteins can be acquired through X-ray diffraction, nuclear magnetic resonance (NMR), Förster resonance energy transfer (FRET), or other biophysical methods (Higurashi et al., 2003, Perham and Wittung-Stafshede, 2007, Buehler et al., 2010, Wu et al., 2012). These methodologies yield valuable structural information that can be used for various applications including drug design and anti-viral agents among others. Nevertheless, these methods have limitations in terms of applicability and time scales.

MD simulation is an *in silico* technique that is routinely used to study biologically important macromolecules such as peptides, proteins, and nucleic acids at the atomic level (Su and Wang, 2009, Mitchell et al., 2009). These simulations are conducted at a time scale ranging from femto- to micro-seconds (Lindahl, 2008). The molecular dynamics are then numerically solved using equations of motion to follow the time evolution and thus determine the kinetic and thermodynamic properties of such molecules (Petrenko and Meller, 2010). The accuracy and level of details acquired from MD simulations are dependent on the size of the biomolecule and the timescale that is used. Given the advances made in computer technology and enhanced algorithmic design, MD simulation approaches towards complex biological systems (e.g. protein complexes) have become an increasingly important asset in many research areas. Hence, the MD simulations have altered the interplay between theory and experiment, providing a basis that aids in the understanding of various molecules that could otherwise not be obtained through other physicochemical techniques. Importantly, examining

the properties of biomolecules by using MD simulation is faster and more cost effective than synthesizing and characterizing it in a real time experiments.

The experimental results in conjunction with the computer modelling data provide a basis for understanding the chemical machinery of various proteins. Details of the structural conformation and molecular interaction of hCpn10 and the scaffold-based hCpn10s needed to be well understood in order to stabilize the protein. Therefore, MD simulations were utilised in order to evaluate how mutations could affect the structural function of the mutant-proteins. This approach was used to identify the domain(s) and/or residue(s) of the hCpn10 that are most likely to accept mutations with maximum stability and thus, identify the correct mutation sites and cluster domains that are important for the development of other hCpn10-based NMEs.

This study was focussed on integrating the data acquired from the MD simulations as well as the subsequent experimental validations. The details on implementing mathematical and physical principles are well documented in molecular modelling literature (Leach, 2001, Frenkel and Smit, 2002). The choice of modelling methods was based on a practical problem, specifically the experimental failure of hCpn10-based variants to develop the proper configurations and to find out why heptamers were not formed in practice. Use of state-of-art simulation techniques have provided more details on the structure and dynamics of hCpn10s that could not be obtained from experimental analyses alone. The MD simulations of the hCpn10 variants provided significant insight into the dynamics of the molecules in order to address a number of important questions. Firstly, how the non-native peptide (in Table 3.1) inserted at the mobile loop domain actually behaved and transformed? Secondly, which residue(s) of this inserted peptide and/or of the hCpn10 molecule play a major role in heptamer stability? Thirdly, despite the stability of native hCpn10, why does/do certain mutation(s) affect the stability of the overall structure? And importantly, is there a conserved heptamer configuration of hCpn10-based NMEs?

The MD simulations generate a number of trajectories, which are a set of atom coordinates. These trajectories are then used for calculations that determine the distance between atoms and/or residues, the prevalence of hydrogen bonds, occurrences of disulfide linkages, structural modifications and/or degradations (i.e. secondary, tertiary, and quaternary configurations), and the magnitudes of binding sites. These observations essentially provide

further biological insights into the molecule that was simulated (Figure 3.7). In addition, parts of the trajectories were highlighted by performing additional simulations in full atomic detail in order to shed light on the interesting parts of the hCpn10 trajectories (Figures 3.7 to 3.9).

Hence, this study focussed on simulating hCpn10 and comparing its dynamics with its mutant variants (Table 3.1). Following the iterative guide-cycle MD simulations were performed in two stages (i.e. Phase I and II) based on the problem in hand (Figure 3.6). The first phase of the MD simulations was performed on two of the original proposed hCpn10 variants that contained either the E-76 or Myc- peptides at the mobile loop domains. The structures and kinetics of folding were studied using molecular simulations in conjunction with free energy and global optimization methods. A number of simulations were conducted at temperatures of 310 K in order to enhance the rate of hCpn10 denaturation, which in turn provided further insights into the forces that stabilized molecule. Furthermore, during the simulated refolding of hCpn10, kinetic intermediates were identified and further studied.

The initial stages of folding were directly observed during the 10 ns MD simulations for hCpn10 (101 residues) without explicit water (Figure 3.7). This illustrated that both the top and mobile loops of hCpn10 were flexible at 310 K. However, none of these loops were flexible enough to interact with the core structure of hCpn10. These findings confirmed that the dimerisation interfaces of hCpn10 were not perturbed in the monomeric form of the protein (Figures. 3.7 to 3.9).

The electrostatic potential is thought to be the second most important factor of investigative interest. Naturally, the electrostatic interactions play an essential role in the structure and function of proteins, DNA, cell membranes, and other biomolecules. One important aspect for understanding these interactions is that biologically active molecules are always surrounded by water molecules and dissolved ions, adding more degree of freedom of these molecules in solvent. The electrostatic potential can be projected on a molecular surface for the identification of possible binding sites. Also, it is important in determining the induced charge distribution on the molecular surface which accounts for the change in polarization charge across the dielectric boundary. Therefore, the electrostatic potentials of hCpn10 monomers were determined and the electrostatic potentials generated by hCpn10 monomer interfaces were illustrated to be complementary and further supported the development of a heptamer conformation (Figure 3.10).

One of the major problems, of which comprehensive data is available, was the series of events that hCpn10 undergoes upon binding with a contiguous monomer. Studies at a 1 ns relaxation of the hCpn10 illustrated the kinetics of cluster motif "*RSA*/*GKV*" rebinding and the motion paths of the mobile loop, which in turn, added a further understanding of the discriminating mechanisms that could prevent dysfunctional binding of alternative sequences. The sequence motif "*RSA*" and "*GKV*" are located at *N*- and *C*-termini of the mobile loop, and were retained in a few of the hCpn10 variants (Figure 3.9). The packing of the amino acids within the two motifs improved during the MD simulations, indicating that the fold of the native mobile loop is optimal for a heptamer. The hCpn10 monomer seemed to prefer a tightly packed configuration where the *Lys39* and *Arg19* side chains can optimise the polar interactions within the loop as well as the hydrophobic contacts of their aliphatic chains.

Phase I MD simulations were performed on hCpn10 variant termed CE76, whereby hCpn10 mobile loop was substituted with an anticoagulant E-76 peptide. Simulation data indicated that CE76 failed to develop into a heptamer (Figure 3.22). MD simulation of CE76 monomer confirmed an overall stable molecule during the 1 ns simulation. However, the mobile loop, which contained the E-76 sequence (*AL**C**DDPRVDRWY**C**QFVEG*) and a disulphide bond between *Cys23* and *Cys33* residues, demonstrated numerous conformational changes from the initial model whereby the E-76 loop contacted the surface of the core protein formed by the β-barrel like structure. This clearly indicated that the E-76 loop would block interactions between different monomers and therefore, heptameric assembly. A number of critical interactions did occur within the E-76 loop where polar residues *Arg30* and *Asp24* interacted with *Glu74* of the β-core, and residue *Glu37* interacted with *Lys79* of the β-core. In addition, weak hydrophobic interactions did occur between residues *Val100*, *Leu71*, and *Gln42* of the β-core and the CE76. Due to the flexible nature of the E-76 loop, collision occurred more frequently with β-strands 4 and 5 within the β-core of the CE76 model. In addition, an observed distortion on the CE76, relative to hCpn10 occurred at the dimerisation interfaces (Figure 3.10B). A significant finding was that there were two negatively charged regions near the binding site, indicating a lower electrostatic potential for dimerisation of two (or more) CE76 monomers. Therefore, due to the collision of the E-76 loop with the protein core, and the reduction of electrostatic potential could possibly explain the inability of CE76 to form a heptamer *in vitro* (Figure 3.22).

MD simulations (I) were applied to second variant called CPMyc, which contained the Myc-peptide within the mobile loop region of hCpn10 (Table 3.1). Although Myc-tag peptide is typically used for purifying recombinant proteins, it was utilised in the MD simulations as the peptide contained a number of charged and polar residues. Also, the success of CPMyc as scaffold molecule in binding to 9E10 antibody could open a future window in the design of vaccines, by using the avid hCpn10 to display different epitope(s) for quick immune response against a specific illness. Under these circumstances, the Myc-tag has been proven beneficial in gaining a further understanding of the dynamics of hCpn10-based variant when a peptide is inserted into the mobile loop. Following the same methodology used in the previous simulation of CE76, the CPMyc monomer was in general a stable molecule, however the substituted mobile loop with the Myc-tag was found to be highly flexible with a continuous change in conformation during 1 ns simulation (Figure 3.12). The high flexibility of the Myc-loop was likely due to the high contents of charged residues (i.e. ***E****QK*LIS***EED****L*NG) at the substituted loop domain. Furthermore, the MD simulations showed that the Myc-loop was interacting with the core β-barrel of contiguous CPMyc monomer where β-strands 4 and 5 of the protein were moving towards the mobile loop and disrupting the dimerisation interfaces of the adjoining monomers (Figure 3.12). Based on the simulation, the CPMyc model was unlikely to form full-size heptamers which was clearly confirmed through *in vitro* experiments.

The second phase of the MD simulations was conducted on the modified CE76 and CPMyc variants. Following on from the investigations made on using hCpn10 for displaying various peptides, four redesigned models were evaluated in this second phase (II). The modifications were made by inserting the various linkers in Table 3.2 between the native hCpn10 (β-barrel core) and the insertion sites of the non-native peptides (at *N*- and *C*-junctures of the mobile loop). The inserted linkers were enriched with alanine(s)/proline(s) residues, however, in the case of the E5-linker used for the E-76 peptide five native loop residues were retained at the *N*- and *C*-termini, respectively. The key aim of using these linkers was to apply a certain degree of extinction and rigidity in the inserted peptide sequences, and prevent interaction with the β-barrel core of the hCpn10 protein. The importance of inserting linkers with multiple proline (*Pro*) residues was to impose some constraint and thereby reduce the entropic rate of the inserted non-native peptides, due to the nature of proline that provide rigidity to polypeptide chain by imposing certain torsion angles on the segment of protein

structure, which play crucial roles in protein folding (Ho et al., 2005, Eyles and Gierasch, 2000, Shi et al., 2002, Bhat et al., 2003, Greenfield, 2006).

Phase II of the MD simulations was performed in three different modes (as simulation 1, 2, and 3) for each variant. $CE76_{P1}$ is the first redesigned model to be examined where the CE76 was modified by inserting a P1-linker (Table 3.2). MD simulation 1 of the $CE76_{-P1}$ indicated that the P1-linker could likely stabilize the E-76 loop and prevent interactions with the protein core (Figure 3.13). In contrast to the unmodified CE76 model, the $CE76_{P1}$ core was more stable and no significant distortions were observed for any of β-strands 4 and 5. For enabling the purification process, a *His*-tag was *C*-terminally inserted. However, due to the high flexibility of this tag, there was a possibility of interacting with the β-barrel core which might hinder the heptameriztion development which is contrary to hCpn10. However, it could also probably prevent or reduce the interactions between the E-76 insert and the protein β-core. The electrostatic potential generated by the $CE76_{P1}$ monomer interface was not altered by the addition of a P1-linker, but it did support heptamerisation due to a slight compatibility in surface charges. Overall, the ability of the $CE76_{P1}$ model to form a heptamer was apparent, and further verified *in vitro* when the $CE76_{P1}$ model produced a fully functional heptamer following a buffer screening approach for detecting optimum conditions (i.e. $CE76_{P1}$ ~ isoelectric point (pI) 7.1). The $CE76_{P1}$ was stable and functional at low pH 5.5, which was resolved by UV detections of SE-HPLC (Figure 3.22) and *in vitro* activity determinations (Chapter 4).

$CE76_{P2}$ was a modified construct that contained a P2-linker (Table 3.2) and underwent a series of simulations. MD simulations on $CE76_{P2}$ were performed in three environments as described above. These simulations verified that the P2-linker independently succeeded in stabilising the E-76 inserted loop. The P2-linker prevented any interaction of the inserted-loop with the protein core and therefore allowing heptamerisation to occur in a similar format to hCpn10. The $CE76_{P2}$ β-barrel core was stable during the MD simulations and there were no major distortions observed in any of the β-strands as observed in the unmodified CE76 model. The $CE76_{P2}$ was also *C*-terminally *His*-tagged, however, unlike the $CE76_{P1}$, simulations determined that this tag did not interact significantly with the E-76 insert (Figure 3.15). In addition, the distributions of electrostatic potential that were generated by the $CE76_{P2}$ contiguous monomers supported heptamer formations (Figure 3.14B). Consequently, the P2-linker in $CE76_{P2}$ model was predicted to form a heptamer. Remarkably, the $CE76_{P2}$

was expressed and purified as a full heptamer, whereby the UV profiles of $CE76_{P2}$ illustrated the stability of the heptamer configurations (Figure 3.22).

Third examined model variant was the $CE76_{E5}$ where the E-76 loop was inserted further downstream of the mobile loop between *Val25* and *Glu34* (Table 3.2). MD simulations illustrated that the E5-linker could not prevent interactions between the highly polar E-76 loop and the protein β-core in all simulation conditions. An example is, in a dielectric environment of ε=40, it did not shield the electrostatic interactions sufficiently enough to prevent the interactions between the polar E-76 loop and the protein β-core. The heptamerisation of the $CE76_{E5}$ might therefore be dependent on optimising the buffering conditions, for example, maintaining a pH level that was utilised for the $CE76_{P1}$ model. The E5-linker sequence "*AETVT/PEKSQ*", of hCpn10 not only failed to prevent interactions between the mobile loop and the β-core, but furthermore, the linker sequences distorted the structure of the $CE76_{E5}$ model that was significantly distinct from hCpn10 (Figures 3.16 and 3.17). MD simulations of $CE76_{E5}$ also illustrated that both *N*- and *C*-termini were highly flexible, however the *His*-tagged *C*-terminus did not interact significantly with mobile loop despite its high mobility. Nonetheless, the charged compatibility that was generated by the $CE76_{E5}$ monomer interface supported heptamer formation (Figure 3.14C). Generally, the potential of the $CE76_{E5}$ variant (pI ~ 6.65) to develop into a heptamer is not certain at this stage and will most likely be dependent on the protein media and storage conditions (e.g. adjusted pH level). *In vitro* studies of the $CE76_{E5}$ model, however, demonstrated that the protein was unstable and formed soluble aggregates and other configures as clearly shown in its UV profile of SE-HPLC despite the medium pH was adjusted for optimum (Figure 3.22).

The last redesigned model to be examined was the $CPMyc_{P1}$ whereby the Myc-loop was stabilised by single proline-linker (P1) (Table 3.2). The MD simulations of the modified $CPMyc_{P1}$ variant illustrated that the protein was stable overall with no apparent distortions in any of the core β-strands. MD simulations indicated that the P1-linker could independently stabilise and retain the inserted Myc-loop from interacting with the protein core (Figure 3.18). This is, however, different from the simulated $CE76_{P1}$ model and this is possibly due to the shorter size of the Myc-peptide (12 residues). Similar to other models, the MD simulations of both the *N*- and *C*-termini were highly elastic and the *His*-tag did loosely interact with the Myc-loop. Conversely, this might be necessary in order to stabilise the β-barrel core. The removal of the *His*-tag from the *C*-terminus during simulations with a lower

dielectric constant (ε=40), however, eventually distorted the CPMyc$_{P1}$ model. Interestingly, the electrostatic distributions that were generated by the CPMyc$_{P1}$ adjacent monomers at the dimerisation sites might in fact hinder oligomerisation and heptamer formation. This could be due to the different conformational folds of the *N*- and *C*-termini (Figure 3.14D). Generally, the CPMyc$_{P1}$ variant is likely to form a heptamer if the *His*-tag can adopt a conformation where it does not interfere with dimerisation between adjacent monomers. Yet, the recombinant CPMyc$_{P1}$ was successfully formed a heptamer as evident in the SEC profile measured at 280 nm, despite the unsupportive MD data of the distribution of the electrostatic potential (Figure 3.23). This, however, could be more relevant to the greater choice of buffering systems, rather than the dependent on the molecular behaviour alone for the development of heptamer configuration.

A force-mode approach is an alternative approach and was applied in order to improve the development of heptamer for the scaffold based-hCpn10 variants. This method is based on genetically designing the hCpn10 variants with force linking six of the seven contiguous monomers with three glycine residues (~ 3.4 Å) from the *C*- to *N*-termini. About 3 Å is the distance between adjacent monomers termini. Therefore, 3x*Gly* would have no or minor affect in molecular assembly. This approach was applied to the original construct of CE76 (Table 3.1) called FH-CE76**,** and similarly applied to the native hCpn10 termed FH-Cpn10. Despite the success of this approach in forming a heptamer for both FH-Cpn10 and FH-CE76, there were technical drawbacks in this regard, particularly with the FH-CE76. Firstly, the large molecular size of 84 kDa as shown in Figure 3.19 meant that such a protein would more than likely to form IBs (i.e. aggregates). Secondly, extracting FH-CE76 from the IBs as unfolded proteins, followed by subsequent refolding via buffer exchange would essentially be a time consuming exercise. Lastly, the FH-CE76 was unstable and formed soluble aggregates following purification and therefore this protein could not be analysed via SE-HPLC. Therefore, a forced-heptamer approach cannot be applied for a molecule like CE76. However, the FH-Cpn10 was more stable with similar properties and characteristics of native hCpn10. Thus, the forced-heptamer approach is still applicable however in a molecular based manner.

3.7 Conclusion

The aim was to establish a fundamental basis for designing a molecular scaffold based on hCpn10 for future therapeutic and diagnostic/analytical applications. Computational modelling is an advanced technology that provides the basis in understanding the kinetics and dynamics of various biomolecules. This is particularly important as it provides a basis to effectively engineer the hCpn10 NMEs with various functional peptides that will eventually form a heptamer in similar conformation to its wild-type counterpart. The molecular modelling simulations of all binders based on hCpn10 were slightly similar to the hCpn10. While the β-barrel architectures were retained in both native hCpn10 and its most variants, the mobile loop exhibited a large conformational change, most likely as a direct result of the side chain replacements with either charged or hydrophobic residues. This observation reveals remarkable structural plasticity of the inserted loops such as E-76 peptide and the Myc-tag which in turn prevented the dimerisation that was essential for heptamer formation. Remarkably, such uncontrolled plasticity of the mobile loop had been successfully retained and minimized through the insertions of various linkers at the *N*- and *C*-junctures of the inserted-loop, most likely due to the proline(s) content of those linkers. Subsequently, all computationally predicted models of Cpn10 based binders were synthesized and validated *in vitro*. The use of the P1- and P2-linkers, containing several alanine/proline residues, restricted the mobility of the inserted loop, and furthermore, facilitated the formation of a heptamer. The P2-linker would be more suited for higher molecular weight peptides that are inserted into the mobile loop. Contrary to the P1- and P2- linkers, E5- was ineffective in conferring an overall stability in the protein. Although the mutant hCpn10 variant of $CE76_{E5}$ formed a heptamer, the protein rapidly degraded following purification. This could be due to significant alterations occurring within the protein core and the loss of the critical β-stands which are important for conferring stability within the molecule. This was clearly predicted by the MD simulations and revealed experimentally.

This study showed the potential of hCpn10 to be modified with the promise to be used as NME scaffolds and has proven that MD simulation is a speedy and suitable tool to assist in such molecular design.

Chapter 4. Development of a Factor VIIa Inhibitor Based on a High Affinity Peptide Ligand Displayed on Human Chaperonin-10 Molecular Scaffold.

4.1 Introduction

Advances in recombinant DNA technology have facilitated the production of large quantities of various peptides and proteins for the treatment of a variety of disease indications. There are more than two hundred approved biologics by FDA used for the treatments of various diseases that have been developed through recombinant DNA technology (CDER, 2014), including a few that are approved for in the management of the blood coagulation cascade including tissue factor pathway inhibitor (TFPI) and rNAPc2 (Enjyoji K et al., 1995, Bergum et al., 2001).

At present the underpinning approach towards managing the clotting process is using small molecule, protease inhibitors, that directly bind to the enzymatic active site (Leung et al., 2000). Recently, a serine protease inhibitor that selectively targets the extrinsic coagulation cascade (i.e. the TF-dependents) was isolated from a random peptide phage display library, hence heralds a new class of serine protease inhibitor, named as an E-series peptide (Dennis et al., 2000). E-76 peptide, derived from the parent E- series, predominantly binds to the exosite of Factor VIIa (FVIIa) with high affinity. This in turns causes a large conformational change within the active site, which subsequently prevents the activation of Factor X (FX) to FXa, and thus blocks the initiation of the extrinsic pathway.

In general, therapeutic peptides are small in size (≤ 20 kDa), may known to have potent biological activity, and some with high binding affinity towards a particular molecular target *in vitro*. Nevertheless, peptides administered *in vivo* have a number of limitations, one of which is due to their small size, resulting in a shorter half-life as they are rapidly cleared by the renal system (Diao and Meibohm, 2013). Furthermore, peptides are highly susceptible to protease degradation, and as such, multiple doses of the peptide are required if there is to be any potent therapeutic activity *in vivo* (Sato et al., 2006). In order to address these limitations, peptide fragments can be introduced into protein scaffolds through recombinant DNA technology in order to increase the half-life, stability of the peptide, and thus prevent

proteolytic degradation. These protein scaffolds include such entities as DARPins, lipocalins, and others which are discussed in Chapter 1.

The research described in this thesis focuses on hCpn10, an essential heptameric mitochondrial protein that facilitates the proper folding of translated polypeptides as a molecular scaffold. The hCpn10 (70 kDa) of seven identical monomers each is composed of two flexible loops, namely, the β-hairpin roof and mobile loop, and a rigid hydrophobic β-barrel core. As with other molecular scaffolds, the loops have an extended interface that facilitates binding towards the target molecule (Korndorfer et al., 2003).

Based on the properties of the hCpn10 molecule, the research described in this chapter focussed on utilising the hCpn10 as NME displaying the E-76 peptide (termed CE76) at the mobile loop site (Figure 4.1). Optimisation of CE76 with respect to stability was performed through molecular dynamic simulations with the insertion of a number of proline residues, as discussed in Chapter 3. Herein, based on experimental studies, $CE76_{P1}$ was the most suitable structure to be further characterised. $CE76_{P1}$ showed remarkable inhibitory properties towards FVII activation, and subsequently reduced the activation of FX, with IC_{50} of 6.8 nM and 1.5 nM, respectively. Also, $CE76_{P1}$ found to prolonged the TF-dependent coagulation pathway by 7-fold in PT assays. In significant, the NME $CE76_{P1}$ possessed a potent anticoagulant property that could be applied as future biologics for the treatment of blood disorders.

4.2 Research aims

The overall aims of the work presented in this study chapter are:

- To demonstrate proof of concept by designing the CE76 as a therapeutic biologics agent based on the hCpn10 molecule; and
- To conduct *in vitro* biological activity studies on the CE76 in relation to the E-76 peptide.

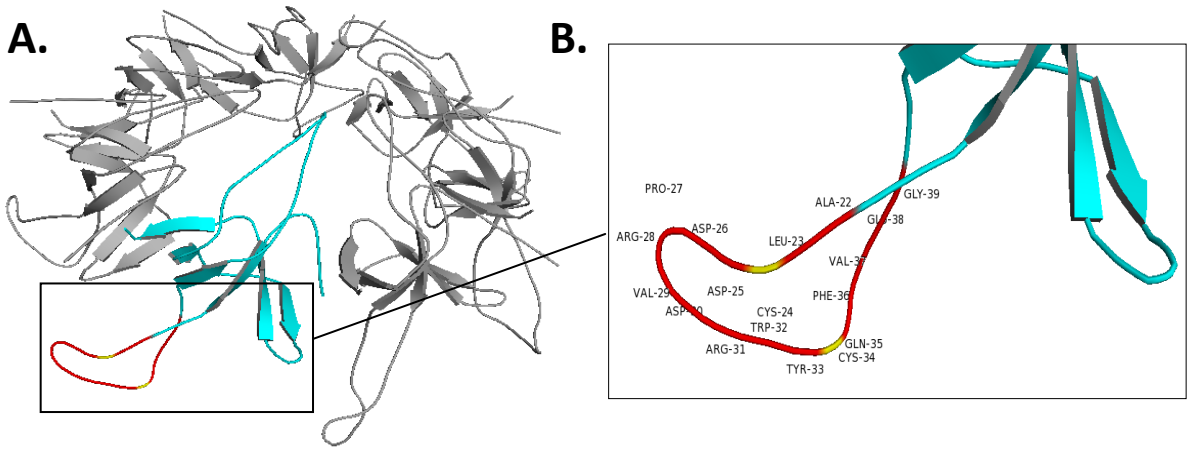


Figure 4.1 Three-dimensional model of CE76

3D homology model of CE76 shown in ribbon representation **(A)** Model of full size CE76 heptamer; each monomer consists of main hydrophobic β-barrel core and two loops (top β-hairpin and bottom mobile loop). **(B)** Substituted mobile loop with E-76 peptide (red, 18 amino acids) between at *Ala21* and *Gly38*, a disulfide bond form between *Cys24* and *Cys34* (yellow), refer to text for full details. The homology model based on wt Cpn10 (PDB ID. 1WE3) as detailed in Chapter 3. This was modelled using PyMOL software.

4.3 Materials and mMethods

4.3.1 Materials

Table 4.1 details the relevant materials used in the research discussed in this study and their manufacturer/supplier.

Table 4.1 List of materials for the research study

Item(s)	From
Proteins/Peptide:	
NovoSeven®RT, Factor VIIa (rFVIIa)	NovoNordisk®
Tissue Factor (rTF)	Hyphen BioMed™
E-76 peptide (Biotin-*ALCDDPRVDRWYCQFVEG*) control	Auspep™
Antibodies:	
mAb-Mouse Anti-FVII Human	Sigma-Aldrich®
HRP-Goat Anti-Mouse	Novex®, *Life* technologies™
HRP-conjugated Mouse Anti-6xHis	Novex®, *Life* technologies™

Item(s)	From
Chemicals & Reagents:	
Thromborel®S	SEIMENS™
TriniCLOT™ aPTT HS	TrinityBiotech™
Chromozym *t-PA*	Roche™
S-2765™ (chromogenic substrate)	*CHROMOGENIX* ™
S-2222™ (chromogenic substrate)	*CHROMOGENIX* ™
3,3′,5,5′-Tetramethylbenzidine (TMB)	Sigma-Aldrich®

4.3.2 Protein expression, purification, and characterisations

All CE76 variants in this study were expressed as soluble protein and purified (≥ 95%) according to Chapter 2: General Materials and Methods.

4.3.3 Protein binding detections by ELISA

A SpectraMax-M4 plate reader was used to measure the binding of CE76 variants in a 96-well miroplates. A typical microtitre plate was coated overnight with 100 nM rFVIIa (NovoSeven®RT), diluted in PBS, 0.05 % (v/v) Tween 20 (PBST, pH 7.4), total of 200 µL per well. The ELISA plate wells were blocked by PBST, 2 % (w/v) non-fat dry milk, followed by three rounds of washes with PBST for removal of unbounded rFVIIa. Purified *His*-tagged CE76 and its variants (summarised in Table 4.2) were incubated in various concentrations (from 0.1-100 nM), and in parallel with *His*-tagged hCpn10 (control), and mAb-mouse anti-FVII human (control), all were incubated for 1 hour at 25 °C. After incubations, plates were washed with PBST for three rounds to confirm the removal of unbound proteins. The histidine-tagged proteins were detected by 1:5000 dilution (in PBST) of either HRP-conjugated mouse anti-6xHis antibody or HRP-goat anti-mouse antibody, and further incubated (in dark) for 30 min at RT, followed by three rounds of washes with PBST, then single wash with PBS buffer prior to developing the reaction with 100 µL of 3,3′,5,5′-Tetramethylbenzidine (TMB, Sigma-Aldrich®) per well. The enzyme reaction terminated after 15 min with 50 µL of 1 M H_2SO_4. Reaction results were measured at 450 nm (A_{450}) using spectrophotometer. Data of blank well (absence of rFVIIa) was used for adjusting background activities and data was fitted by non-linear regression from triplicate independent experiments.

4.3.4 Binding affinity and kinetic analysis by Bio-layer Interferometry (BLI)

Binding affinity and kinetic profile measurements were conducted using BLI Octet® Red instrument (ForteBio). The buffer contained 50 mM Tris-HCl (pH 7.6), 150 mM NaCl with 5 mM $CaCl_2$ and was used for all dilutions and monitoring processes. Purified *His*-tagged CE76 variants were captured via Ni-NTA biosensors, such that the response due to immobilization was approximately 1-2 nm. Biosensors with captured proteins were then dipped into multiple concentrations of rFVIIa (0, 0.1, 0.5, 1, 5, 10, 25, 50, 75, and 100 nM), to measure the association phase, with 1000 rpm of plate shaking. The dissociations phase was monitored in buffer only. All data were fit to 1:1 or 1:2 binding model of stoichiometry proteins by global fitting of multiple kinetic traces, and analysed by Data Analysis 7.0 software. Data were monitored for eight different concentrations simultaneously, and zero nM/control trace was used for blank run and subtraction. The kinetic measurements were the average of triplicate measurement sets.

4.3.5 FVII and FX enzyme activity assays

The amidolytic activity of FVIIa was measured using the chromogenic substrates Chromozym *t*-PA or Spectrozyme FXa, as previously described (Andersen et al., 2012). In brief, a 96-well plate was blocked overnight in 200 µL Assay Buffer (AB) containing; 20 mM HEPES (pH 7.8), 150 mM NaCl, 5 mM $CaCl_2$, 0.1 % PEG 8000, 0.5 % BSA, at 4 °C. The amidolytic activity was determined in total reaction of 200 µL in AB, by incubating 50 nM or 100 nM rFVIIa with 0.5-500 nM TF (Thromborel®S), for 30 min at 37 °C. Chromogenic substrate was added to a final concentration of either 1 mM Chromozym *t*-PA, 1 mM S-2288 chromogenic substrate, or 3 mM S-2288 for assay without TF. The initial reaction rates for substrate hydrolysis were monitored at 405 nm for 20 min using Multiskan™ GO Microplate Spectrophotometer (Thermo Scientific™), at RT, and the rates were determined by linear regression. Similarly, incubation of TF with substrate in the absence of rFVIIa was used as control.

FX activation by TF-FVIIa complex was also measured using chromogenic assay as described (Andersen et al., 2012). The activity of FX was determined by incubating 100 µL AB containing 200 nM FX with 100 nM rFVIIa, or with 10 nM rFVIIa and 200 nM TF

(Thromborel®S), incorporated into phospholipid as recommended (Kelley et al., 1997) of 6 µL TriniCLOT™ aPTT HS, for 1 hr at 37 °C. FX activity was impeded by adding 5 mM EDTA. The amount of FXa generated was measured relative to hydrolysis activity of 0.5 mM S-2765 chromogenic substrate, at 405 nm by using Multiskan™ GO Microplate Spectrophotometer, at RT.

4.3.6 FVIIa and FX inhibition assays

Inhibition of amidolytic activity of FVIIa by $CE76_{-P1}$ was determined as a function of inhibitor concentrations, in the presence and absence of TF. In summary, a 0.001-1000 nM (in 10-fold series dilution in AB) of $CE76_{P1}$, was incubated with 10 nM rFVIIa and 100 nM TF, for 10-20 min, at 37 °C, prior to adding 1 mM Chromozym *t*-PA or 1 mM S-2288, then measuring the initial reaction velocities at 405 nm by Multiskan™ GO Microplate Spectrophotometer, at RT. The linear rate of increased absorbance was expressed as fractional activities (V_i/V_0), in which the velocity of the substrate hydrolysis in the presence (Vi) or absence (V_0) of inhibitor $CE76_{P1}$. The relative velocities (V_i/V_0) were plotted against the $CE76_{-P1}$ concentration (nM).

Inhibition of FX activation by TF-VFIIa complex was performed as described (Sen et al., 2011). FX activation by TF-FVIIa was determined as a function of $CE76_{P1}$ (inhibitor) concentrations. A total reaction volume of 100 µl AB containing 175 nM FXa with 10 nM rFVIIa, and 150 nM TF was incorporated into phospholipid 6 µL TriniCLOT™ aPTT HS, and incubated with $CE76_{P1}$ (0.01–1000 nM) for 30 min at 37 °C, prior to adding 1 mM S-2765 chromogenic substrate. Controls were made in the absence of $CE76_{P1}$, rFVIIa, or FX. The rate of substrate hydrolysis was measured at 405 nm over 15 min by Multiskan™ GO Microplate Spectrophotometer at RT. The rate of FXa generation was expressed as fractional activities (V_i/V_0), as previously described. All experiments were performed in triplicate (Appendix 4).

4.3.7 Plasma clotting time (PT/aPTT) assays

The prothrombin time (PT) and activated Partial Thromboplastin Time (aPTT) assays (detailed in Figure 4.2) were performed in pooled citrated-treated human plasma from healthy volunteers (n=5), as described (Lefkowitz, 2013). Clotting times were determined as a

function of protein concentrations at 37 °C in triplicate, using Hyland-Clotek™ clotting analyser. The recalcification time of pooled plasma was 2-4 minutes. PT and aPTT assays were triggered by using Thromborel®S (human placental thromboplastin) and TriniCLOT®aPTT HS (phospholipids mixture) for initiating coagulation of extrinsic and intrinsic cascades, respectively.

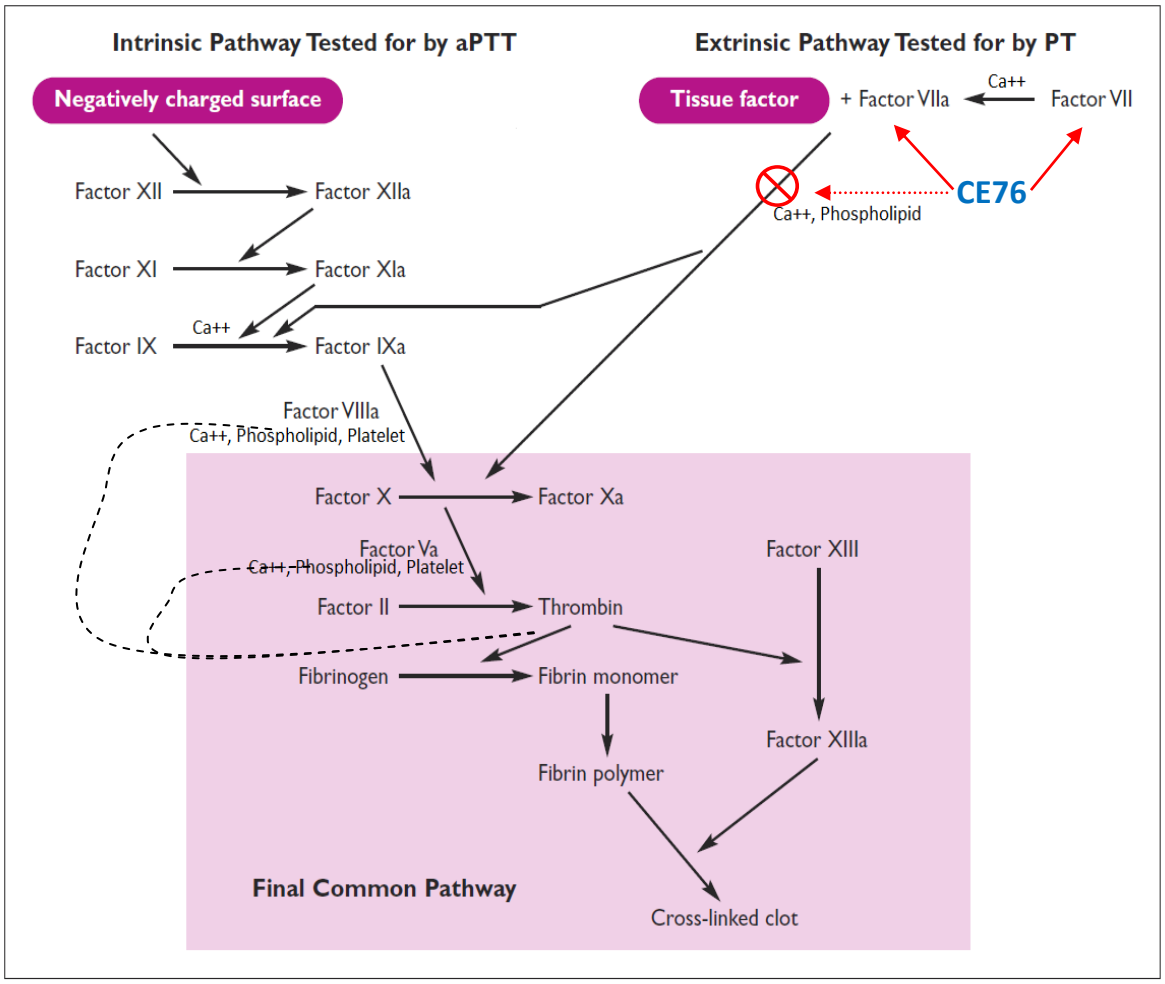


Figure 4.2 Normal blood coagulation cascade and pathways PT/aPTT tests

PT/aPTT tests are based on the type of initiations. The solid arrows indicate activation of a coagulation factor by the predecessor factor while the dotted lines indicate activation of FV and FVIII by thrombin. The extrinsic pathway is initiated upon vascular injury or trauma. This in turn leads to exposure of tissue factor (TF) that binds to phospholipids/Ca^{2+}, and subsequently activates FVII (i.e. FVII → FVIIa). The intrinsic pathway is initiated when blood makes contact with the damaged vessel (i.e. exposed to negatively charged surfaces) resulting in the activation of FXII (i.e. FXII → FXIIa). The two pathways independently lead to the activation of FX to FXa which subsequently leads to the conversion of prothrombin to thrombin through a hydrolytic reaction. Fibrin polymer (i.e. fibrin-clot) is the end result of the blood cascade which is synthesised via thrombin-mediated conversion of the fibrin monomers. NME CE76 act as anticoagulant and solid arrows (red) indicate binding to both

FVII and FVIIa while the dotted lines (red) indicate inhibit extrinsic activation of FX. Image adapted from Lefkowitz (2013).

PT assay, in brief, was performed by adding 100 µL citrated plasma to a glass clotting tube and incubated for 10 min in the heating block of Hyland-Clotek™, at 37 °C. CE76 variants (50 µL, as inhibitors) were added in various concentrations (0.001 – 100 µM, in 10-fold dilutions in saline) to a final volume of 150 µL. Plasma clotting was triggered by adding 10-100 µL Thromborel®S, and clotting times were determined. hCpn10, saline and heparin were used as controls.

aPTT assay, was performed similarly to PT assay with sight modifications. A 100 µL citrated plasma, 100 µL TriniCLOT®aPTT HS, with various concentrations of CE76 variants (0.001 – 100 µM, in 10-fold dilutions in saline), were added to glass clotting tubes, with a final volume of 250 µL and incubated for 2 min at 37 °C in the heating block of Hyland-Clotek™. Plasma clotting was initiated by adding 50 µL of 2 mM $CaCl_2$. Time prolongations of both PT and aPTT (in sec) were converted and plotted as folds (T_i/T_0).

4.3.8 Thromboelastography (TEG)

Blood clotting time was measured using TEG assay. Herein, TEG were performed on pooled and citrated human blood from healthy donors, as described (Fischer et al., 2011). An aliquot blood (250-300 µL) was pipetted into warmed chambers of a dual-channel TEG® 5000 Analyser. To commence coagulation, it was recalcified with 20 µL of 0.2 M $CaCl_2$ and 20 µL saline, for 6-10 minutes. TEG parameters (detailed in Figure 4.3) were all measured. The speeds of fibrin-clot formation were determined as a function of protein inhibitor concentration as previously described (Fischer et al., 2011). In brief, 20 µL inhibitor (0-100 µM, 10-fold dilutions) of $CE76_{P1}$ versus E-76 peptide (control), and 10 µL of 0.2 M $CaCl_2$ with 2 µL clotting trigger (1:10 or 1:100 dilution), was added to TEG chamber, followed by 280 µL citrated blood in a total reaction volume of 340 µL, made up of normal saline, and measured at 37 °C in duplicate. TEG assays were divided into two sections in which either Thromborel®S or TriniCLOT® aPTT HS was added for the initiation of extrinsic or intrinsic blood coagulation cascades, respectively. The prolongations time (R) of fibrin-clot formations (min) were plotted in a chart versus protein/peptide inhibitor concentrations (µM).

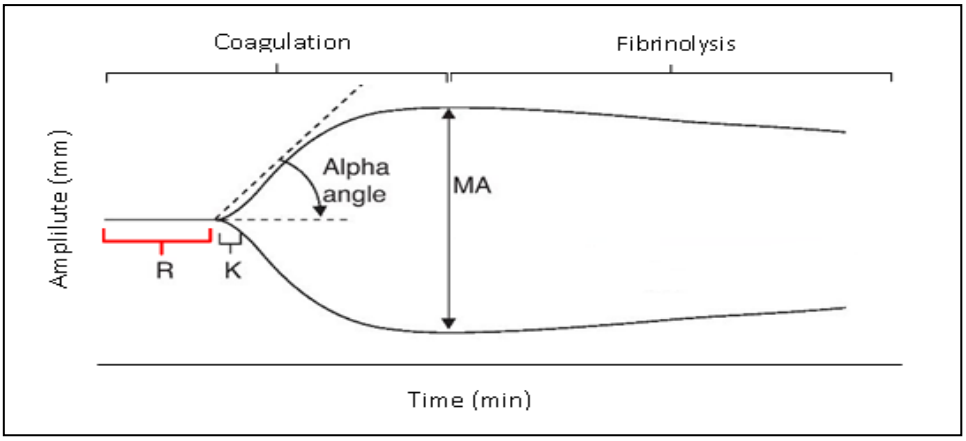


Figure 4.3 Thromboelastography (TEG) normal scheme

The important TEG parameters for the coagulation process are as follows: (**R**), the time of formation of the fibrin strand polymers (i.e. fibrin-colt); (K), the speed at which the clot forms; (α, alpha), the slope drawn from R to K; and (MA) is the strength of the clot. Diagram adapted from Galvdez and Cortes (2012).

4.4 Results

4.4.1 Protein expression and purification

The CE76 and its variants (Table 4.2) were designed *in silico* by inserting E-76 peptide between *A21* and *G38* of hCpn10, and optimized for prokaryotic expression. All variants were expressed in *E. coli* BL21 (DE3) as soluble protein with up to 30 % of cellular protein content.

Table 4.2 indicates inserted linkers (red, Italic), loop amino acids size, isoelectric point (pI), and monomer Mwt in kDa. Whereby, native hCpn10 mobile loop (form *Ala21* to *Gly38*, blue) was substituted with E-76 peptide (back).

Table 4.2 Mobile loop amino acid sequence of hCpn10 and its variants

Construct	Mobile Loop AA Sequence (*N*- to *C*-)	ML Size	pI	Mwt kDa$^{\psi}$
hCpn10*	**AAETVTKGGIMLPEKSQG**	**18**	**8.91**	10.8
CE76	**ALCDDPRVDRWYCQFVEG**	18	7.21	12.2
CE76$_{P1}$	***AAP*ALCDDPRVDRWYCQFVEG*PAA***	24	7.21	12.68
CE76$_{P2}$	***PP*ALCDDPRVDRWYCQFVEG*PP***	22	7.21	12.59
CE76$_{E5}$	***AETVT*ALCDDPRVDRWYCQFVEG*PEKSQ***	28	6.65	13.27

* Wild-type.
ψ Calculated Mwt in kDa per monomer.

Proteins were expressed with a *C*-terminus histidine-tag and were subjected to two steps of purification. It was firstly captured by affinity chromatography of IMAC (Figure 4.4), followed by final polishing step using gel-filtration chromatography (Figure 4.5). Proteins were adjusted to a working concentration of 1 mg/mL, determined by spectrophotometric and BCA assays.

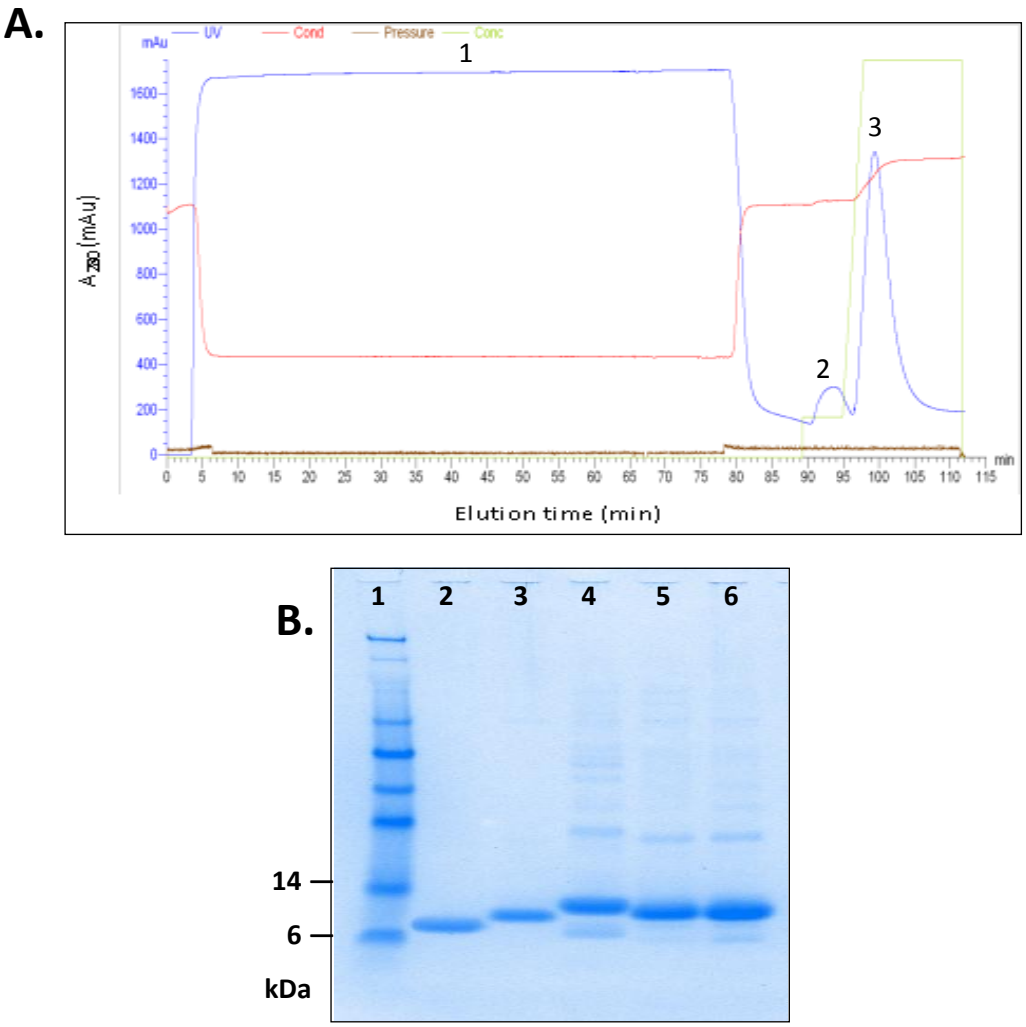


Figure 4.4 IMAC purification and SDS-PAGE analyses of CE76 variants

Representative IMAC chromatogram and SDS-PGE analysis. **(A)** IMAC purification of $CE76_{P1}$ chromatogram showing the A_{280} absorbance of (1) cell lysate loading, (2) followed by a wash step using 10 % of 0.5 M imidazole, and (3) gradient elution to 100 % of 0.5 M imidazole (pH 7.6) 20 mM Tris-HCl, 150 mM NaCl. **(B)** SDS-PAGE of IMAC purification pooled factions, Lane1: SeeBule® Plus2 Mwt std marker, Lane2: hCpn10 (10 kDa, control), Lane3: CE76 (12.2 kDa), Lane4: $CE76_{E5}$ (13.27 kDa), Lane5: $CE76_{P2}$ (12.59 kDa), Lane6: $CE76_{P1}$ (12.68 kDa).

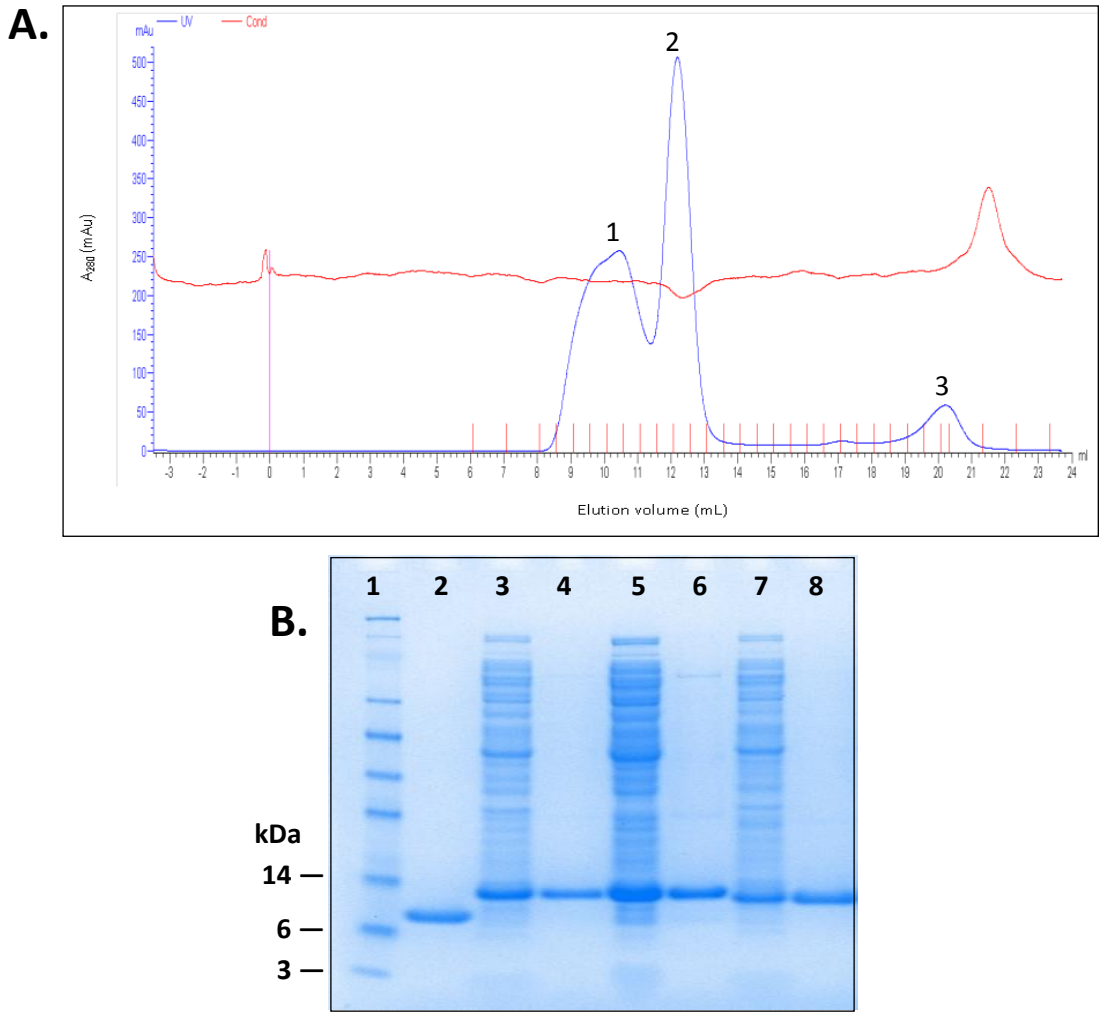


Figure 4.5 Gel filtration purification and analyses of CE76 variants

Representative GF chromatogram, form a polishing step utilised for all CE76 variants; samples were re-concentrated and loaded onto a HiLoad 16/600 Superdex® 200 PG prepacked column. **(A)** GF A_{280} chromatogram of $CE76_{P1}$ purification showing; (1) soluble-aggregates, (2) $CE76_{P1}$ heptamer, and (3) buffer residues and contaminants. **(B)** SDS-PAGE analysis of all CE76 variants; Lane1: SeeBule® Plus2 Mwt std marker, Line2: hCpn10 (10 kDa, control), Lanes 3 to 8 correspondingly show cell lysate (post-expression) and GF fractions (post-purification) of: Lanes 3-4 $CE76_{P1}$ (12.6 kDa); Lanes 5-6 $CE76_{E5}$ (13.2 kDa); Lanes7-8 $CE76_{P2}$ (12.5 kDa).

Protein purity and the approximate Mwt were determined by SDS-PAGE (Figures 4.4 to 4.6). In addition, the native (i.e. non-denatured) Mwt for each examined variants was verified by SE-HPLC (Figure 4.6). Calculated Mwt of all hCpn10 variants were also compared to observed Mwt that was determined by MALDI-TOF mass spectrometry, summarised in Table 4.3. Based on the experimental data that proved $CE76_{P1}$ was a highly stable molecule, therefore it was mostly selected for further characterisation utilising various *in vitro* assays/bioassays and other functional studies.

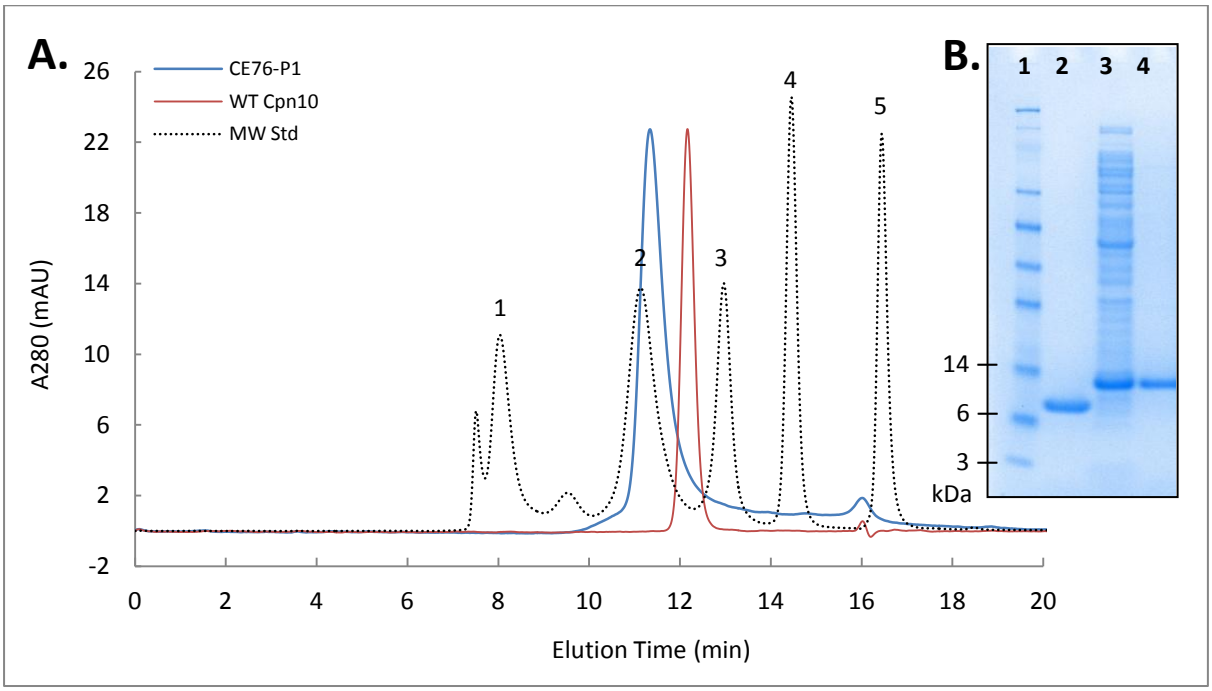


Figure 4.6 SE-HPLC and SDS-PAGE analyses of recombinant CE76$_{P1}$

(A) SE-HPLC A280 absorbance chromatogram of CE76$_{P1}$ (88 kDa, blue) apparently forming heptamer in comparison to hCpn10 (control, 70 kDa, red). Mwt standards (dashed line) are: (1) Thyroglobulin (670 kDa); (2) Bovine gamma globulin (158 kDa); (3) Ovalbumin (44 kDa); (4) Myoglobin (17 kDa); (5) Vitamin B12 (1.35 kDa). **(B)** SDS-PAGE of CE76$_{P1}$ fractions, Lane1: Mwt std marker, Lane2: hCpn10 (10 kDa, control), Lane3-4: CE76$_{P1}$ (12.6 kDa) cell lysate and SE-HPLC, respectively.

Table 4.3 below shows calculated versus observed Mwt of recombinant protein samples obtained from SE-HPLC (from Figure 4.6). Iin addition, peptide mass fingerprinting (PMF) from MALDI-MS revealed all CE76 variants were ≥ 82 % homologous to human Cpn10 protein by matching observed peptide masses (subjected to trypsin digestions) to the theoretical masses from MASCOT protein sequence database.

Table 4.3 Molecular weight and molecule identification of CE76 variants

Construct	Mass (Da)		PMF molecule identity
	m/z_{cal}	m/z_{obs}	
CE76	12206	12204	82 % hCpn10
CE76$_{P1}$	12685	12683	82 % hCpn10
CE76$_{P2}$	12595	12593	82 % hCpn10
CE76$_{E5}$	13278	13275	90 % hCpn10

4.4.2 FVIIa-CE76 binding activity

The binding properties of the CE76 variants to FVIIa were first investigated using ELISA. CE76 variants were bound to immobilised rFVIIa, and were detected using an Anti-His-tag HRP conjugated antibody (Figures 4.7 and 4.8). The binding curves in Figure 4.8 are indicative that the CE76 variants bind strongly to rFVIIa, further suggesting that the E-76 peptide was successfully incorporated into the mobile loop of hCpn10. In addition, the specificity of the CE76s were further verified by the lack of interaction observed with hCpn10. Also, rFVIIa was probed with anti-human FVII antibody (control) that resulted in a strong positive signal, comparable to the CE76 variants. It should be noted that although previous stability studies indicated that $CE76_{E5}$ was compromised, binding with rFVIIa was nevertheless strong (Figure 4.8).

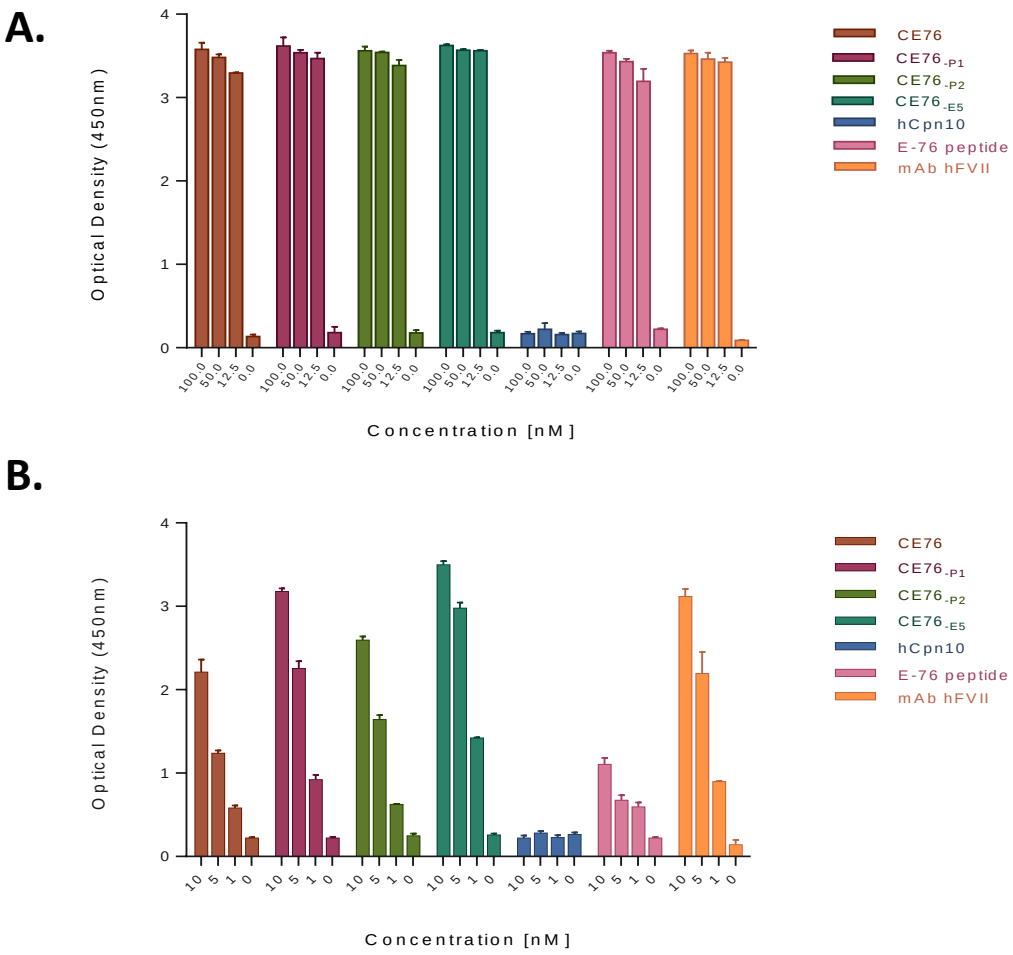


Figure 4.7 Quantitative measurement of CE76 variants binding to rFVIIa

Binding of CE76 variants to immobilised rFVIIa by ELISA. **(A)** High protein concentration range (100, 50 and 12.5 nM) for designed samples CE76, $CE76_{P1}$, $CE76_{P2}$, $CE76_{E5}$, along with controls; hCpn10 (Cpn10 scaffold less E76 peptide, negative control), E-76 peptide (naked E-76 peptide,

positive control) and anti-FVIIa mAb (positive control), respectively. **(B)** Low protein concentration range (10, 5 and 1 nM) for same samples. For CE76 specificity controls, no binding was observed to BSA, rFX or skim milk proteins (data not shown). Data represented are the mean and standard error of mean (n=3) measured at OD_{450}.

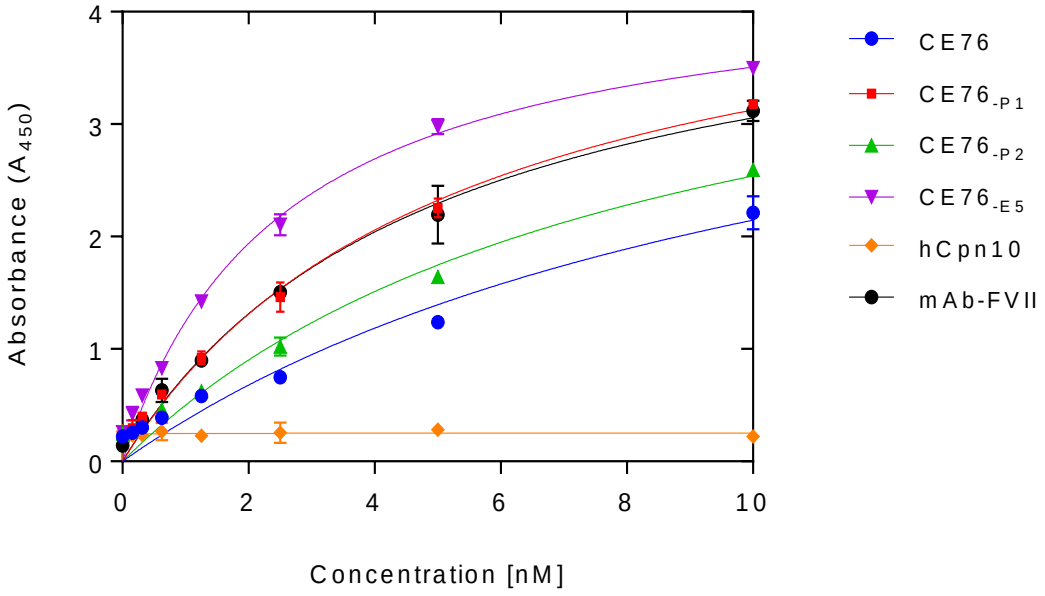


Figure 4.8 CE76 variants binding to rFVIIa

The binding of CE76 variants to rFVIIa were determined by ELISA. Protein binding to immobilised rFVIIa measured by microplate spectrophotometer at 450 nm, as following: CE76 (●, blue); $CE76_{P1}$ (■); $CE76_{P2}$ (▲); and $CE76_{E5}$ (▼) applied in series dilutions. For binding evaluation a mAb anti-human FVII (●, black) and hCpn10 (♦), positive and negative controls, respectively, were used. Data represented graphically as absorbance (A_{450}) plotted versus multiple concentrations (nM) of the applied proteins as mean ± standard error of mean (n=3), and curves were fitted by nonlinear regression.

Further binding studies were conducted on CE76 variants and rFVIIa in order to determine the kinetics of CE76s to rFVIIa and obtain binding profiles through real-time BLI using an Octet® Red system. The CE76 variants were captured onto the Ni-NTA biosensor tips and exposed to varying concentrations of rFVIIa, whereby up to 8 different concentrations can be analysed simultaneously (Figure 4.9). The observed association (k_a) and dissociation (k_d) rates of CE76 variants are summarised in Table 4.4. In essence, the calculated binding affinity (K_D) of CE76 varied between 1 to 2-fold higher relative to the control E-76 peptide (Table 4.4). This further suggests that the CE76 variants have conferred binding avidity, with K_D increased by two-fold (Table 4.4).

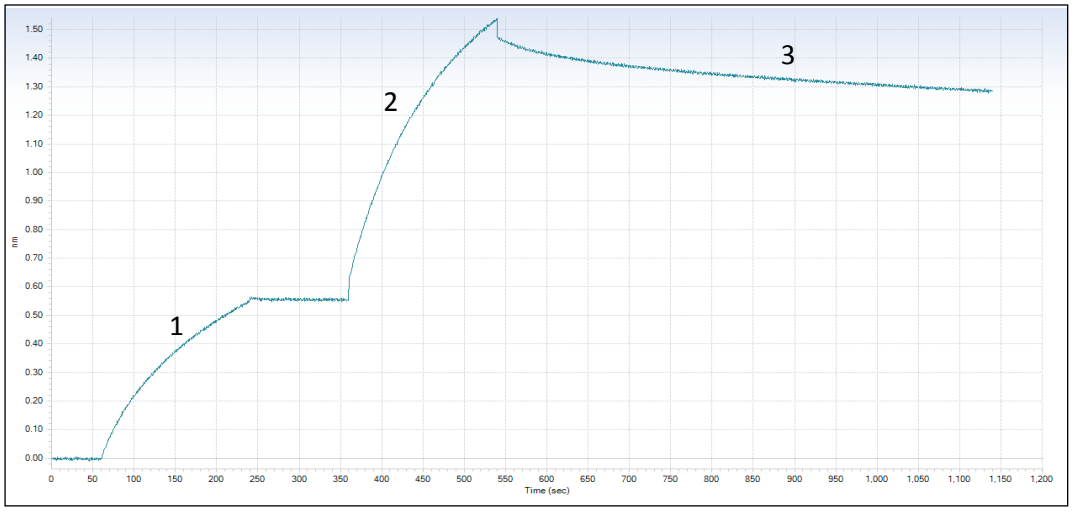


Figure 4.9 Association/dissociation (BLI) trace of $CE76_{P1}$ binding to rFVIIa

Representation of a Ni-NTA biosensor (BLI, ForteBio) as following: 1. Loaded of $CE76_{P1}$; 2. Association, and 3. Dissociation of rFVIIa, respectively. To assess non-specific binding, a Ni-NTA sensor without $CE76_{P1}$ was tested, whereby rFVIIa did not bind; also there was no binding detected of rFVIIa and hCpn10, nor rFX and $CE76_{P1}$. The calculated K_D summarised in Table 4.4. Data analysed by ForteBio® Data Analysis 7.0 software.

Table 4.4 summarises measurement of the association and dissociation rates of CE76 variants and E-76 peptide (control) binding to rFVIIa. The rates obtained from triplicate experiments.

Table 4.4 Summary of kinetic binding data determined for CE76 variants

Construct	Ligand	K_D (nM)	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})
hCpn10	rFVIIa	-*	-	-
E-76 peptide	rFVIIa	6.17 (±0.24)[a]	1.26 $x10^{+4}$ (±0.88)[a]	7.79 $x10^{-5}$ (±0.21)[a]
$CE76_{P1}$	rFVIIa	2.95 (±1.57)[a]	1.91 $x10^{+4}$ (±0.42)[a]	5.63 $x10^{-5}$ (±0.66)[a]
$CE76_{P1}$	rFX	-*	-	-
$CE76_{P2}$	rFVIIa	4.72 (±0.83)[a]	2.83 $x10^{+4}$ (±0.29)[a]	1.33 $x10^{-4}$ (±0.24)[a]
$CE76_{E5}$	rFVIIa	5.50 (±1.59)[a]	1.87 $x10^{+4}$ (±0.52)[a]	1.03 $x10^{-4}$ (±0.83)[a]

* No/insignificant binding activity detected.

[a] Errors refer to mean and ± standard error of mean from triplicate runs.

4.4.3 Inhibition of FVIIa/TF-FVIIa activity

Further assays were conducted in order to determine the specific activity of $CE76_{P1.}$ As discussed previously, the $CE76_{P1}$ has strong affinity towards rFVIIa and this subsequently led to the potent inhibition of FX, which was assessed in a dose dependent manner. Furthermore, $CE76_{P1}$ has a calculated median inhibitory concentration of 1.5 ±0.64 nM of FX (Figure 4.10A) that was equivalent in value to the E-76 peptide (control), as previously reported by Dennis et al. (2003). However, maximum inhibitory concentrations varied where 95 % of FX activity was inhibited by $CE76_{P1}$ relative to E-76 which inhibited 90 % of FX activity, despite small differences in affinity between the two. Furthermore, $CE76_{P1}$ was also found to be a potent inhibitor of amidolytic activity, which entails the hydrolysis of the small protein substrate Chromozym *t*-PA, (i.e. to determine FVII activity) with IC_{50} 6.8 ±0.55 nM (Figure 4.10B). Interestingly, the $CE76_{P1}$ inhibited the amidolytic hydrolysis with IC_{50} 8.4 ±0.6 nM despite an absence of tissue factors (Figure 4.10B). Overall, values for $CE76_{P1}$ inhibition of FX and amidolytic activities, in the absence of TF, was 1.1 nM and 9.7 nM, respectively, were similar to IC_{50} and maximal inhibition of the original E-76 peptide. This essentially and illustrates that the inhibitory activity of $CE76_{P1}$ works independently from the TF mediated triggering of the extrinsic clotting pathway.

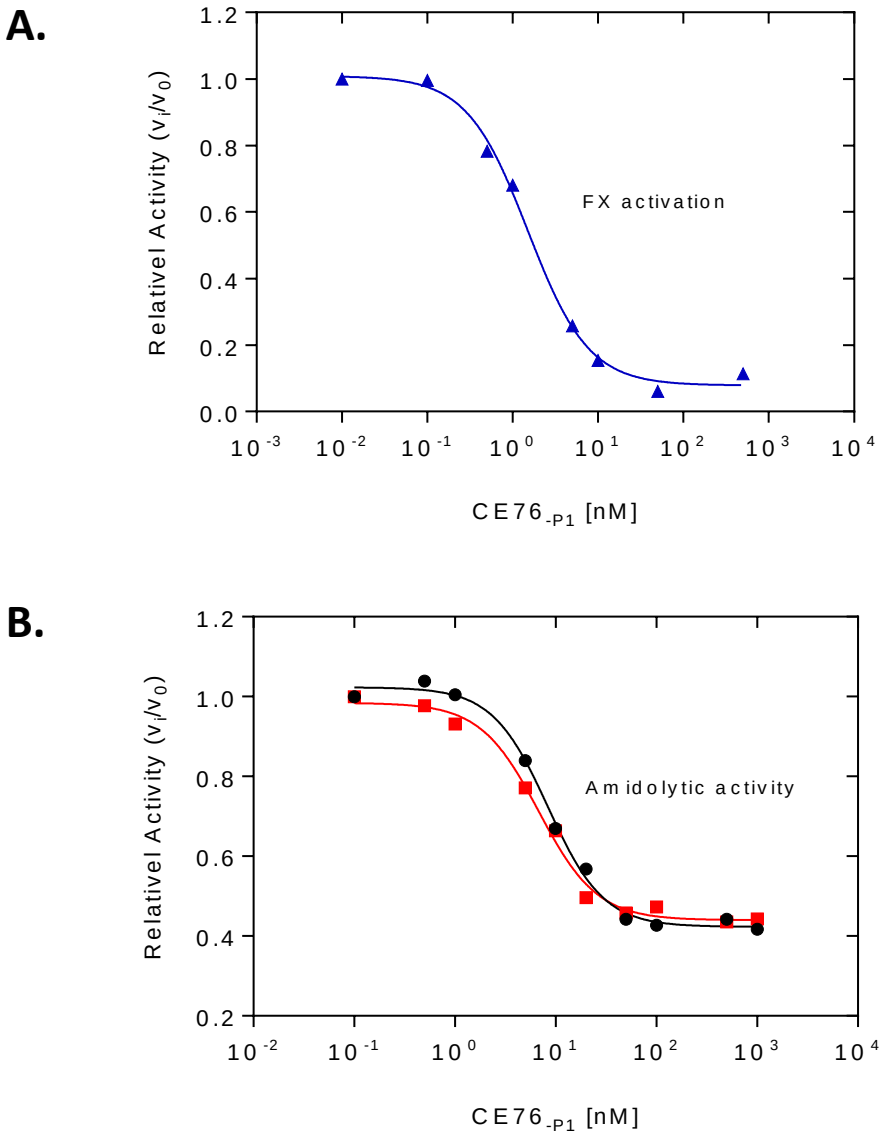


Figure 4.10 Inhibition of TF-FVIIa catalysed activation of FX and amidolytic activity

Representations of FX and amidolytic inhibition activities by $CE76_{P1}$. **(A)** The IC_{50} value for inhibition of FX activation (blue) by $CE76_{P1}$ was 1.5 nM. **(B)** The IC_{50} values for inhibition of amidolytic hydrolysis by $CE76_{P1}$ in the presence (red) or absence (black) of TF were 6.8 and 8.4 nM, respectively. Inhibition rates of both FXa and amidolytic activity expressed as fractional rate (V_i/V_0), and determined by nonlinear regression analysis of the data fit to a four-parameter equation for calculating the IC_{50} and maximal percent inhibitions.

4.4.4 Effects of CE76 on clotting times in plasma

The efficacy of the CE76 variants, as summarised in Table 4.2, were assessed in further detail using human plasma, donated by healthy volunteers (n= 5). The inhibitory effects of all CE76 variants were examined using two common clotting approaches, namely, a prothrombin time (PT) and activated partial thromboplastin time (aPTT), summarized in Figure 4.2. The PT assay is normally used to determine the deficiencies of Factors VII and X in the TF-dependent extrinsic pathway while the aPTT assay is used to determine the deficiencies of Factors IX, X, XI, and XII in the TF-independent intrinsic pathway.

Based on the ability of the CE76 variants to specifically binding to FVII, FVIIa, and/or TF-FVIIa-complex, most CE76 variants were significantly prolonged clot formation in the PT assays by 4-fold compared to control E-76 peptide (Figure 4.11A). On the other hand, the CE76 variants (mainly $CE76_{P1}$) did not prolong clot formation in any of the aPTT assays as expected (Figure 4.11B). This further indicates that $CE76_{P1}$ had specific interactions with the FVII/FVIIa of the extrinsic pathway only. It is significant to note that at higher concentrations, the $CE76_{P1}$ considerably prolonged the clot formation times, in contrast to the E-76 peptide that reached saturation and plateaued at 3-fold of the control hCpn10 or saline (i.e. non-inhibitors). It is more than likely that the extended clot formation time at higher concentrations is correlated with the higher avidity of the $CE76_{P1}$ as there are seven binding sites (i.e. E-76 peptide) displayed within $CE76_{P1}$ (Figures 4.1 and 4.11A).

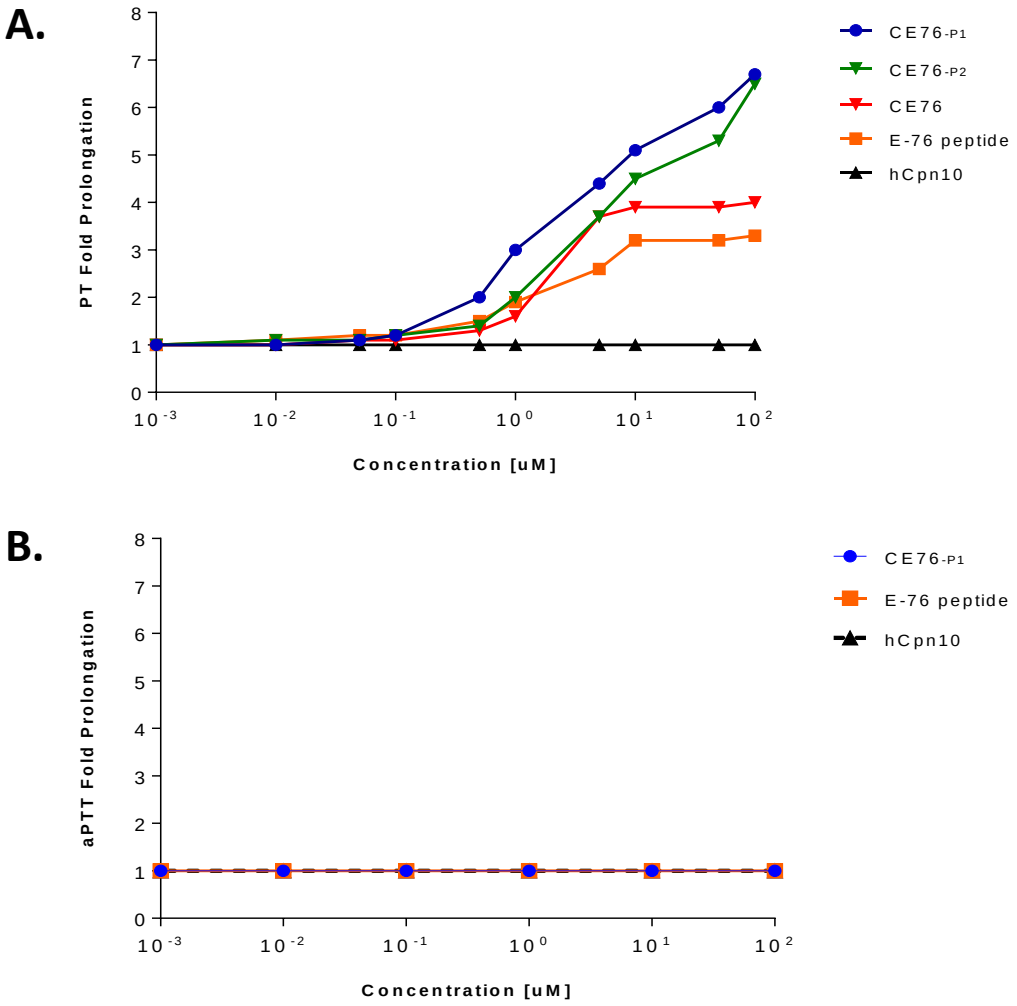


Figure 4.11 Prolongation of PT and aPTT in human plasma

PT and aPTT assays performed in a pool of healthy human plasma (n=5) using Hyland™ Clotting Analyser. **(A)** Prolongation of plasma clotting times upon initiation by Thromborel® S (Human thromboplastin) and Ca^{2+} in PT assays, measuring the effect of CE76 variants versus controls E-76 peptide and hCpn10. **(B)** Prolongation of plasma clotting times upon initiation by TriniCLOT® aPTT HS (Phospholipids Mixtures) in aPTT assays, measuring the effect of $CE76_{P1}$ versus E-76 peptide and hCpn10. The uninhibited clotting times were 15 and 40 sec for the PT and aPTT, respectively, prolongations time (T_1/T_0) plotted as (fold) versus CE76 variants (µM). hCpn10 and saline were used in both assays as a negative control.

4.4.5 Effects of $CE76_{P1}$ on human blood

The anticoagulant effects of $CE76_{P1}$ were further investigated via TEG in human blood. Citrated blood samples were acquired from healthy donors (n=3). TEG is normally used to monitor significant coagulopathy in patients by using whole blood. The TEG assesses the interaction of platelets within the coagulation cascade factors, starting from the time the platelet-fibrin interaction is initiated, right through to platelet aggregation, clot strengthening, fibrin cross linkage, and eventually clot lysis, as summarized in Figure 4.4.

The focus of this research study is to determine the reaction time (R-Time) in the final development of a fibrin clot through both coagulation cascades when blood is exposed to $CE76_{P1}$. In order to speed up the clotting process, the citrate-treated blood was initiated by either thromboplastin/Ca^{2+} for TF-dependent extrinsic factors or initiated by phospholipids for TF-independent intrinsic factors (Figure 4.2). Interestingly, a dose of 1 µM of $CE76_{P1}$ was sufficient to significantly delay the development of a fibrin-clot up to three-fold compared with 1 µM of the original E-76 peptide, which delayed blood clotting by approximately 5 mins (Figure 4.12). Conversely, the $CE76_{P1}$ had no direct impediment effects on the coagulation processes of the TF-independent intrinsic factors when initiated by partial thromboplastin and phospholipids (Figure 4.13). Therefore, the anticoagulant effect of $CE76_{P1}$ is only at prolonging clot formation which triggered by the extrinsic pathway factors in a dose dependent manner. Importantly to declare, that data obtained for this experiments were mostly from duplicate runs due to number of issues; first, time scale in which TEG cycle required about 25 to 45 min per run for two samples therefore only measuring the effect of $CE76_{P1}$ versus three controls (i.e. E-76 peptide, hCpn10, and Heparin). Second, limited number of blood donor, in which TEG results were found to be more accurate from newly collected blood samples.

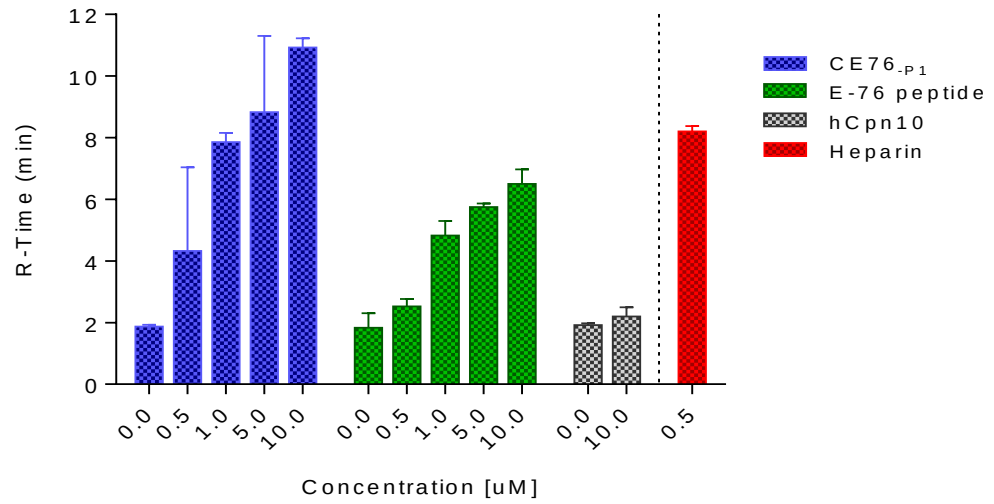


Figure 4.12 A measurement of the reaction time (R-time) for $CE76_{P1}$ versus E-76 peptide

Plot of the clotting times (R, min) versus multiple concentrations (µM) of $CE76_{P1}$ and E-76 peptide. The R-time for uninhibited blood clotting was 2 min by hCpn10 (10 uM), while inhibited blood clotting by heparin (0.5 U/mL) was approximately 8 min. Citrated blood samples were from healthy donors (n=3), and for escalating clotting process a Thromborel® S (Human thromboplastin) was used. The blood clotting times were measured by using TEG® 5000 Analyser (Thromboelastography®, HAEMONETICS®).

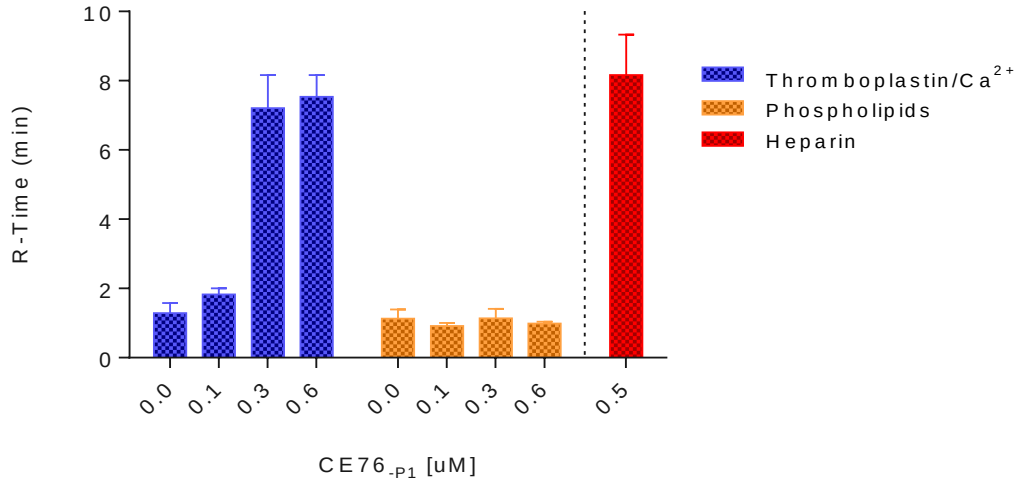


Figure 4.13 Measurement of anticoagulant activity of $CE76_{P1}$ by TEG

Graphical representation illustrated the effect of $CE76_{P1}$ on R-time for the development of fibrin-clot of citrated human blood from healthy donors (n=3). The anticoagulant activity of $CE76_{P1}$ determined upon initiation of extrinsic/TF-dependent pathway by Thromborel®S/Ca^{2+} versus initiation of intrinsic/TF-independent pathway by TriniCLOT® aPTT HS. Shows $CE76_{P1}$ (blue) was adequate to pause blood clotting upon initiations by TF/Ca^{2+}. No significant anticoagulant effects (orange) were detected upon initiations by phospholipids. Heparin (0.5 U/mL, control) shows a clear anticoagulant activity of 8 min upon initiation of either pathways. R-time (min) measured using TEG®5000 Analyser.

4.5 Discussion

Blood coagulation is a process wherein a series of proteolytic reactions leads to the activation of thrombin, which plays a key role in the formation of a fibrin clot (Gray et al., 2012, Lefkowitz, 2013). The coagulation cascade is composed of two separate pathways, namely, the intrinsic and extrinsic, which are initiated by different factors (detailed in Figure 4.2). The extrinsic pathway commences when the circulating zymogen FVII, an inactive serine protease, is exposed to tissue factor (TF, a nonvascular protein) which is released upon vessel injury or trauma (Lefkowitz, 2013). FVII is activated upon binding to TF forming FVIIa, which results in the formation of the TF-FVIIa complex. This complex subsequently activates three important serine proteases within the cascade, and the conversion of FX, FIX and FVII into their active form FXa, FIXa and FVIIa, respectively. This in turn, leads to the generation of thrombin and eventually a fibrin clot (Gray et al., 2012, Lefkowitz, 2013). It is interesting to note that only 10 nM of activated FVII is required to initiate the blood coagulation process (Maun et al., 2003), and therefore, zymogen FVII or activated FVIIa serves as an attractive therapeutic target when the extrinsic pathway is initiated with the TF-FVII complex.

Classical anticoagulant medications such as warfarin and heparins are widely used in regulating the clotting pathway (Klement and Rak, 2006). Warfarin works by impairing the functions of vitamin-K dependent proteins such as FVIIa, FIXa, FXa and thrombin as well as the activated anticoagulant proteins such as protein C and protein S (Klement and Rak, 2006). This is done by inhibiting the post-translational modification of these proteins, which are required for the interaction with the phospholipid membrane, thus leading to the disruption of the coagulation cascade. Heparin modulates the coagulation cascade by directly interacting with the plasma protein antithrombin III (ATIII) (Gray et al., 2012). This subsequently leads to the enhanced inactivation of a number of serine proteases, namely, FIIa, Xa and IXa. Heparins accelerate the ATIII that inhibit FXa and thrombin and plays an important role by activating the protein C pathway, in the presence of thrombomodulin (Enjyoji K et al., 1995). Based on the broad interaction of these drugs against a number of coagulation proteins, drug dosages need to be strictly monitored as it can disturb the balance between thrombosis and haemostasis.

Based on the non-selective interaction of warfarin and heparin, a number of approaches have been undertaken in order to develop inhibitors that are specific towards FVII or TF-FVIIa

complex, which in turn will prevent activation of FX. An example is the design of a TF-inhibitor that can block the interactions of TF and FVII, calcium, or phospholipids; or by interfering with the FX activity through TF-mutants, monoclonal antibodies, or synthetic peptide inhibitors (Paborsky et al., 1995, Kelley et al., 1997). Likewise, a FVIIa-inhibitor was developed from a Fab antibody, that worked by inhibiting the catalytic activity of FVIIa through interaction with the FVIIa-active site (Biemond et al., 1995). In addition, a naturally-occurring inhibitor of the TF-FVIIa complex such as tissue factor pathway inhibitor (TFPI), a protease inhibitor with three Kuntiz-type inhibitory domains, was developed through recombinant DNA technology (Enjyoji K et al., 1995). At last, a recombinant NAPc2 was also developed that inhibits the activity of the TF-FVIIa complex by forming a canonical inhibitory loop within the FVIIa-active site when bound to the exosite of zymogen FX or FXa (Bergum et al., 2001).

Recently, peptide members of a new class of serine protease inhibitor with a specificity towards FVIIa were successfully isolated from a random peptide library via phage display and were classified as an E- (Dennis et al., 2000) and A-series (Dennis et al., 2001). Studies conducted by Dennis et al. (2000), have reported the characterisation of the anticoagulant peptide, classified as E-76 (an 18 residue peptide) that binds to FVIIa, both displayed on the surface of filamentous phage and as a synthetic peptide, and determined that the peptide effectively inhibited the TF-dependent coagulation cascade, and ultimately the activity of FX. The interaction of TF with FVIIa is important as it facilitates the binding to FX and anchorage of FVIIa to cell membrane. The inhibitory properties of the E-76 peptide clearly resulted from its non-competitive inhibition of FX activation (Dennis et al., 2000). The resolved crystal structure of FVIIa-E-76 complex through X-ray crystallography revealed that the E-76 peptide binds to the exosite of FVIIa rather than the active site within the protease domain. The binding of the E-76 peptide to FVIIa caused a significant conformational change within the 140's loop, which constitutes a part of the activation domain. The studies further indicate that the E-76 peptide may alter the way in which FVIIa binds to FX as well as disrupt the hydrogen bond that is linked to the oxyanion hole in the active site. Therefore, the E-76 peptide inhibits TF-FVIIa activity by an "allosteric switch" which in turn may give rise to both modes of inhibitions (steric and allosteric) for FX activation (Dennis et al., 2000).

Although peptides show significant promise as potential therapeutics through *in vitro* analyses, their utility in *in vivo* (physiological) settings are severely limited, as they are

subject to either rapid protease degradation or clearance by the renal system (Diao and Meibohm, 2013, Sato et al., 2006). Therefore, in order to confer greater stability and to decrease the metabolic clearance rate so as to enhance the intended therapeutic modality, such peptides can be attached to various molecular scaffolds (Chapter 1), which can ameliorate their rapid clearance or degradation *in vivo*, while increasing the overall binding strength (or K_D) through the avidity effect of multiple peptides within the molecular scaffold.

Eukaryotic Cpn10 is an essential mitochondrial protein that assists in protein folding (Hartl and Hayer-Hartl, 2002). It is an oligomeric protein that is structurally conserved across many species (Perham and Wittung-Stafshede, 2007). The quaternary structure of Cpn10 is a heptameric fold (70 kDa) that is composed of seven identical monomers that are stabilised through a number of hydrogen bonds. Each Cpn10 monomer (10 kDa) consists of a rigid hydrophobic β-barrel core, a β-hairpin roof loop, and a mobile loop that extends below the core of Cpn10 (Figure 4.1). This flexible "mobile" loop region is a hallmark structure that plays a crucial role in interacting with its cognate Cpn60/GroEL (Lin et al., 2004).

The mobile loop within Cpn10 lacks any well-defined secondary structure and is subject to various conformational changes (Chapter 3). The hCpn10 mobile loop consists of 18 residues, namely, *AAETVTKGGIMLPEKSQG*, and substitution or a deletion (e.g. ΔL-Cpn10) of this region does not compromise the overall structure of the heptameric fold (Chapter 3). Basically, the loop domain forms an extended interface of the designed molecular scaffold, which facilitates binding to the target molecule. Based on the unique features of hCpn10, there are opportunities to design a newly and highly effective FVIIa inhibitor by substituting the mobile loop region with the E-76 peptide and thus create additional binding sites on par with hCpn10.

The research reported in this study shows that the hCpn10 mobile loop domains are able to be substituted with the anticoagulant E-76 peptide, *AL**C**DDPRVDRWY**C**QFVEG*, in order to generate a NME based on the hCpn10, that binds FVIIa with nano-molar affinity, with structural stability and longer half-life than that of the E-76 peptide alone. Furthermore, this research study is focussed on the optimal designing of a molecular scaffold based on the hCpn10, and incorporating the E-76 peptide (so-called CE76).

The mechanism by which the CE76 binds to FVIIa is beyond the scope of this study. However, extensive analyses have been reported elsewhere (Dennis et al., 2000, Dennis et al., 2001). Furthermore, structural optimisation of CE76 was conducted through molecular modelling, whereby three CE76 variants were created with the insertion of a number of peptide linkers and designated accordingly as $CE76_{P1}$, $CE76_{P2}$ and $CE76_{E5}$ (Chapter 3). The research necessitated the expression and purification of the various CE76 variants and further testing for their anticoagulant properties. However, due to a higher yield and perceived stability of the $CE76_{P1}$, this variant was selected for most functional characterisation.

Obtaining a high yield of CE76 proved to be challenging, however this was overcome by inducing high expression from the pET30a vector using IPTG inducer. CE76 variants included a *C*-terminus histidine(x6)-tag in order to facilitate purification. The CE76 variants were expressed in *E. coli* BL21 (DE3) as soluble proteins that constituted roughly 25-40 % of total cell proteins, as determined by the BCA assay.

Maintaining the CE76 in a soluble form was also a challenge due to its low pI (≤ 7.21), which is close to the pI of most bacterial proteins. Furthermore, retaining the protein in a stable form was difficult due to the presence of a disulphide bond within the E-76 peptide. In order to confer more stability within the protein, multiple buffer screening trials were performed, whereby aggregation (leading to precipitation) for each protein was examined at 320 nm and 390 nm (Chapter 2). CE76 variants were found to be more soluble, stable (i.e. no aggregation), and functional (i.e. conservation of binding to FVIIa) in 100 mM sodium acetate buffer pH 5.5, where pH was two units below the pI.

All CE76 variants were purified from bacterial lysates through two purification stages, whereby the first purification step was capturing by IMAC followed by a second purification step of polishing by gel-filtration. The polishing step ensured that a high purity of heptameric CE76s could be acquired (see Figure 4.5). CE76 purity and Mwt were verified by MALDI-TOF MS and SE-HPLC analyses (Table 4.3 and Figure 4.6, respectively). Importantly, the three CE76 variants, namely $CE76_{P1}$, $CE76_{P2}$ and $CE76_{E5}$, were analysed by analytical SE-HPLC in order not only to verify conservation of native heptameric configurations, but also to monitor the stability with respect to aggregation, for example the variant molecules stored over varying time periods at different temperatures, namely -20 °C, 4 °C, 25 °C and 37 °C. As predicted, two CE76 variants (i.e. both $CE76_{P1}$ and $CE76_{P2}$) exhibited stable and native-

like heptameric conformations under such storage conditions as illustrated by analytical SE-PHPLC measured at 280 nm.

Since the E-76 peptide was isolated from a random peptide library against clotting Factor VIIa (Dennis et al., 2000), the newly synthesised CE76 was at first tested against human rFVIIa. Initially, the CE76 was tested for binding reactivity to rFVIIa through conventional ELISA. The results showed that all CE76 variants bound to rFVIIa with concentrations as low as 1 nM (Figures 4.7 and 4.8). In contrast, hCpn10 control (up to 1 µM) showed no significant binding to rFVIIa. This was to ensure that the interactions between the scaffold-based hCpn10 and rFVIIa had not occurred independently from the E-76 peptide.

The equilibrium dissociation constant (K_D) was also determined for the CE76 variants in real-time by BLI-Octet®. Octet is as versatile complement of commonly used Biacore® but it operates in a more multiplexed format in order to detect and quantify the kinetics of molecular interactions (Abdiche et al., 2008). Each *His*-tagged CE76 variant was captured by the optical biosensors that were coated with Ni-NTA. Subsequently, the binding affinity of rFVIIa was measured in a concentration dependent manner. The $CE76_{P1}$ had significantly higher affinity binding for FVIIa than other CE76 variants, with a K_D 2.95 nM (Table 4.4). Importantly, the binding profile of $CE76_{P1}$ to rFVIIa was as similar to the original E-76 peptide (Dennis et al., 2000). Similar to the ELISA data, there was no binding detected for hCpn10 to rFVIIa, further confirming the specificity of CE76 variants to rFVIIa (detailed in Table 4.2). Importantly, in all previous detection methods, the CE76 variants required the presence of Ca^{+2} ions (in the form of 5 mM $CaCl_2$) in order to enable the binding to rFVIIa, although TF was not as crucial. Furthermore, the CE76 variants did not hinder the interaction between TF to rFVIIa due to interaction of CE76s with the exosite of FVIIa. Therefore, the CE76 variants were only capable of binding to the coagulant FVIIa in the presence of Ca^{+2}. In comparison, TF is only essential for activating zymogen FVII and for mounting active FVIIa onto the cell membrane phospholipids that results in the activation of FXa.

Further characterisations of the inhibitory properties of CE76 against FVIIa were conducted. Accordingly, $CE76_{P1}$ is the optimal candidate NME and was chosen for such characterisation. In comparison with the E-76 peptide (control), there was a slight improvement in the inhibition of TF-FVIIa mediated activation of FX in both the maximal extent of inhibition and the affinity of $CE76_{P1}$ (Figure 4.10A). The $CE76_{P1}$ inhibited roughly ~ 95 % of FX

activation, in addition the IC_{50} of $CE76_{P1}$ was 1.5 nM. Furthermore, it was important to determine if the extension of the *C*-terminus tag of 4x*Gly*/6x*His* residues was required for the inhibition of FX activation, and therefore, a similar assay was conducted with hCpn10 that contained same tag. This verification was deemed necessary since charged residues like histidine with a pKa of 6.04 may interfere with the interaction between $CE76_{P1}$ and FVIIa, which in turn may increase or decrease the level of FX activation. Consistent with these finding, the histidine extensions did not adversely affect FX activation (data not shown). Therefore the effect of the inserted E-76 domain was independent of the scaffold-hCpn10 molecule of $CE76_{P1}$. In addition, the $CE76_{P1}$ provided no substantial improvement to the calculated affinity from half-maximal inhibitory concentration, in which the IC_{50} values of 1.5 nM is slightly equivalent to the control E-76 peptide of 1.1 nM.

Moreover, the ability of $CE76_{P1}$ to inhibit the amidolytic activity, which determines FVII activity, was also investigated and measured using the small chromogenic substrate chromozym *t*-PA. This in turn may give further insight into assessing whether the binding of scaffold-designed $CE76_{P1}$ could affect the enzymatic active site, of the active serine domain, of FVII. The data for inhibition of amidolytic activity is illustrated in Figure 4.10B. The amidolytic activity assays have a larger dynamic range since the maximal inhibition by $CE76_{P1}$ at saturating concentrations was only 60 % instead of 95 % as found in the FX activation assays. Consistent with previous data for FX activation, a slight improvement of the maximal extent of inhibition by $CE76_{P1}$ to 60 % was observed. Furthermore, the maximal inhibition of amidolytic activity at saturating concentrations by $CE76_{P1}$ of TF-FVIIa and FVIIa-only were determined to be 60 and 62 % inhibitions, respectively (Figure 4.10B). This indicates that TF had no substantial effect when using the activated form of FVII, nor does the $CE76_{P1}$ obstruct the TF binding to FVII. Importantly, within the amidolytic activity, the IC_{50} value that was determined for $CE76_{P1}$ was 6.8 nM, whereas in the absence of TF, the IC_{50} value was slightly higher at a value of 8.4 nM (Figure 4.10B).

The scope of this chapter is also focussed on examining the anticoagulant properties of the designed CE76 variants through prothrombin time (PT) of TF-dependent clotting assays in citrated human plasma (Kamal et al., 2007). Remarkably, the $CE76_{P1}$ specifically prolonged the TF-dependent coagulation pathway 8-fold over normal clotting time, and thus potently inhibited FX activation. Furthermore, the $CE76_{P1}$ illustrated its capacity to prolong clot formation in the PT-coagulation assays in a dose-dependent manner (Figure 4.11). This is

presumably due to the increased avidity of the $CE76_{P1}$ as the molecule scaffolds displays seven E-76 peptides causing indistinct multiple binding. In comparison, the maximal PT fold-prolongation for the control E-76 peptide plateaued at a 3-fold over normal clotting time, which was significantly lower than the observed in the $CE76_{P1}$ as well as other CE76 variants (Figure 4.11A).

The properties of the $CE76_{P1}$ variants were assessed further through the activated partial thromboplastin time (aPTT) of TF-independent (Kamal et al., 2007). As evident from these analyses, neither the $CE76_{P1}$ nor the control E-76 peptide had significant effect in prolonging clot formation in the TF-independent coagulation pathways (Figure 4.11B). This gives further credence that $CE76_{P1}$ is highly specific at prolonging clot formation only in the TF-dependent extrinsic pathway through its interaction with the FVII/TF-FVIIa complex. In contrast and as expected, no anticoagulation effect was observed in either the PT or the aPTT when using wild-type hCpn10. It is important to observe the influence of hCpn10 molecule in the overall binding to any of the serine protease factors of both extrinsic and intrinsic pathways.

The potent activity of $CE76_{P1}$ and its specificity towards FVIIa within the extrinsic pathway is indeed significant. The higher avidity of the $CE76_{P1}$, as evident from the extended prolongation within the PT assay relative to the E-76 peptide, can be attributed to the heptameric nature of the designed scaffold structure. However, the multiple binding sites of the $CE76_{P1}$ towards its target FVIIa do not indicate an absolute binding of molar ratio of 1:7, which requires further detailed studies and is beyond the scope of this study. This observation does not occur when examining the control E-76 peptide with molar ratio of 1:1 binding that reached saturation prolongation at certain concentrations over the PT-assay (Figure 4.11A).

Due to the limited access to animal models (e.g. rabbit model of venous thrombosis) for conducting comparative *in vivo* anticoagulant studies for $CE76_{P1}$, an alternative *in vitro* thromboelastography (TEG) was used. TEG is normally used to monitor significant coagulopathy in patients by using whole blood, whereby TEG enables the assessment of function made from single blood sample, documenting the interaction of platelets with the protein coagulation cascade from the time of the initial platelet-fibrin interaction, through platelet aggregation, clot strengthening and fibrin cross linkage, to eventually clot lysis (Donahue and Otto, 2005, Martini et al., 2008, MacDonald and Luddington, 2010). Thus, the anticoagulant effect of $CE76_{P1}$ was investigated *in vitro* using blood samples from healthy

donors. The reaction time (R), known as coagulation development time, was measured from the initiation of the first fibrin clot (Figure 4.3). In order to speed up the clotting processes, the citrated blood was triggered by thromboplastin/Ca^{2+}, for initiating the clotting reactions of extrinsic pathway factors of FVII and FX**.** Yet, similar to the observation of effects on PT-assays above, the $CE76_{P1}$ significantly extended the R-time by 11-fold over the uninhibited blood clotting, whereas the maximal R-time prolongation for the control E-76 peptide was only 5 mins when clotting was initiated by thromboplastin/Ca^{2+} (Figure 4.12). Furthermore, the specificity of $CE76_{P1}$ towards the extrinsic pathway was evident as no inhibitory activity was observed through the TF-independent coagulation pathway when the clotting reactions were triggered by phospholipids/Ca^{2+} for initiating of the intrinsic pathway Factors of IX, X, XI, and XII (Figure 4.13).

4.6 Conclusion

The aim of this study was to investigate the designing of a NME based on the hCpn10 scaffold with specificity for FVIIa. The goal was to optimise the CE76 structure for stability and for binding capability to FVIIa. This study is not aiming to replace the classical well known anticoagulant drugs but rather is providing an alternative and a new class of biologic for specific medical applications; for example, administration post angioplasty to prevent platelet aggregation and thrombosis events, the latter being not an uncommon occurrence post angioplasty.

The anticoagulant CE76 was successfully designed by displaying a functional E-76 peptide (targeting coagulating FVIIa) within the mobile loop between *Ala21* and *Gla38* of hCpn10. The CE76 was further optimized through the introduction of additional peptide linkers (detailed in Chapter 3), which conferred more stability and functionality within the overall protein complex.

In this study, CE76 variants ($CE76_{P1,}$ $CE76_{P2,}$ and $CE76_{E5}$) production, purification, and *in vitro* bioassays were performed and executed. Importantly, the $CE76_{P1}$ potently inhibited the FVIIa-catalysed, and subsequently inhibited the activation of FX. Furthermore, the $CE76_{P1}$ significantly prolonged the clotting time of both human plasma and blood, indicating a potential use as a new class of biologic-based anticoagulant. The success of this approach in

developing a molecular scaffold-based on hCpn10 represents a framework for developing new molecular entities (NMEs) for multiple therapeutic and/or diagnostic purposes.

Remarkably, the success of these experiments could result in $CE76_{P1}$ being trialled as an anticoagulant biologic to balance thrombosis caused by surgery, and haemostasis caused by classical unspecific anticoagulant, primarily for postsurgical patients. This will be an extraordinary future goal.

Chapter 5. The Development of a Novel Cell Surface CD44 Detection Probe based on Displaying P7-Peptide Lligand on hCpn10-Molecular Scaffold

5.1 Introduction

In 2010, there were around 30 million reported new cases of cancer globally with about 1.4 million cases reported in the United States (US) and it is forecasted that there will be around 10.3 million cancer-related death worldwide by 2020 (Alwan, 2011, Bonvillian, 2011). A study conducted between 1970 and 2006 in the US has indicated that out of 100,000 patients, cancer mortality rates have dropped from 21 to 11 % in men, and from 12 to 6 % in women, which is an insignificant drop compared to other diseases (Jemal et al., 2010). In the United Kingdom (UK), the cancer related death rate dropped by 20 % for women and 27 % for men between 1990 and 2011 (Wise, 2013). This can be attributed to advances made in early disease detection which in turn translates to better outcomes for patients diagnosed with cancers of the breast, cervix, colon and rectum (Battaglia et al., 2011, Kaluna-Czaplinska and Jozwik, 2014, Ricci-Vitiani et al., 2007, Cappello et al., 2003, Weigel and Dowsett, 2010). Hence, early detection of cancer can improve patient recovery and survival outcomes and ease the financial burden so often associated with cancer treatments which is estimated to be $US 20,000 to 100,000 per patient annually (Hassett and Elkin, 2013, Samaranayake et al., 2009).

Cancer is a very complex disease and as such conventional diagnostic methods such as palpation, tissue biopsy analysis, or blood tests (Seoane Leston and Diz Dios, 2010), may not always be conclusive. Thus, if clinicians are suspicious of potential malignancies, advanced techniques such as magnetic resonance imaging (MRI), digital mammography, as well as conducting genetic testing through either microarrays or real time polymerase chain reaction (RT-PCR) are then utilised (Kaluna-Czaplinska and Jozwik, 2014, Ryu et al., 2013, De Abreu et al., 2013, Bohndiek and Brindle, 2010, Zoon et al., 2009, Bendall et al., 2000). Despite its high accuracy and reliability of results, many patients have limited access to these diagnostic tests.

Cancer biomarkers are alternative tools in the diagnosis and management of cancer. Biomarkers are usually derived from endogenous proteins such as cell surface receptors or metabolites that are secreted in the circulatory system (Kaluna-Czaplinska and Jozwik, 2014, Weigel and Dowsett, 2010) such as the serine protease kallikrein-related peptidases 3 (KLK3, also known as prostate-specific antigen: PSA) that are found circulating in the blood and used as a biomarker for prostate cancer (Hong, 2014). Another example is the transmembrane growth factor receptor HER2 that is normally expressed from secretory epithelia used as diagnostic biomarker for breast cancer (Hicks and Tubbs, 2005).

Cancer biomarkers, in general, can be detected by using less sophisticated visualising techniques such as immunofluorescence and confocal microscopes, or through more advanced techniques such as flow cytometry/fluorescence-activated cell sorting (FACS), array scanner or other automated advanced imaging instruments (Duan et al., 2014, Rahman et al., 2013, Wang et al., 2012, Moadel, 2011, Song and Naeim, 2004). These techniques are based on binding kinetics between, for example, antibody and antigen (Schmetzer et al., 2012, Oshikiri et al., 2006) presumably based on the premise that an overexpressed cell surface biomarker is present in cancer cells. Thus, biomarkers are typically used in the diagnosis, treatment or monitoring of various cancers such as in the detections of HER2 in breast cancer patients to guide, manage and most likely identify patients that can benefit from anti-HER2 Herceptin® antibody and assist in monitoring patients' recovery progress (Hicks and Tubbs, 2005).

CD44 is a cancer biomarker that is overexpressed in different tumours such as in breast, prostate, colon and brain and also in leukaemia (Akisik et al., 2002, Khoursheed et al., 2002, Kaaijk et al., 1995, Olsson et al., 2011, Takaishi et al., 2007, Dhir et al., 1997). CD44 is a polymorphic membrane glycoprotein (i.e. cell surface receptor) with variant isoforms and Mwt that ranges from 85-250 kDa (Olsson et al., 2011, Ponta et al., 2003, Kaaijk et al., 1995, Eibl et al., 1995). CD44 is involved in adhesion with other cells and the extracellular matrix, B cell and T cell immune responses, inflammation and tumour metastases (Olsson et al., 2011, Negi et al., 2012, To et al., 2010, Vincent and Mechti, 2004, Ponta et al., 2003).

Recently, CD44 was also detected as a cell surface marker of cancer stem cells (CSCs) and is found to be highly expressed in several types of cancers (Balic et al., 2013, Jaggupilli and Elkord, 2012, Negi et al., 2012, Ishimoto et al., 2010, Hurt et al., 2008). CSCs are similar to normal tissue stem cells but are thought to be insensitive to chemo- and radiotherapies due to cells being quiescent or slow-cycling. CSCs are found to be responsible for tumour initiation (Hurt et al., 2008). Thus, as a cell surface protein, CD44 serves as an ideal target in the development of highly effective therapeutics e.g. antagonist and/or development of diagnostics e.g. an early detector probe for CD44.

In this study, the anti-CD44 P7-peptide was displayed on hCpn10 as NMEs and developed as a diagnostic probe for CD44, termed CP7. The P7 peptide which was previously isolated using phage display screening (Park et al., 2012) was incorporated into the mobile loop of hCpn10 between the *Ala21* and *Gly38* residues. The heptameric nature of hCpn10 facilitates the potential display of seven loops exposing the P7 peptide (Figure 5.1).

Indeed there is a growing demand for a reliable diagnostic or therapeutic probe and as such, hCpn10 scaffold-based CP7 was assessed in a number of *in vitro* bioassays to determine the binding efficacy of this protein against recombinant CD44 (rCD44) and its binding to breast adenocarcinoma, MDA-MB-486 cells, that overexpress CD44. CP7 could potentially be utilised as a diagnostic probe against a number of cancers with known CD44 overexpression.

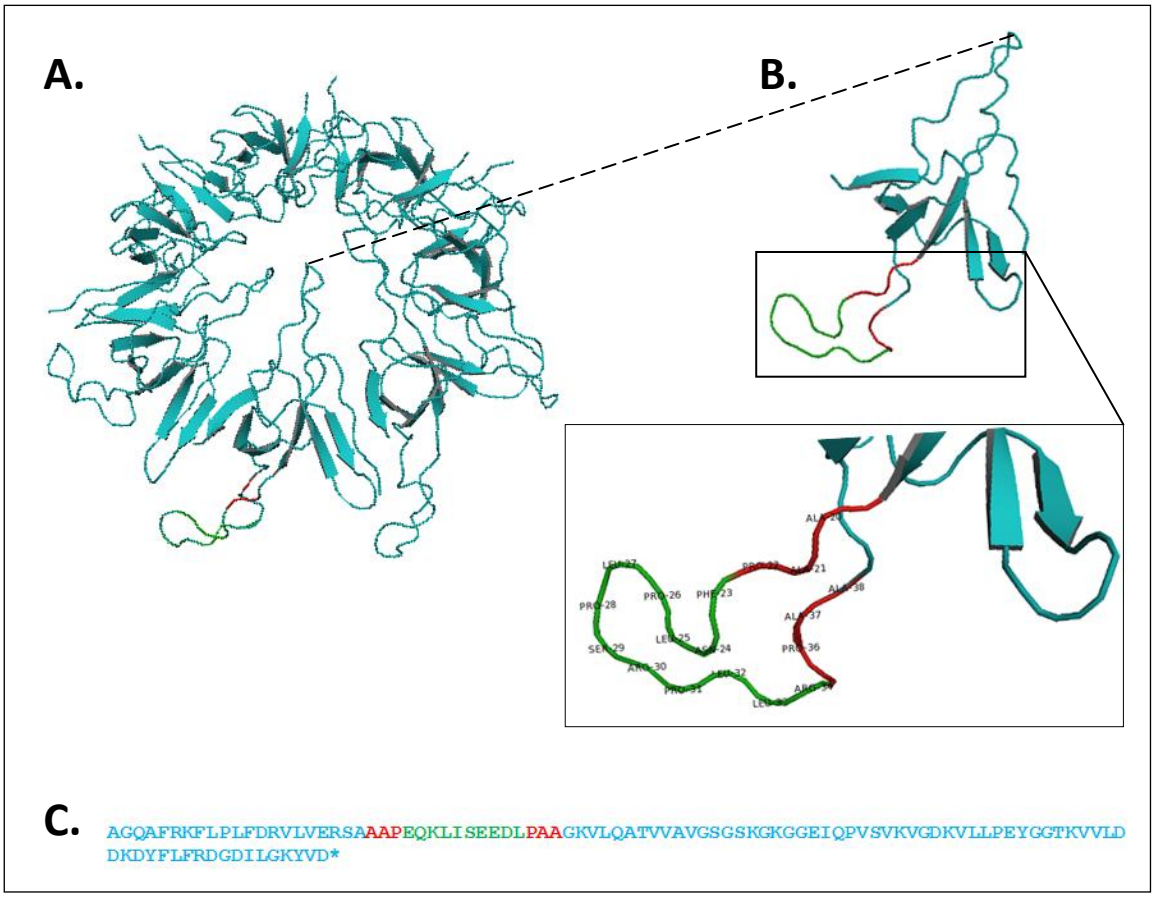


Figure 5.1 Three-dimensional model of CP7

Scaffold-based hCpn10 displaying CD44 specific P7 peptide. **(A)** 3D Model of Heptamer CP7. **(B)** Monomer CP7 highlighting the substituted mobile loop. **(C)** Amino acids sequence of CP7 (blue), mobile loop substituted with P7-peptide (green), and P1-linker (red). The 3D model of CP7 was based on PDB ID. 1WE3 (detailed in Chapter 3).

5.2 Research aims

- To design and produce a hCpn10 based scaffold (CP7) displaying a peptide recognising CD44 for utilisation as a diagnostic probe for cancer detection,
- To determine the oligomeric structure of CP7 and formation of hCpn10 heptamers, and
- To investigate the utility of CP7 as a CD44 cancer detector by determining the binding of CP7 to recombinant and native forms of CD44.

5.3 Material and methods

5.3.1 Materials

Table 5.1 details the relevant materials used in the research discussed in this study and their manufacturers/suppliers.

Table 5.1 List of specific applied materials

Item(s)		From
Cell/Media/Reagent		
Breast Adenocarcinoma (MDA-MB-468)		ATCC® HTB-132™
Gibco® GlutaMAX™ (1x)		*Life* technologies™
Gibco® RPMI 1640 Medium		*Life* technologies™
Gibco® Fetal Bovine Serum (FBS)		*Life* technologies™
Corning® Ultra-Low attachment cell culture flasks (250 mL/25 cm^2)		Sigma-Aldrich®
Corning® Costar® cell culture plates 12-well (flat bottom)		Sigma-Aldrich®
Alexa Fluor® 488 Protein Labeling Kit		Invitrogen™
VECTASHIELD® Mounting Medium with DAPI		Vector Lab. Inc.
Antibodies	**Label**	
Mouse monoclonal anti-human Cpn10	--	abcam®
Rabbit Anti-human Cpn10	FITC	antibodies-online.com
His (C-term) mAb Mouse (Novex®)	FITC	*Life* technologies™
Mouse monoclonal Anti-Human CD44	--	Sigma-Aldrich®
Rabbit Anti-Mouse IgG (H+L)	Alexa Fluor® 488	*Life* technologies™
Mouse mAb Anti-Human CD44	FITC	BD Pharmingen™
Peptide/Proteins		
P7 peptide (*HHHHHHAAPFNLPLPSRPLLR*)		*Biobest*
Recombinant Human CD44 (rCD44)		abcam®
Recombinant Human CD44 (rCD24)		abcam®
Bovine serum albumin (BSA)		Sigma-Aldrich®

5.3.2 Production and purification of recombinant CP7

Refer to general materials and methods in Chapter 2 for the production and purification of recombinant CP7. The purities of CP7 variant used in this study were ≥ 95% assessed by SDS-PAGE and SE-HPLC.

5.3.3 Determination of CP7 binding affinity and specificity for CD44

The binding affinity and specificity of the designed CP7 for CD44 was assessed using BLI-Octet, using recombinant CD44 (rCD44) as well as rCD25 and BSA as control. Using 96-well plate (200 µL/well), *His*-tagged CP7 (1, 10, and 100 nM in PBS) was immobilized and captured by Ni-NTA (NTA) Biosensors of BLI-Octet system for 120-180 sec, at 37 °C, with 1000 rpm shaking. The rCD44 was applied in a concentration-dependent manner (0.01-100 nM in PBS) for 180 sec, for binding and allowing the measurement of the association rate (k_a). This was then washed-out for 10 min in PBS, at 1000 rpm and 37 °C to measure the dissociation rate (k_d). Equilibrium dissociation constant (K_D) of CP7 was calculated using Octet Version 7.0 software, with a 1:1 binding model and global fitting tools. No protein binding was detected for controlled rCD24 and BSA and was therefore used for subtraction. Binding ability of hCpn10 (*His*-tagged) that captured by Ni-NTA biosensor for rCD44 was also not detected. The average K_D value reported in this study was the mean ± standard error of mean from three independent experiments.

5.3.4 Fluorescent conjugation of CP7

CP7 was covalently labelled with fluorophore DyLight 488 (DyLight™ protein label kit) following the manufacturers recommendation. In summary, 8 µL Borate Buffer (0.67 M, pH 8.5) was added to 100 µL of CP7 (1 mg/mL) in PBS. It was then added to DyLight pre-measured dye vial and was gently vortex before incubated for 1 hr at RT (in dark). 100 µL of protein/dye mixture was added into spin column and centrifuged at 1000xg for 1 min. Labelled CP7 was collected by adding 100 µL of labelling reaction to the column and centrifuged for 1 min at 1000xg. The labelled protein final concentration of 10 µM was determined by Equation 5.1. Protein absorbance was measured at 280 nm, the maximum absorption (A_{max}) of DyLight 488 dye is 493 and the correction factor (FC) is 0.147. Protein

molar extinction coefficient ($\varepsilon_{protein}$) of CP7 is 31290 M^{-1} cm^{-1}(while, ε of CP7 monomer is 4470 M^{-1} cm^{-1}). Purified 488-CP7 was stored in single-use aliquots at -20 °C.

$$\text{Protein concentration (M)} = \frac{A_{280} - (A_{max} \times CF)}{\varepsilon_{protein}} \times \text{Dilution factor} \qquad \textbf{Equation 5.1}$$

5.3.5 Cell lines and cell culture

MDA-MB-468 breast cancer cells overexpressing CD44 were cultured in log-phase growth in 20 mL Gibco® RPMI 1640 Medium supplemented with Gibco® GlutaMAX™ with 10 % Fetal Bovine Serum (FBS) at 37 °C in a humidified atmosphere (5 % CO_2), in Corning® Ultra-Low attachment cell culture flasks (250 mL/25 cm^2), to a final concentration of 1 x 10^6 cells/mL.

5.3.6 Fluorescence imaging analysis

Protein binding to cellular CD44 was investigated using fluorescence microscopy (EVOS®, Advanced Microscopy Group) by adding directly labelled 488-CP7 to cultured CD44 MDA-MB-468 cells (1 x 10^6 cells/mL) in 12-well Corning® Costar® cell culture plates. A 0.1, 0.5, 1, 5, and 10 µM of 488-CP7 per well incubated for 1 hr at 4 °C, and in separate CD44 MDA-MB-468 cells treated with primary antibody FITC-Mouse mAb Anti-Human CD44 (1:200 dilution, to final concentration of 5 µg/mL) used as control. Cells were washed with PBS three times, were observed under fluorescence microscope and images were viewed and analysed using Microsoft office picture manager software.

5.3.7 Immunofluorescence microscopy

CD44 expressed MDA-MB-468 cells were treated with CP7 (0.1-10 µg/mL), washed with pre-chilled PBS twice and then fixed with ice-cold methanol (5 min). Cells were blocked with PBS (pH 7.4), 1 % BSA, and 0.1 % Triton X-100, for cells permeabilisation and block non-specific interactions. Cells were then incubated with 1-10 µg/mL secondary antibody FITC-Rabbit Anti-human Cpn10 (1:1000 - 1:100 in series dilutions) at 4 °C for 1 hr. MDA-MB-468 cells were treated with 5-10 µg/mL control primary antibody Mouse mAb Human CD44 (1:200 and 1:100 dilution) and then followed by secondary antibody 1 µg/mL Alexa Fluor® 488 Rabbit Anti-Mouse IgG as control. Similarly cells were treated with 10 µg/mL hCpn10 followed by secondary 5 µg/mL FITC-Rabbit Anti-human Cpn10 antibody at 4 °C

for 1 hr. Cells were then finally washed with PBS three times and mounted onto microspore slides using antifade agent VECTASHIELD® mounting medium with DAPI for nuclear staining, then covered using coverslips.

In a direct approach, MDA-MB-468 cells were treated with labelled 488-CP7, following the procedures above with slight modifications. Cells were incubated with primary 0.5, 1, 5, and 10 uM 488-CP7 at 4 °C for 1 hr, then washed with pre-chilled PBS twice, and fixed with ice-cold methanol and blocked with PBS (pH 7.4), 1 % BSA, and 0.1 % Triton X-100. The control sample was treated with primary antibody FITC-Mouse mAb Anti-Human CD44 (1:100 dilution to final concentration of 10 µg/mL). Cells were washed with PBS three times and mounted onto microspore slides using VECTASHIELD® mounting medium with DAPI, and then coverslips were added. All slides were analysed using EVOS® fluorescence microscope.

5.3.8 Flow cytometry

Cell surface expression of CD44 were characterised by one-colour immunofluorescent staining using BD Accuri™ C6 Flow Cytometer System (BD Biosciences). MDA-MB-468 cells were grown as described previously, then resuspended at a concentration of 5 x10^6 cells/mL, prior to washing with pre-chilled Flow-cytometry Buffer (FB) of: PBS (pH 7.4), 0.1 % Sodium azide, 1 % BSA. For indirect detection approach, a 50 uL aliquot of cells added to equal volume of CP7 (1:2000 and 1:1000) to final concentration of 0.5-1 µg/mL and separately a primary Mouse monoclonal Anti-Human CD44 (1:1000, to final concentration of 1 µg/mL), both incubated at 4 °C for 1 hr on rotation. Cells were then washed with equal volume of FB. Cells were then resuspended on a 50 µL FB contained fluorescein isothiocyanate-conjugate secondary antibody; FITC-conjugated Rabbit Anti-human Cpn10 (1:2000 dilution) and Alexa Fluor® 488-Rabbit Anti-Mouse IgG (1:2000 dilution), respectively, and incubated at 4 °C (dark) for 30 min. Cells were then washed twice and resuspended in 500 µL FB containing 50 µL 4 % PFA. Negative controls consisted of grown cells treated identically except for primary CP7 which was replaced with hCpn10, that did not bind to MDA-MB-468 cells when detected with secondary FITC-conjugated Rabbit Anti-human Cpn10. Therefore, it was used for subtraction of nonspecific binding and gating.

For direct detection, a 50 uL aliquot of cells were added to equal volume of fluorophore-conjugate of 488-CP7 (1:1000 and 1:500 to final concentration of 1-2 µg/mL) and primary control FITC-mouse anti-human CD44 mAb (1:1000 to final concentration of 1 µg/mL), incubated at 4 °C for 1 hr in dark. Cells were then washed twice with FB, and then resuspended on 500 µL FB contained 50 µL 4 % PFA. Viable cells ≥ 10,000 analysed by using BD Accuri™ C6 flow cytometer. In this study, excitation was at 495 nm and fluorescence emission was detected by using the 520 nm filter for FITC/DyLight 488. Data were analysed using CFlow® Sampler software.

5.4 Results

5.4.1 Preparation of CP7 - expression and purification

The nucleotide sequence representative of CP7 with a *C*-terminus Histidine-tagged was cloned into the pET30a bacterial expression vector and was transformed into BL21(DE3) strain of *E. coli*. The expression of CP7 was induced by IPTG and protein expression conditions were further optimised as outlined in Chapter 2. The cells were harvested and resuspended in 50 mM Tris-HCl (pH 7.6), 0.5 M NaCl, and mechanically lysed with a single pass using EmulsiFlex-C5 homogeniser (at 15,000 psi). Cell lysates were recovered and mixed with a cocktail of protease inhibitor tablets and kept at low temperature (at 4 °C) to avoid protein degradation via proteases.

The resulting supernatant (from 14000xg, 15 min, at 4 °C) containing soluble fractions of *His*-tagged CP7 were subsequently captured by IMAC. With the use of IMAC, soluble fractions of CP7 were captured by loaded onto a HisTrap™ FF (5 mL) column and then eluted through a stepwise gradients of 10 and 100 % 0.5 M imidazole for elution. Protein fractions were collected and resolved by SDS-PAGE (4-12 % Bis-Tris) and visualised by staining gel with coomassie (Figure 5.2). The eluent of CP7 (pI 9.52) was buffer exchanged prior to polishing using gel-filtration chromatography. The final eluent CP7 was reconcentrated and stored in high homogeneity in 50 mM Tri-HCl (pH 7.6), 150 mM NaCl at working concentration of 1 mg/mL and stored at 4 °C for further protein characteristics.

CP7 was analysed by SDS-PAGE using 4-12 % Bis-Tris gels which confirmed the purity and expression levels of the sample (Figure 5.2). CP7 (≥ 95% purity) was further analysed for purity by size exclusion HPLC and exhibited a single peak that corresponded to an apparent molecular weight indicative of a heptamer, when compared to Mwt standards. The actual Mwt of hCpn10 is 70 kDa while the apparent Mwt according to the SE-HPLC chromatogram is between 90-100 kDa (Figure 5.3).

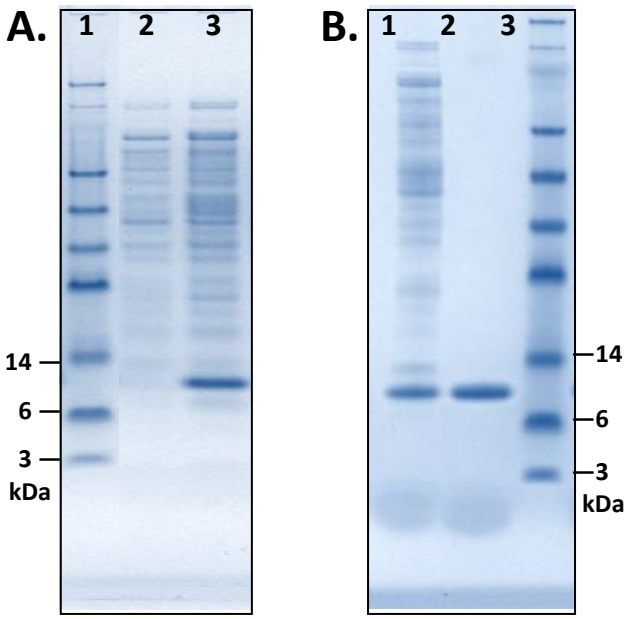


Figure 5.2 SDS-PAGE analysis of CP7 production and purification

Analysis of protein expression and purification of CP7 **(A)** Protein expression, Lane1, SeeBlue® Plus2 Mwt marker; Lane2, CP7 pre-induction; Lane3, CP7 (12kDa) post-induction. **(B)** IMAC protein purification, Lane1, soluble fractions cells lysed; Lane2, CP7 (12kDa) HisTrap™FF elute; Lane3, SeeBlue® Plus2 Mwt marker.

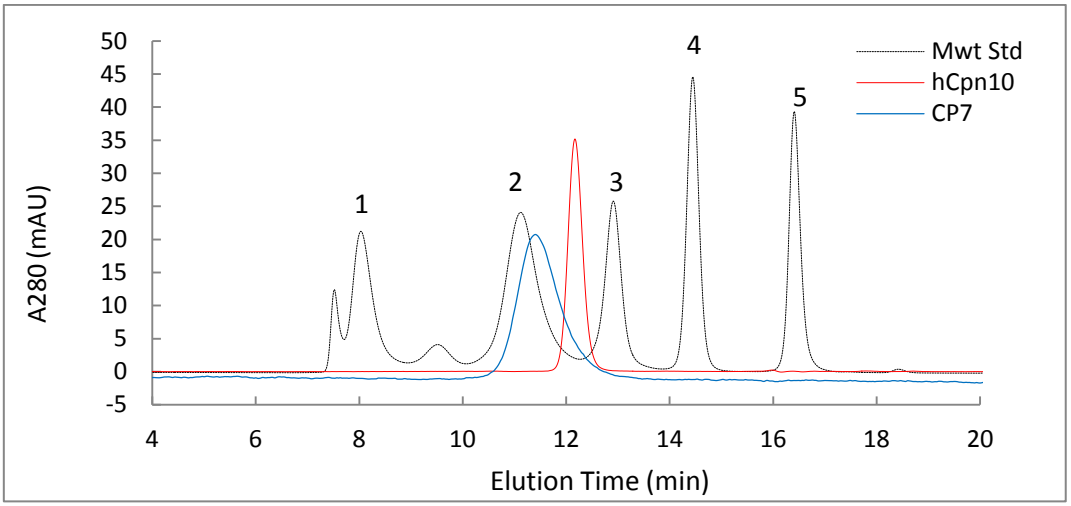


Figure 5.3 Analytical SE-HPLC for recombinant CP7 versus hCpn10

HPLC-based size analysis of CP7 and wild-type hCpn10. The elution time of each protein was detected in mAU at 280 nm wavelength. The dashed line indicates the elution time of various proteins in a Mwt standard mix and corresponds to: (1) Thyroglobulin (670 kDa); (2) BGG (158 kDa); (3) Ovalbumin (44 kDa); (4) Myoglobin (17 kDa); and (5) Vitamin B12 (1.35 kDa). These standards were used to determine the apparent Mwt of CP7 (Blue, 88 kDa) and hCpn10 (Red, 70 kDa).

5.4.2 Characterization of CP7 for CD44 binding

The specificity and binding affinity of CP7 and CD44 were assessed *in vitro* using the BLI-Octet technology which is a label-free technique that analyses real-time interaction between various molecules (Figure 5.4). The use of the BLI-Octet system has many advantages over traditional ELISA methods as it is more accurate, can be easily replicated and only requires a small sample size. Furthermore, the Octet system can determine the binding kinetics which cannot be achieved through conventional ELISA.

The CP7 variant has a potential use as a detection probe for certain cancers overexpressing CD44. The binding characteristics against rCD44 as well as rCD24 and BSA, which acted as a control, were assessed in further detail. The binding affinity of CP7 to rCD44 was determined to be 1.32 x10^{-9} M (± 0.08), (Figure 5.4).

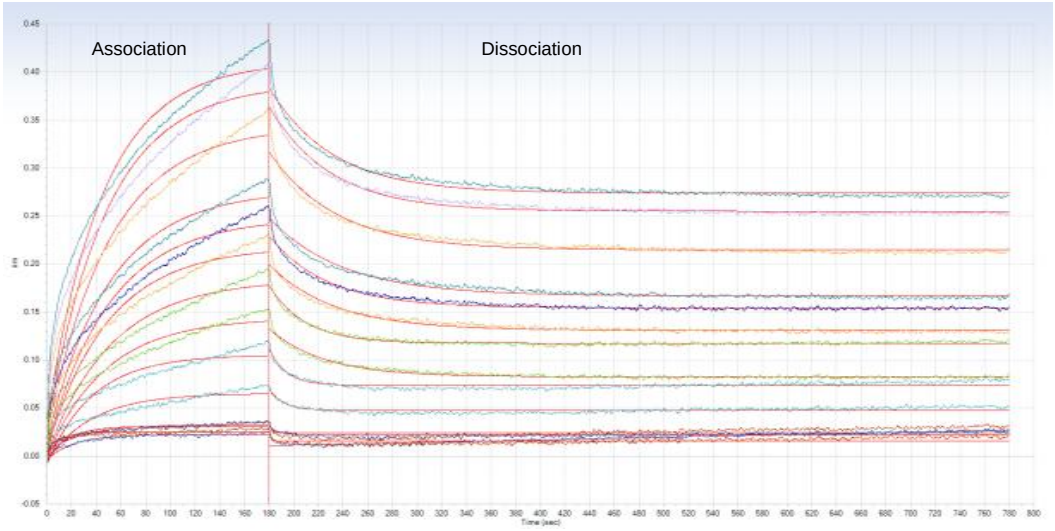


Figure 5.4 Kinetic analysis of CP7 binding to rCD44 determined by BLI-Octet

The binding curves depict the capture of *His*-tagged CP7 by Ni-NTA biosensors and the association of rCD44 at series concentrations (0.01, 0.05, 0.1, 0.5, 1, 5, 10, 25, 50, 75, 100 nM). The determined K_D of CP7 was 1.32 x10^{-9} (± 0.08) M while the association (k_a) and dissociation (k_d) rates calculated from binding curves were 4.29 x10^{5} (± 0.16) M.s^{-1} and 5.66 x10^{-4} (± 0.19) s^{-1}, respectively. Controls rCD24 and BSA were used to evaluate non-specific binding. Coloured curves represent experimental data and red curves represent the statistical fitting of the curves using 1:1 binding model. Octet Data Analysis 7.0 software was used for the kinetic and data analyses.

In addition, the binding characteristics of CP7 were further assessed in MDA-MB-468 breast cancer cell that overexpress CD44 through immunofluorescence microscopy using direct and indirect approaches. In the indirect approach, CP7 interaction with endogenous CD44 was detected by probing the complex with secondary rabbit anti-human Cpn10 antibody conjugated to FITC. A green fluorescein staining was observed after the cells were fixed with 4 % PFA followed by a series of intensive washes with PBS (Figure 5.5). A negative control of cells was probed with hCpn10 followed by the secondary FITC-rabbit anti-human Cpn10 antibody which showed no detection of fluorescence. In order to compare with the two positive controls, P7 peptide and anti-human CD44 antibody, the cells were incubated with *His*-tagged P7 peptide followed by FITC-Mouse anti-His (*C*-term) while monoclonal mouse anti-human CD44 antibody was detected by secondary Alexa Fluor® 488-Rabbit anti-mouse antibody. Both controls were visualised under the fluorescence microscopy (Figure 5.5).

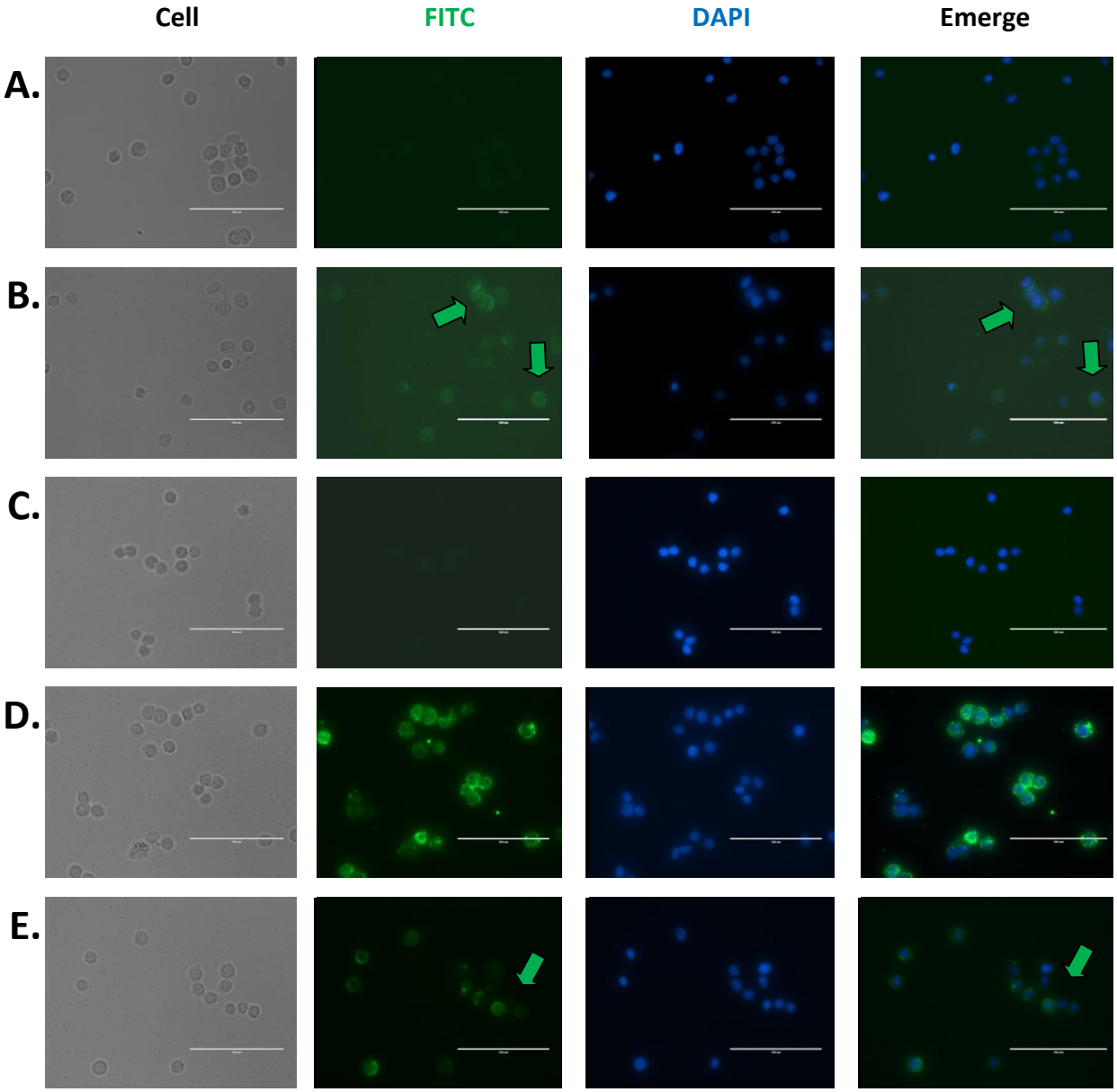


Figure 5.5 Immunofluorescence images of CD44 MDA-MB-468 cancer cells

Immunofluorescence images show breast cancer cells detected indirectly by fluorescein isothiocyanate-conjugate secondary antibodies. **(A)** Unlabelled cells only and counterstained with DAPI (blue) for nuclei staining. **(B)** Control P7 peptide, followed by FITC-mouse anti-His_6 (green, $2^{secondary}$) and nuclear stain, DAPI (blue). **(C)** Control hCpn10, followed by FITC-rabbit anti-Human Cpn10 mAb (green, $2^{secondary}$), and nuclear stain, DAPI (blue). **(D)** Designed **CP7**, followed by FITC-rabbit anti-Human Cpn10 mAb (green, $2^{secondary}$), and nuclear stain, DAPI (blue). **(E)** Control mouse anti-human CD44 mAb, followed by Alexa Flour® 488-rabbit anti-mouse IgG (green, $2^{secondary}$), and nuclear stain, DAPI (blue). Using EVOS® fluorescence microscopy and white bar shows scale of 100µm.

The binding of CP7 to CD44 was also detected directly by conjugating fluorescein DyLight® 488 to CP7 using DyLight® protein labelling kit for 488-CP7 (Figure 5.6). The binding of the fluorescently labelled 488-CP7 to MDA-MB-468 cells was also examined under the fluorescence microscopy (Figures 5.7 and 5.8). As previously described, cells were fixed with 4 % PFA and incubated with labelled 488-CP7. After incubation, the cells were visualised and there appeared to be low fluorescence intensity in relation to the control of FTIC-Mouse anti-human CD44 mAb (Figure 5.7). However, this could be due more to the technical challenges of directly conjugating fluorophore to CP7. The fluorescein-488 did not affect the inserted P7 peptide from binding to targeted CD44 nor did it affect intermolecular stability of CP7 which was confirmed by SE-HPLC.

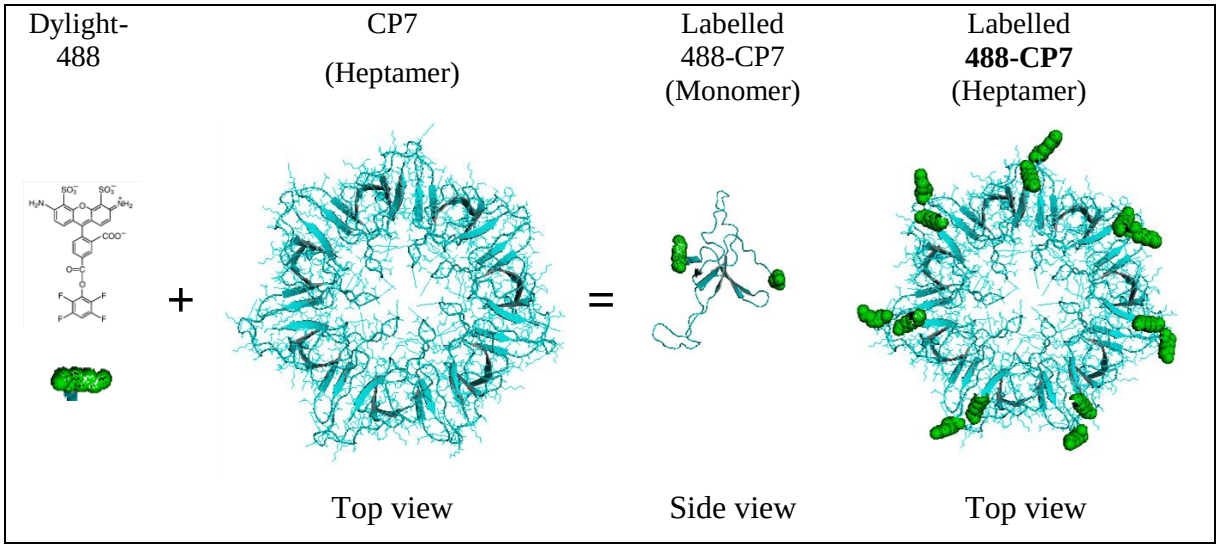


Figure 5.6 Labelling scheme for CP7 with fluorescein DyLight® 488

Process of labelling and possible conjugation sites at the *N*- and *C*-termini of CP7 using DyLight® protein labelling kit.

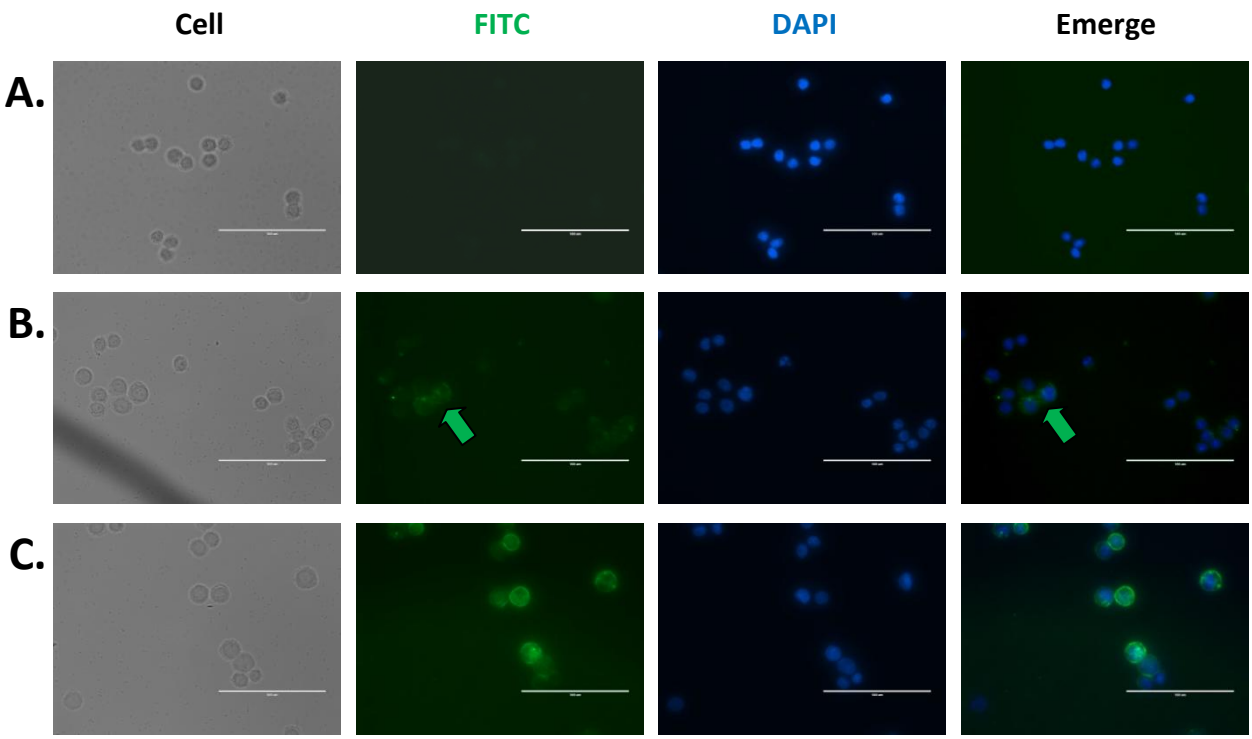


Figure 5.7 Immunofluorescence images of MDA-MB-468 cancer cells

Immunofluorescence images of CD44 MDA-MB-468 cancer cells using primary-conjugated CP7. **(A)** Unlabelled cells only and counterstained with DAPI (blue) for nuclei staining. **(B)** Labelled 488-CP7 (green, $1^{primery}$), and nuclear stain, DAPI (blue). Showing low fluorescence intensity due to labelling preparation optimization, despite confirmed binding to cells using FACS. **(C)** Control FITC-Mouse mAb anti-human CD44 (green, $1^{primery}$), and nuclear stain, DAPI (blue), using fluorescence microscopy (white bar shows scale of 100 µm).

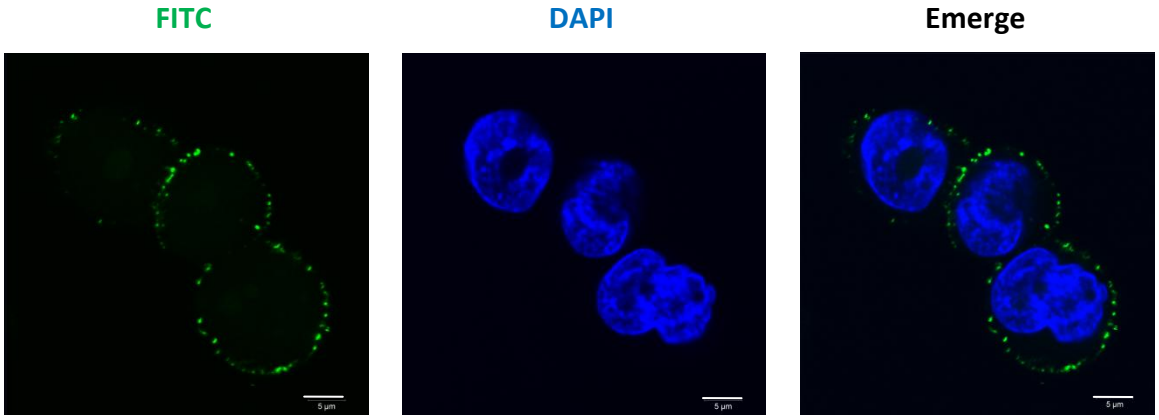


Figure 5.8 Immunofluorescence images of MDA-MB-468 cancer cells by confocal microscopy

Immunofluorescence images of CD44 MDA-MB-468 cancer cells directly detected by labelled 488-CP7 (green, $1^{primery}$), and nuclear stain, DAPI (blue), using confocal microscopy (white bar shows scale of 5 µm).

Prior to conducting the immunofluorescence assays of CP7 against the MDA-MB-468 cells (from both indirect and direct approaches), the interaction of the CP7 molecules with the breast cancer cells was assessed using flow cytometry in order to evaluate the specificity and binding efficiency of CP7 against CD44 on the cell surface. The same approaches of direct and indirect probing discussed above was conducted in the flow cytometry analyses and a similar observation was made whereby CP7 had a stronger binding affinity to the CD44 MDA-MB-468 cells relative to the anti-human CD44 antibody (Figure 5.9). Overall, the unlabelled CP7 and fluorescently labelled 488-CP7 both clearly showed a strong binding affinity to CD44 in both the experimental protocols (Figures 5.9 and 5.10).

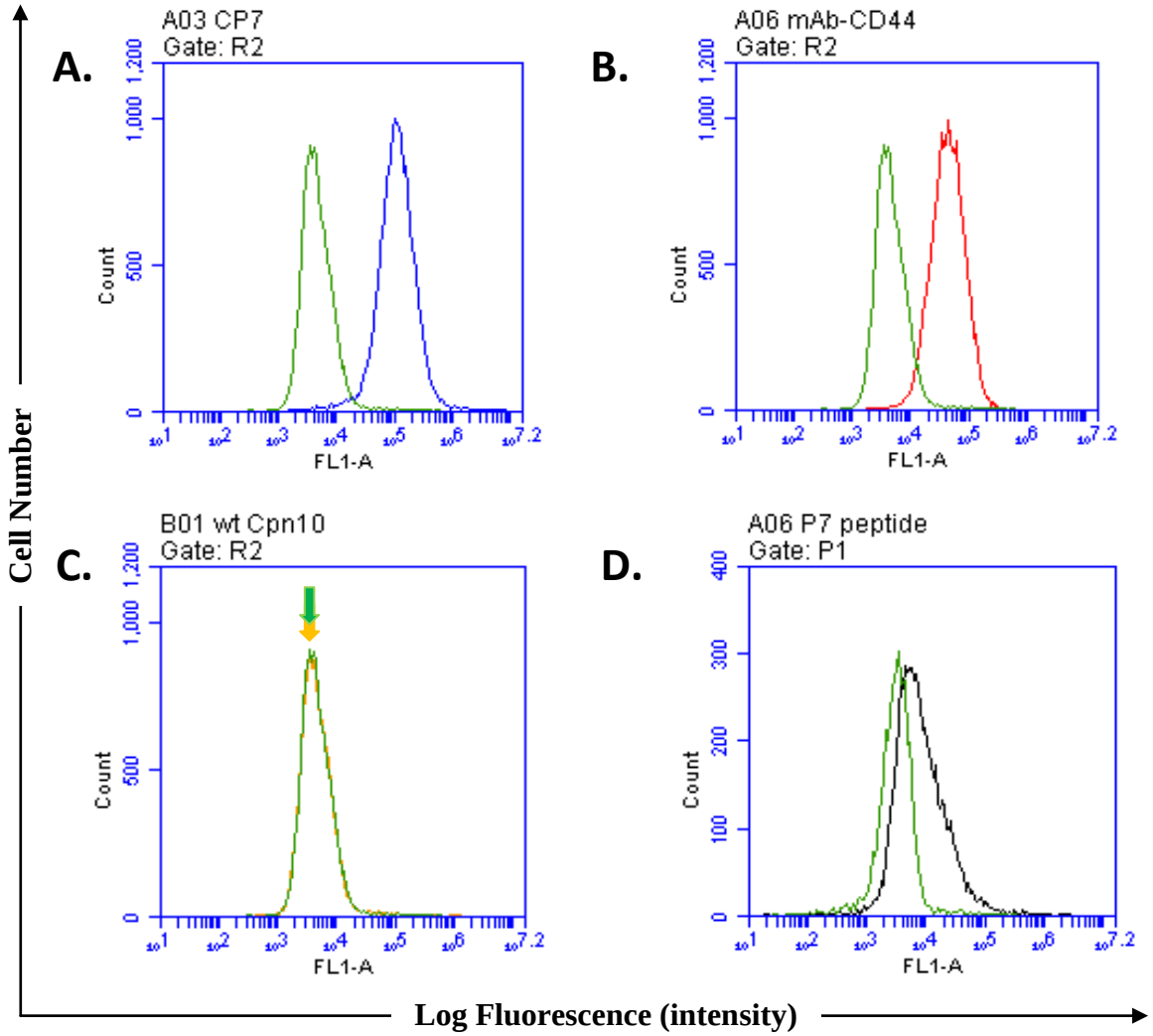


Figure 5.9 Flow cytometric analyses of CD44 expression on MDA-MB-468 cell surface

Overlay histograms of MDA-MB-468 cells indirectly detected by secondary fluorescein isothiocyanate-conjugate antibodies using Accuri™ C6 Flow Cytometer. Fluorescence intensity was expressed as a 4-decade log scale plotted versus cells number. **(A)** Cells incubated with **CP7** ($1^{primary}$), followed by FITC-Rabbit anti-Human Cpn10 (blue, $2^{scondary}$). **(B)** Cells incubated with Mouse mAb anti-CD44 Human (control antibody, $1^{primary}$) followed by Alexa Flour® 488-Rabbit Anti-Mouse IgG

(red, $2^{scondary}$). **(C)** Cells incubated with hCpn10 ($1^{primary}$), followed by FITC-Rabbit anti-Human Cpn10 (orange, control, $2^{scondary}$). Unlabelled Cells (green) used for illustrating the non-specific binding and also for gating. **(D)** Cells incubated with P7 peptide (control, $1^{primary}$), FITC-Mouse Anti-His_6 (green, $2^{secondary}$). Data are representative of one from three separate experiments. Acquisition of > 5,000 events was performed.

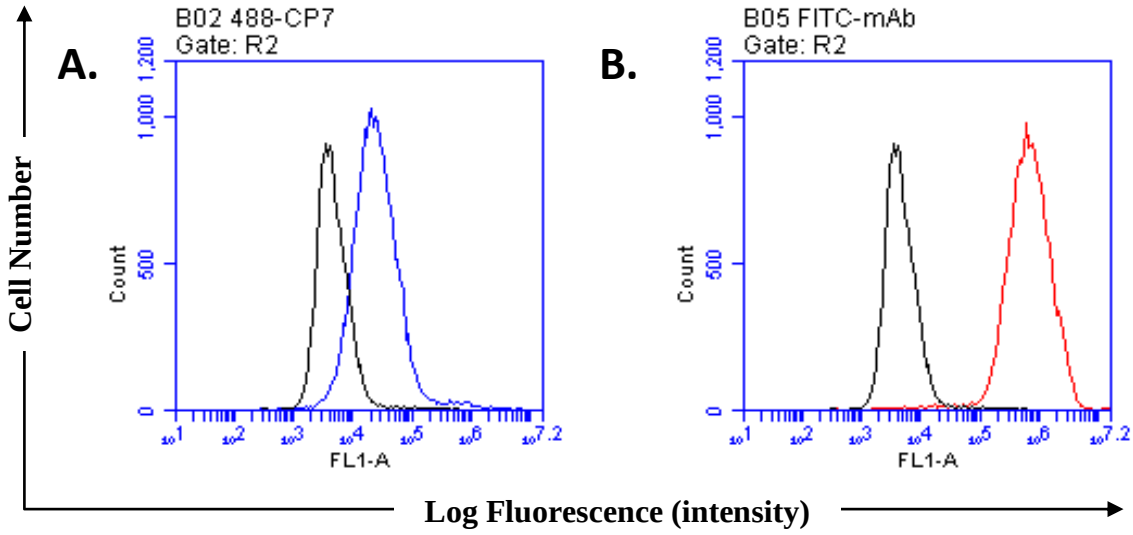


Figure 5.10 Flow cytometric analyses of CD44 expression on MDA-MB-468 cell surface

Overlay histograms of CD44 MDA-MB-468 cells stained with primary detectors using Accuri™ C6 Flow Cytometer. **(A)** Cells directly detected by labelled 488-CP7 (blue, $1^{primary}$); and **(B)** Cells directly detected by FITC-Mouse mAb anti-CD44 Human (red, $1^{primary}$, control); unstained cells (Black) illustrating the non-specific binding and gating. Samples were prepared in similar manner as of Figure 5.9.

5.5 Discussion

By 2020, it is expected that the total number of cancer-related deaths will be over 30 million worldwide (Jayamohan et al., 2013, WHO, 2002). One of these types of cancer is breast cancer which is a serious disease affecting 1.3 million people worldwide and accounts for an annual mortality of over 500,000 (Zoon et al., 2009, Hortobagyi et al., 2005). Survival rates of cancer sufferers have, however, improved significantly as advancements have been made in early detection of this cancer. This is not only important for breast cancer, but also for all types of cancers where an early diagnosis during Stages I or II can ensure a higher five-year survival rate over 50 % (Walters et al., 2013, Wise, 2013, Jemal et al., 2010, WHO, 2002). Relapses of cancer patients due to poor prognosis are responsible for the decrease of the survival rates of breast cancer patients from 80 to 60 % with local recurrence, and from 80 to 10 % for metastatic cancer (To et al., 2010, Jaggupilli and Elkord, 2012).

Much research has been invested in the discovery and detection of a number of candidate biomarkers as they have the potential to aid clinicians in the management of various cancers whereby staging, progression and potential regression of the tumour can be monitored. CD44 is a cell surface protein that is vital for a number of cellular functions such as cell adhesion, migration, interaction with the endothelium and subsequent progression towards metastasis (Zoeller, 2011, Naor et al., 2002). CD44 has been identified as a biomarker where high expression levels have been observed in breast, prostate, colon, leukaemia, head and neck, and brain cancers (Trapasso and Allegra, 2012, Jaggupilli and Elkord, 2012, Negi et al., 2012, Akisik et al., 2002, Khoursheed et al., 2002, Kaaijk et al., 1995, Olsson et al., 2011, Takaishi et al., 2007, Dhir et al., 1997).

Importantly, CD44 was also found to be highly expressed in cancer initiation cells of CSCs of several solid tumours (Ishimoto et al., 2010, Hurt et al., 2008). Thus, the detection of CD44 cancer cells is crucial as CD44 CSCs are responsible for initiation of tumours post chemo- and radio-therapies due the inherent resistant nature of CSCs and have quiescent or slow-cycling properties (Balic et al., 2013, Negi et al., 2012, Ishimoto et al., 2010, Hurt et al., 2008). Studies conducted by To et al. (2010) have determined that CD44 is expressed at high levels in a population of CSCs in breast cancer and as such, this protein serves as a promising marker for detection of cancer.

Recent work conducted by Park et al. (2012) using phage display technology resulted in the isolation of a peptide P7, with high affinity and specificity towards CD44. This study focussed on enhancing the binding properties of the P7 peptide by engineering P7 into the mobile loop of hCpn10, generating a NME designated CP7 (Figure 5.1). Utilising P7, the development of CP7 described in this research, with specificity towards CD44, would have utility as a diagnostic probe, as this could provide an opportunity to detect, for example cancer during its early stages and improve therapeutic outcomes. The design of CP7 was initially optimized *in silico* through MD simulation studies, which indicated on a theoretical basis that P7 could be stabilised within the hCpn10 scaffold using a P1-linker for CP7, as discussed in Chapter 3. Therefore, all results and discussions of studies refer to the optimized $CP7_{CD44\text{-}P1}$, and designated CP7.

CP7 was designed and modelled *in silico* by incorporating P7 peptide (*FNLPLPSRPLLR*, 12 residues) into the mobile loop of hCpn10 and produced using recombinant DNA technology. The P7 peptide was inserted between *Ala21* and *Gly38* with the addition of a P1-linker (i.e. at the *N*- and *C*-junctures of the loop) for the stabilization of the inserted non-native P7 oligopeptide (Figure 5.1). Also, a *C*-terminus histidine-tag was incorporated into CP7 to facilitate a high purity yield of this variant (Figure 5.2). CP7 was produced as a soluble protein and assembly into a heptamer, which was confirmed by SE-HPLC (Figure 5.3).

The functional characteristics of CP7 were further verified for their potential utility as a diagnostic probe for CD44/biomarker. However, in order to be effective, CP7 must first have a higher binding affinity and avidity towards CD44 and furthermore, the CP7 variant must be highly stable under various harsh laboratory conditions. The binding kinetics of CP7 with human rCD44, were initially measured using a BLI-Octet system at various concentrations (Figure 5.4). CP7 exhibited higher affinity towards rCD44 in the nanomolar range (1.32×10^{-9} M), compared to that of P7 peptide alone. Also, CP7 illustrated high specificity towards rCD44 by lacking interactions with rCD24 and BSA (Figure 5.4).

Two general methods are regularly applied to detect biomarkers, based on a direct or indirect approach. The direct method involves the use of a conjugated antibody/protein such as a FITC-labelled mAb that recognizes a specific antigen. The method is simple and fast as it involves just one step and less number of wash procedures. However, the sensitivity of the primary-labelled antibody/protein is low due to low signal amplification. Indirect methods involve probing the target antigen with an unlabelled antibody/protein followed by incubating the bound antibody with a labelled secondary antibody that recognizes the primary antibody/protein. In contrast to the direct approach, this approach is more sensitive and produces higher signal intensity. Both direct and indirect approaches were then applied to CP7 variant to assess its interaction with endogenous CD44 in cancer cells.

In this study, the efficacy of CP7 variant was assessed further using the adenocarcinoma breast cancer cells MDA-MB-468, that is known to overexpress the transmembrane CD44. Other cell lines, such as MCF7 breast cancer cells, may require treatment with radiation to enrich the expression of CD44 (Lagadec et al., 2010, Phillips et al., 2006). In fact, it has been reported that radiation was found to result in a doubling of CD44 expression. In this study, co-localization of CD44 and CP7 variant was observed in untreated MDA-MB-468 cells

(i.e. no radiation were required). The immunofluorescence imaging and flow cytometry approaches both revealed that CP7 variant was specific to CD44 in MDA-MB-468 cells, as evidenced by a higher fluorescence intensity detected in both analyses (Figures 5.5 to 5.10).

Initially, the indirect approach was applied when MDA-MB-468 cells were probed with CP7 and stained with a rabbit anti-human Cpn10 mAb that was conjugated to FITC. The results of the immunofluorescence assay and flow cytometry reveal that CP7 bound with high affinity to CD44 (Figures 5.5 and 5.9). Furthermore, CP7 had higher fluorescence intensity relative to both controls, the P7 peptide (*AAPFNLPLPSRPLLR*) and mouse monoclonal anti-human CD44 antibody, possibly due to two reasons. Firstly, CP7 variant contains seven identical monomers and consequently seven binding sites with displayed P7 peptides, which in turn increases the avidity of CP7 for binding to CD44 on the cell surface; secondly, it is highly probable that the secondary mAb of FITC-rabbit anti-human Cpn10 bound to one or more subunits of the CP7 variant, and is actually one of the seven identical subunits.

In the direct approach, CP7 was directly conjugated to a DyLight 488 fluorophore, called 488-CP7. The DyLight 488 was coupled at the lysine residues of CP7 (Figure 5.6) and was subsequently used as a primary probe for CD44 but with less washing steps. The interactions were verified through flow cytometry and immunofluorescence microscopy imaging, as previously discussed. Despite the binding detection, the signal intensity of labelled 488-CP7 was low in relation to the indirect approach described above (Figures 5.7, 5.8 and 5.10). This could be due to the sub-optimization of the conjugation conditions of the fluorophore 488 to CP7 variant or due to the low number of free lysine residues that are found in both *N*- and *C*-termini of the CP7. This step, however, will require further optimisation in the future before *in vivo* studies can be conducted in order to assess the efficacy of directly labelled CP7's binding to CD44 for example of breast, brain, colon, leukaemia, or prostate cancers.

5.6 Conclusion

CD44 has emerged as a promising biomarker for the potential detection of different type cancers as well as of CSCs. This study focused on designing a diagnostic probe by incorporating the P7 peptide into the mobile loop of hCpn10 to detect endogenous CD44, overexpressed on MDA-MB-468 cells. The *in vitro* results indicated that CP7 is a stable NME, can be potentially utilised as a low cost diagnostic probe for cancer cells, and can be

used as an alternative to antibodies in various immunologically-based assays. Nevertheless, further validation of CP7 variant is required in order to determine whether this NME has potential clinical value.

The hCpn10 scaffold could have a variety of applications across a broad spectrum, including environmental monitoring, medical diagnostics, purification reagents and therapeutics. Naked peptides have generally low affinity, and the bestowing of avidity through incorporation into the mobile loop of hCpn10 could render previously known peptides useful for the above applications.

Chapter 6. Preliminary Results of Anti-Aangiogenic Candidates

6.1 Introduction

Angiogenesis is a mechanism whereby new blood vessels form by branching out from pre-existing ones. This process is highly regulated by a number of pro- and anti- angiogenic factors, that can for example, be triggered in response to physical trauma in order to facilitate proper wound healing. In addition, angiogenesis can be artificially stimulated in order to treat myocardial and peripheral perfusion (Bridoux et al., 2012, Atluri and Woo, 2008).

Vascular endothelial growth factor receptor (VEGFR) and fibroblast growth factor receptor (FGFR) are receptors involved in a number of diseases and are therefore novel targets for biopharmaceutical development with the intention of managing various cancers. A number of strategies have been pursued in order to develop anti-angiogenic therapeutics for the treatment of cancer. This is based on the premise that any absence of new blood vessels would prevent further tumour malignancy and reduce the chances of metastasis (Figure 6.1). Nevertheless, balancing between physiological and pathological angiogenesis is a significant challenge. In hypoxic conditions, malignant tumours secrete a number of growth factors such as VEGF or FGF in order to trigger the formation of new blood vessels that will eventually vascularise the tumour and facilitate the growth and potentially initiate tumour metastasis (Hummers et al., 2009, Teicher, 2010). Indeed, inhibiting angiogenesis is a major part of cancer treatment, for example, Avastin® is used to treat lung cancer by inhibiting angiogenesis (Ferrara et al., 2005).

VEGF is a pro-angiogenic factor and is well known for its role in activating angiogenesis in the vasculature. VEGF bind to three receptors, namely; VEGFR1, VEGFR2, and neurophilin1 (NR1). VEGF interacts primarily with VEGFR2 in triggering endothelial proliferation, migration and survival (angiogenesis signalling). Creating a VEGFR2 receptor antagonist with high affinity would hinder the interaction between VEGF and VEGFR2. This in turn would prevent phosphorylation and the subsequent initiation of angiogenesis through the mitogen activated protein kinase (MAPK) pathways (Tarrega et al., 2010).

The FGF family is composed of more than twenty-three signalling molecules that induce cell proliferation, differentiation and migration through interactions with the FGF receptors (FGFR1 to 4) on cell surface (Coutts and Gallagher, 1995). Basic FGF (bFGF) is a member of FGF family with high expression levels, along with its cognate receptors, are implicated in malignancy, angiogenesis and metastasis (Yamazaki et al., 1997, Giri et al., 1999). As such, bFGF serves as a potential target for cancer therapy whereby an anti-bFGF binder can effectively inhibit endothelial cell proliferation *in vivo* (Hori et al., 1991, Asano et al., 1995) and thus prevent bFGF-induced angiogenesis. Bevacizumab (Avastin®), an anti-VEGF monoclonal antibody that was approved by the FDA, is currently used in the clinic for the management of cancer whereby angiogenesis is inhibited (Cohen et al., 2007). Other small molecule drugs that are reported to have anti-angiogenic activity are sorafenib (BAY 43-9006) (Greenwald et al., 2013) and sunitinib (SU11248) (Baumann et al., 2012), which target the receptor tyrosine kinases. Nevertheless, these drugs have severe side effects due to their broad spectrum reactivity against a number of receptors kinases.

bFGF and VEGF are critical in promoting tumour angiogenesis and their high expression levels were correlated with vascularity and malignancy (Gacche and Meshram, 2014). VEGF promotes mitogenic activity on the endothelial cells, while bFGF synergistically enhances the mitogenic effect of VEGF. Both molecules have a direct effect on stimulating endothelial and tumour cell proliferation. In essence, this blocker could then potentially terminate the endothelial cell survival, proliferation and migration (Figure 6.1). In this study, two peptides designated P11 and R2 were identified that bind bFGF and VEGF receptors, respectively. These peptides were substituted for the mobile loop of hCpn10, to create NMEs designated $CP11_{bFGF\text{-}P1}$ and $CR2_{VEGF1\text{-}P2}$. The research reported in this chapter details the expression and purification of $CP11_{bFGF\text{-}P1}$, and includes some limited characterisation.

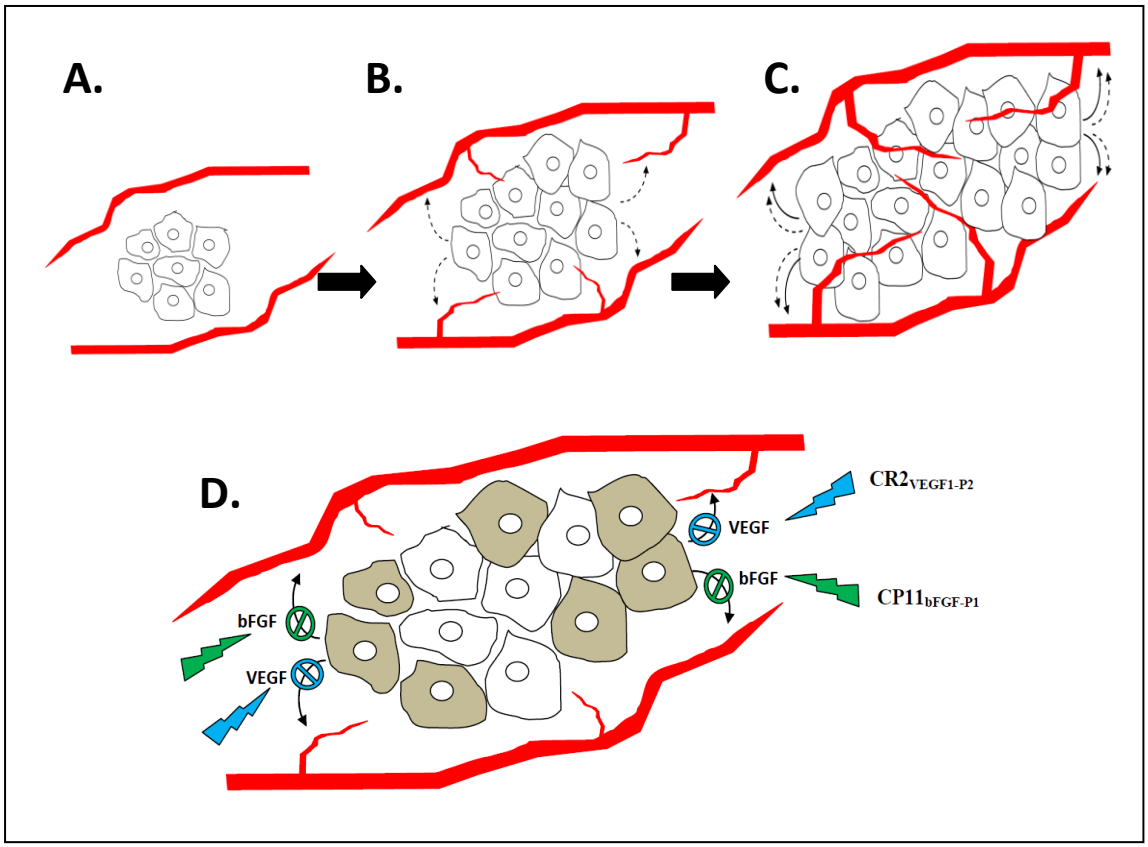


Figure 6.1 The role of angiogenesis factors in stimulating tumour growth

The premature tumour cell development for obtaining a vascular supply. **(A)** The beginning of the tumour mass is small and can obtain nutrients and oxygen from existing blood vessels. **(B)** The tumour grows beyond the capacity of nearby blood vessels and then releases soluble pro-angiogenic factors (e.g. bFGF and VEGF) which promote the development of new blood vessels from major vessels. **(C)** The new sprouting blood vessels will provide more nutrients and oxygen for tumour to grow beyond 3 mm in size. **(D)** Two proposed and designed anti-angiogenic candidates named $CP11_{bFGF\text{-}P1}$ and $CR2_{VEGF1\text{-}P2}$ are currently under development and can be applied for stopping and preventing further angiogenesis sprouting and thus causing tumour growth diminutions.

6.2 Materials and methods

6.2.1 Protein expression and purification

Following the standard methodology of recombinant protein expressions and purifications described in Chapter 2, the variant $CP11_{bFGF\text{-}P1}$ was designed *in silico* by inserting P11 peptide between *A21* and *G38* of hCpn10, expressed in *E. coli* BL21 (D3) as soluble protein, and purified by IMAC (≥ 90%) as determined by SDS-PAGE and GF.

6.2.2 Protein characterisation

Native Mwt of $CP11_{bFGF-P1}$ was determined by Size-exclusion (SEC) using high-resolution Superdex® 200 10/300 GL column following standard methodology of gel filtration chromatography (Section 2.7.2.3, Chapter 2).

6.3 Results and discussion

Given that both FGFR and VEGFR2 are potential target receptors for cancer treatment, two receptor antagonists were designed using hCpn10 as a molecular scaffold. The P11 peptide (Li et al., 2012), which is reactive to FGFR and the R2 peptide (Vicari et al., 2011), which is reactive to VEGFR2, were incorporated into the mobile loop region of hCpn10 to create $CP11_{bFGF-P1}$ and $CR2_{VEGF1-P2}$, respectively. Based on previous studies where hCpn10 was assessed as a potential scaffold for peptide display, it was hypothesised that $CP11_{bFGF-P1}$ and $CR2_{VEGF1-P2}$, would be more potent than the naked peptides, in preventing the interactions of bFGF and VEGF with their cognate receptors (Figure 6.1).

NME $CR2_{VEGF1-P2}$ is part of research study reported elsewhere (Recinos and Mahler, unpublished). Herein, $CP11_{bFGF-P1}$ is only reported. The NME $CP11_{bFGF-P1}$ was successfully expressed in *E. coli* BL21 (DE3) as soluble proteins. *His*-tagged $CP11_{bFGF-P1}$ (11.6 kDa) was purified using IMAC, and was then subsequently characterised using the standard methodology of SDS-PAGE outlined in Chapter 2 (Figure 6.2). SEC chromatography indicated that NME $CP11_{bFGF-P1}$ with the insertion of stabilizer P1-linker (i.e. proline-linker as detailed Chapter 3) at the *N*- and *C*-junctures of the mobile loop was successfully assembled into a native heptamer-like configuration in comparison to hCpn10 (Figure 6.3).

$CP11_{bFGF-P1}$ and $CR2_{VEGF1-P2}$ are proposed as NMEs with anti-angiogenic activity, whereby they are able to block FGFR and VEGFR2 receptors which promote the development of new blood vessels from major vessels. In another therapeutic scenario, $CP11_{bFGF-P1}$ and $CR2_{VEGF1-P2}$ could be utilised in combination at a 1:1 ratio, targeting both receptors simultaneously. The mechanism of action could work in a similar manner as that proposed for the novel dual VEGF-R2/bFGF-R inhibitor Brivanib, which is currently in clinical development (Dempke and Zippel, 2010).

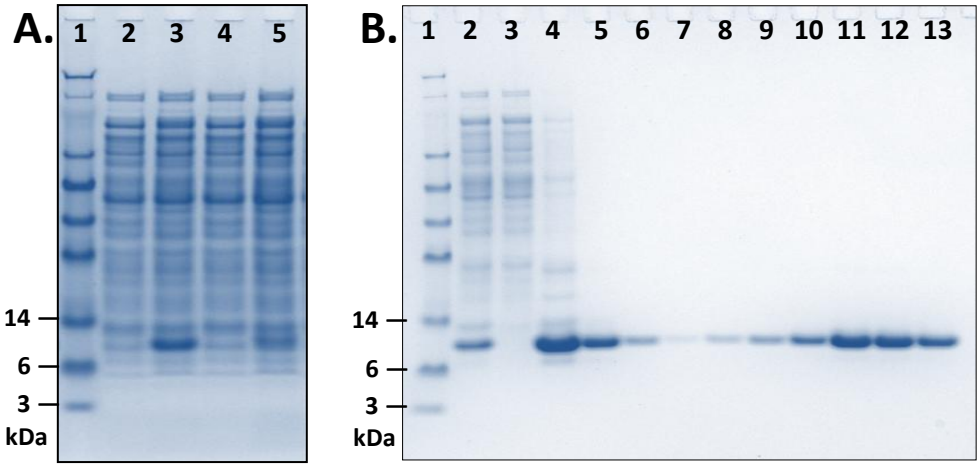


Figure 6.2 SDS-PAGE of $CP11_{bFGF\text{-}P1}$ expression and purification

Representative SDS-PAGE analyses of $CP11_{bFGF\text{-}P1}$. **(A)** $CP11_{bFGF\text{-}P1}$ (11.6 kDa) expression in *E.* coli BL21(D3), detail as following; Lane1: SeeBlue® Plus2 Mwt std marker, Lane2: pre-expression cell lysate (soluble), Lane3: $CP11_{bFGF\text{-}P1}$ post-expression cell lysate (soluble), Lane4: pre-expression cell lysate (insoluble), Lane5: $CP11_{bFGF\text{-}P1}$ post-expression cell lysate (insoluble). **(B)** $CP11_{bFGF\text{-}P1}$ (11.6 kDa) IMAC fractions, Lane1: SeeBlue® Plus2 Mwt std marker, Lane2: pre-load, Lane3: column unbound proteins flow-through, Lanes4-7: 10 % washed 0.5 M imidazole, Lanes8-13: 100 % 0.5 M imidazole purification fractions.

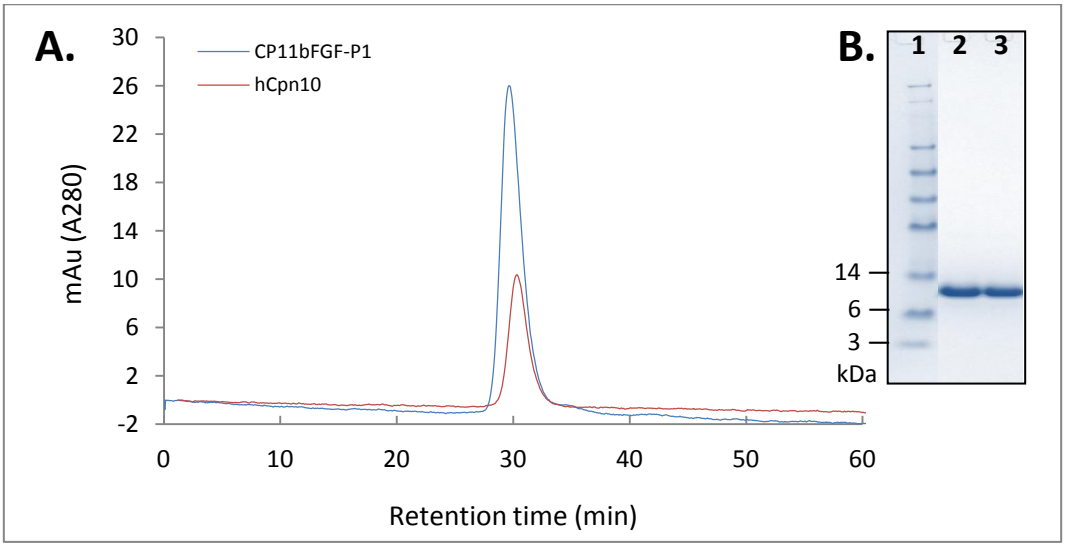


Figure 6.3 Analytical size-exclusion and SDS-PAGE of recombinant $CP11_{bFGF\text{-}P1}$

Representative SEC analysis of $CP11_{bFGF\text{-}P1}$. **(A)** SEC chromatogram of $CP11_{bFGF\text{-}P1}$ (blue, 75.5 kDa) and hCpn10 (red, 70 kDa) determined by Superdex® 200 10/300 GL prepacked column. Notably, $CP11_{bFGF\text{-}P1}$ apparently formed heptamer-like configuration in comparison to hCpn10 (control). **(B)** SDS-PAGE of SEC fractions; Lane1: SeeBlue® Plus2 Mwt std marker, and Lane2-3: reduced and non-reduced $CP11_{bFGF\text{-}P1}$ (11.6 kDa), respectively.

6.4 Future directions

$CP11_{bFGF\text{-}P1}$ and $CR2_{VEGF1\text{-}P2}$ were successfully expressed and purified. Both NMEs were confirmed to form a native-like heptamer by SEC-chromatography. Further characterisation studies associated with the activity and potency of $CP11_{bFGF\text{-}P1}$ and $CR2_{VEGF1\text{-}P2}$ are required. These studies could include *in vitro* assays using traditional methods such as ELISA and real-time BLI to assess the binding between the NMEs and FGF and VEGF receptors and to determine binding parameters (k_a, k_d and K_D). Additional studies can be conducted in order to assess anti-tumour activity through methylthiazole tetrazolium (MTT) colorimetric assays (Nawa et al., 2012). The inhibitory effects on cell cycle progression can be assessed through flow cytometry or by measuring the inhibition of MAPK activation (Tarrega et al., 2010). The anti-angiogenic effects can be assessed on chick embryo chorioallantoic membrane (CAM) (Li et al., 2007, Martinez-Madrid et al., 2009), by studying the level of neovascularisation that is either enhanced by bFGF or VEGF versus inhibited by $CP11_{bFGF\text{-}P1}$ or $CR2_{VEGF1\text{-}P2}$. Further studies could to be conducted *in vivo* using xenograft models to investigate the delay in tumour development. In summary the production of $CP11_{bFGF\text{-}P1}$ and $CR2_{VEGF1\text{-}P2}$ has laid the groundwork for further research and development of these two NMEs as biologic drugs for cancer treatment.

Chapter 7. Conclusion and Future Directions

The research outcomes presented in this thesis have made a valuable contribution in the area of molecular scaffolds, and has indeed shown that human Cpn10 is another molecular scaffold capable of presenting and displaying peptides in a fashion that allows binding of the peptides to target proteins. As has been pointed out, monoclonal antibodies are nature's scaffold for presenting peptides through the CDR loops, that are able to bind to a myriad of antigenic determinants found in nature. Antibodies are established therapeutic and diagnostic agents, and there are now around 35 approved globally and 100s in clinical trials for treatment of disease indications, predominantly chronic diseases such as cancer and inflammatory diseases. Antibodies are the key reagents in diagnostic test kits and other forms of immunoassays. Although antibodies are very successful as therapeutic and diagnostic agents, the intellectual property landscape is complex and is a hurdle in the development of new antibody-based NMEs.

The discovery of other molecular entities that are capable of behaving like antibodies in "beyond antibody" approaches has opened up a new area in protein therapeutics and diagnostics, and has led to the development of novel intellectual property. The common attribute of molecular scaffolds is that they possess domains (typically loop structures, but also flat structures, inherent in the DARPins (ankyrins), for example) where the endogenous loop peptide can be substituted with other peptides. These peptides are typically preselected to bind to an antigenic determinant (for example, from a peptide phage display library), in the same way that antibodies may be selected that bind to antigenic determinants on proteins through antibody phage display technology. Some of the more successful scaffolds include the Fibronectin (Monobody) and lipocalins (Anticalins), both which have candidate NMEs in clinical development. More recently another novel molecular scaffold has been reported termed FynomAb, derived from the human Fyn SH3 domain that is highly conserved among different species (Brack et al., 2014). FynomAb have two loops within the molecule, and peptides can be substituted within these loops that bind to an antigen of choice. FynomAb have no cysteine residues, are of small size (7 kDa) and are highly stable.

Human Cpn10 is a chaperone protein and is associated with protein folding, but has also been reported to have anti-inflammatory properties (Chapter 1). Regardless of the endogenous

activity of hCpn10, the loop structures within the monomer (mobile loop and roof loop) can potentially be substituted with peptides that have been preselected to bind to an antigenic determinant. The mobile loop was chosen in this study, as it has been previously shown that the loop can be deleted without compromising assembly of heptamer from monomer, and preliminary data had suggested that the loop could be substituted with other peptide sequence (Mahler, personal communication). Like FynomAb, hCpn10 does not contain any cysteine residues, which can have some advantages when expressing in *E. coli*. Also like FynomAb, hCpn10 is soluble when expressed in *E. coli*, which simplifies production. An added advantage of hCpn10 as a molecular scaffold is that the monomer assembles into a stable heptamer. The heptameric structure has seven peptides displayed, and so will impart an avidity which considerably increases the apparent affinity of heptamer over monomer, in the same way that avidity contributes to the apparent affinity of the pentameric IgM immunoglobulin.

The development of hCpn10 as a molecular scaffold has not been without technical challenges, and initial studies showed simple loop substitution did not always lead to heptamer formation. Indeed with some examples, ideal heptamer assembly did not occur and resulted in dimers, trimers, tetramers and other higher order oligomers and aggregates. The mixed oligomers and aggregates were partitioned between soluble and insoluble (inclusion body) fractions. In order for hCpn10 to function as a molecular scaffold, stable heptamer formation upon generic peptide substitution is required. To this end, molecular modelling studies revealed the propensity for some peptides within the mobile loop to interact with residues of the β-barrel interface through salt bridging, for example, and disrupt stable heptamer formation. The insertion of alanine/proline linkers at the *N*- and *C*-terminal junctions of the mobile loop was effective in spatially orienting the mobile loop away from the β-barrel interface, facilitating stable heptamer formation. Although some further optimisation of the hCpn10 molecular scaffold may be required, the research outcomes have demonstrated that the mobile loop can indeed be substituted with non-native peptides, to create NMEs with binding activity to other proteins. Thus in a “beyond antibody” approach, the hCpn10 molecular scaffold has been utilised to create NMEs of potential therapeutic and diagnostic utility.

7.1 Molecular modelling and simulations

Molecular modelling and simulations sought to gain a fundamental understanding of the molecular structure of hCpn10 and the molecular dynamics of the mobile loop, so as to understand and assess the potential utility to display various peptides within the mobile loop region. Accordingly studies were conducted to determine if non-native peptides could be inserted into the mobile loop without compromising monomer structure and assembly into heptamer (Chapter 3).

MD simulations of wild-type hCpn10 and its variants revealed that the β-barrel core was highly rigid and stable, whereas the mobile loop with substituted peptides exhibited large conformational change from its native configurations due to the nature of the amino-acid side chains. Removal of the mobile loop did not affect the assembly of the hCpn10 monomer into heptamer (as discussed in Chapter 3), providing further insight into the utility of this protein as a molecular scaffold.

MD simulation studies were conducted on hCpn10 containing non-native peptides within the mobile loop region. As a consequence, the assembly of the hCpn10 variant was compromised in some cases. However as mentioned previously, insertion of multiple alanine/proline linkers (P1 and P2) at the *N*- and *C*-junctures of the mobile loop stabilised the loop (of the non-native peptides) of the hCpn10 variants. Based on these simulation studies, it was apparent that such linkers constrained the inserted peptide loop into native-like conformation in order to better mimic the variants hCpn10 heptamer and facilitate the proper assembly of the recombinant designed protein (Chapter 3).

Structural predictions as a result of the MD studies were verified through expression and purification of the NMEs candidate based on hCpn10, followed by characterisation including the use of SE-HPLC to verify the native heptamer formation. The observed results were consistent with the hypothesis that these peptide linkers are important for the stabilisation of these hCpn10 variants. Thus, the success of this approach in developing a molecular scaffold based on hCpn10 represents a new framework for the development of NMEs for therapeutic and/or diagnostic purposes.

7.2 Development of an anticoagulant biologic

Blood coagulation is a process in which multiple proteolytic reactions lead to the activation of thrombin, a key protein involved in the formation of a fibrin clot. There are two pathways in the coagulation cascade, namely, intrinsic and extrinsic. Factor VII is responsible for the activation of the coagulation cascade via the extrinsic pathway and therefore serves as an attractive target for the development of biologic anticoagulants.

Studies have been conducted using phage display technology to isolate peptides that can act as serine protease inhibitors. One such peptide was previously isolated and showed preferential reactivity to FVII/FVIIa and subsequently inhibited FXa of the extrinsic TF-dependent coagulation cascade. This peptide, subsequently termed E-76, was incorporated into the mobile loop region of hCpn10 for potential therapeutic applications. The 18 residues within the native mobile loop region were substituted with the E-76 peptide between *Ala21* and *Gla38* to form a new scaffold molecule called CE76 (Chapter 4). Introducing the E-76 peptide within the loop of hCpn10 potentially results in an extended half-life compared to that of E-76 naked peptide due to a larger size, and also potentially protects the peptide from proteolysis. In addition, the introduction of this peptide into the mobile loop of hCpn10 further increased the binding avidity of this peptide as seven copies of the E-76 peptide were displayed (Chapter 4).

The anticoagulant CE76 was successfully expressed and purified from *E. coli* cells as a soluble protein and was further characterised. SE-HPLC studies indicated that a number of CE76 variants, namely, $CE76_{P1}$, $CE76_{P2}$, and $CE76_{E5}$ formed a stable heptamer with the introduction of the multiple linkers as indicated through the MD simulation studies, with the exception of $CE76_{E5}$ that had a short-term stability profile (Chapter 4).

$CE76_{P1}$ demonstrated the highest affinity towards FVIIa using biolayer interferometry (BLI). Furthermore, the *in vitro* studies indicated that the $CE76_{P1}$ protein was a potent inhibitor of FVIIa and subsequently FX activity that delayed clot formation in citrated plasma and human blood, as well as based on the measurements of enzymatic activity. The results obtained from these analyses indicate that $CE76_{P1}$ may have potential utility as a biologic-based anticoagulant (Chapter 4). $CE76_{P1}$ could be used to effectively regulate the coagulation cascade and prevent haemostasis or thrombosis, based on the specificity of $CE76_{P1}$ towards the extrinsic pathway. Moreover, $CE76_{P1}$ could be used for patients experiencing post-

surgical thrombosis and reduce the incidence of haemostasis which are typically observed with the use of standard anticoagulants such as warfarin or heparin.

Further characterisation of $CE76_{P1}$ is required, including investigating the mode of action *in vivo* of $CE76_{P1}$ inhibiting FVIIa activity. Furthermore, the half-life of $CE76_{P1}$ should be examined *in vivo* in relation to the E-76 peptide, which is essential for any future study. Most importantly, conducting *in vivo* studies, for example using rabbit models of venous thrombosis, can provide further insight on the utility of $CE76_{P1}$ as a potential biologic anticoagulant.

7.3 Development of a cancer diagnostic probe

Biomarkers are utilised to diagnose and assess disease indications. Implicit in investigating biomarkers is the need to develop diagnostic reagents that have specificity for the target biomarker. CD44 is a cell surface glycoprotein that is found to be overexpressed in a number of cancers such as breast, prostate, colon and ovarian, as well as CSCs in a number of solid tumours, therefore, is an ideal biomarker for diagnosis or targeted treatment.

In this facet of the research, hCpn10 was utilised as a scaffold whereby the mobile loop was replaced with the anti-CD44 P7 peptide, creating an NME designated CP7 (Chapter 5). Experimental binding studies by BLI confirmed that the CP7 bound to rCD44. Furthermore, CP7 bound to MBA-MB-468 breast cancer cells which overexpress CD44, with higher fluorescence intensity compared to that generated when using a mAb anti-CD44 antibody, as observed through fluorescence microscopy imaging and flow cytometry. Also, labelled 488-CP7 bound to MBA-MB-468 in a comparable manner. Thus, *in vitro* studies indicate that CP7 could be utilised as a probe for cancer diagnosis (Chapter 5).

Furthermore, CP7 could be conjugated with radiolabels such as ^{99m}Tc-, ^{64}Cu-, ^{68}Ga- and used as a probe for *in vivo* cancer detections and to monitor patient's response to therapy. The design of NMEs with diagnostic potential based on hCpn10 molecular scaffold may translate to more advanced detection methods, better prognosis, and improved therapeutic outcomes for various cancers.

7.4 Future directions and applications for scaffold based-hCpn10

hCpn10, as a molecular scaffold, meets the important requirements in biologics development in terms of size, molecular and thermal stability, production scalability, and has advantages as previously described. The viability of using hCpn10 as a molecular scaffold has been demonstrated by substituting peptides E-76 (binds Factor VIIa), P7 (binds CD44), P11 (binds FGFR) and R2 (binds VEGFR2) for the mobile loop region of hCpn10, while maintaining native heptameric structure. The success of displaying the E-76 peptide within the mobile loop of hCpn10 for use as an anticoagulant and the display of the P7 peptide as a diagnostic probe is indicative that hCpn10 can serve as peptide display platform for the development of various biologics and diagnostic agents with a higher binding avidity towards different biological targets, compared to naked peptides.

In other applications, hCpn10 could be utilised in vaccine development whereby an antigen of choice can be incorporated into the scaffold in order to induce an effective immune response. Other applications could involve utilising hCpn10 as a multivalent biologic against a variety of biological targets; for example, cross-linking cell surface receptors, or inducing a cytotoxic T-cell response against a malignant cell. hCpn10 can be used to develop a platform technology for the discovery of new biologic binders using phage display technology. In essence, a library of hCpn10 binders could be developed by introducing sequence diversity within the mobile loop and displaying this molecule on the surface of filamentous phage. For managing autoimmune diseases, hCpn10 can be used as a decoy molecule by displaying peptide fragments that are target of auto-antibodies. Regardless of its application, hCpn10 is easily scalable for production as a soluble protein in prokaryotic production systems, which generally have favourable process economics compared to mammalian cell production systems.

An open question is the potential immunogenicity of hCpn10 variants with potentially antigenic peptides within the mobile loop. hCpn10 has been shown to be non-immunogenic, supported by data that was acquired from Phase I and II clinical trials, where hCpn10 was tested for therapeutic efficacy against rheumatoid arthritis and multiple sclerosis (Chapter 1). However substituting non-endogenous peptides would likely change the immunogenicity profile, and each NME based on hCpn10 would need to be evaluated for immunogenicity through clinical trials.

List of References

Adiche, Y., Malashock, D., Pinkerton, A. & Pons, J. 2008. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical Biochemistry,* 377**,** 209-217.

AbsoluteAntibody. 2014. *Antibody Structure* [Online]. Available: http://absoluteantibody.com/antibody-resources/antibody-overview/antibody-structure/ [Accessed 18-08-2014].

Aguilar, X., Weise, C. F., Sparrman, T., Wolf-Watz, M. & Wittung-Stafshede, P. 2011. Macromolecular Crowding Extended to a Heptameric System: The Co-chaperonin Protein 10. *Biochemistry,* 50**,** 3034-3044.

Akisik, E., Bavbek, S. & Dalay, N. 2002. CD44 variant exons in leukemia and lymphoma. *Pathology and Oncology Research,* 8**,** 36-40.

Alwan, A. 2011. *Burden: mortality, morbidity and risk factors*, Geneva, World Health Organization.

Andersen, L. M., Andreasen, P. A., Svendsen, I., Keemink, J., Ostergaard, H. & Persson, E. 2012. Antibody-induced Enhancement of Factor VIIa Activity through Distinct Allosteric Pathways. *Journal of Biological Chemistry,* 287**,** 8994-9001.

Asano, M., Yukita, A., Matsumoto, T., Kondo, S. & Suzuki, H. 1995. Inhibition of tumor-growth and metastasis by an immunoneutralizing monoclonal-antibody to human vascular endothelial growth-factor vascular-permeability factor (121). *Cancer Research,* 55**,** 5296-5301.

Atluri, P. & Woo, Y. J. 2008. Pro-angiogenic cytokines as cardiovascular therapeutics - Assessing the potential. *Biodrugs,* 22**,** 209-222.

Attucci, S., Gauthier, A., Korkmaz, B., Delepine, P., Ferrer-Di Martino, M., Saudubray, F., Diot, P. & Gauthier, F. 2006. EPI-hNE4, a proteolysis-resistant inhibitor of human neutrophil elastase and potential anti-inflammatory drug for treating cystic fibrosis. *Journal of Pharmacology and Experimental Therapeutics,* 318**,** 803-809.

Balic, M., Williams, A., Lin, H., Datar, R. & Cote, R. J. 2013. Circulating Tumor Cells: From Bench to Bedside. *Annual Review of Medicine, Vol 64,* 64**,** 31-44.

Bandeiras, T. M., Hillig, R. C., Matias, P. M., Eberspaecher, U., Fanghanel, J., Thomaz, M., Miranda, S., Crusius, K., Puetter, V., Amstutz, P., Gulotti-Georgieva, M., Binz, H. K., Holz, C., Schmitz, A. A. P., Lang, C., Donner, P., Egner, U., Carrondo, M. A. & Muller-Tiemann, B. 2008. Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin. *Acta Crystallographica Section D-Biological Crystallography,* 64**,** 339-353.

Battaglia, T. A., Burhansstipanov, L., Murrell, S. S., Dwyer, A. J., & Caron, S. E. 2011. Assessing the Impact of Patient Navigation Prevention and Early Detection Metrics. *Cancer,* 117**,** 3553-3564.

Baumann, K. H., Du Bois, A., Meier, W., Rau, J., Wimberger, P., Sehouli, J., Kurzeder, C., Hilpert, F., Hasenburg, A., Canzler, U., Hanker, L. C., Hillemanns, P., Richter, B., Wollschlaeger, K., Dewitz, T., Bauerschlag, D. & Wagner, U. 2012. A phase II trial (AGO 2.11) in platinum-resistant ovarian cancer: a randomized multicenter trial with

sunitinib (SU11248) to evaluate dosage, schedule, tolerability, toxicity and effectiveness of a multitargeted receptor tyrosine kinase inhibitor monotherapy. *Annals of Oncology,* 23**,** 2265-2271.

Beck, A., Wurch, T. & Corvaia, N. 2008. Therapeutic Antibodies and Derivatives: From the Bench to the Clinic. *Current Pharmaceutical Biotechnology,* 9**,** 421-422.

Bendall, L. J., Bradstock, K. F. & Gottlieb, D. J. 2000. Expression of CD44 variant exons in acute myeloid leukemia is more common and more complex than that observed in normal blood, bone marrow or CD34(+) cells. *Leukemia,* 14**,** 1239-1246.

Bergum, P. W., Cruikshank, A., Maki, S. L., Kelly, C. R., Ruf, W. & Vlasuk, G. P. 2001. Role of zymogen and activated factor X as scaffolds for the inhibition of the blood coagulation factor VIIa-tissue factor complex by recombinant nematode anticoagulant protein c2. *Journal of Biological Chemistry,* 276**,** 10063-10071.

Bernstein, J. A. & Moellman, J. J. 2012. Progress in the Emergency Management of Hereditary Angioedema: Focus on New Treatment Options in the United States. *Postgraduate Medicine,* 124**,** 91-100.

Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A. 1999. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 1898-1903.

Bhat, R., Wedemeyer, W. J. & Scheraga, H. A. 2003. Proline isomerization in bovine pancreatic ribonuclease A. 2. Folding conditions. *Biochemistry,* 42**,** 5722-5728.

Biemond, B. J., Levi, M., Tencate, H., Soule, H. R., Morris, L. D., Foster, D. L., Bogowitz, C. A., Vanderpoll, T., Buller, H. R. & Tencate, J. W. 1995. Complete inhibition of edotoxin-induced coagulation activation in chimpanzees with a monoclonal Fab fragment against factor VII/VIIa *Thrombosis and Haemostasis,* 73**,** 223-230.

Binz, H. K., Amstutz, P., KOHL, A., Stumpp, M. T., Briand, C., Forrer, P., Grutter, M. G. & Pluckthun, A. 2004. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology,* 22**,** 575-582.

Binz, H. K., Amstutz, P. & Pluckthun, A. 2005. Engineering novel binding proteins from nonimmunoglobulin domains. *Nature Biotechnology,* 23**,** 1257-1268.

Birch, J. R. & Onakunle, Y. 2005. Biopharmaceutical proteins - Opportunities and challenges. *In:* SMALES, C. M. J. D. C. (ed.) *Methods in Molecular Biology.*

Bloom, L. & Calabro, V. 2009. FN3: a new protein scaffold reaches the clinic. *Drug Discovery Today,* 14**,** 949-955.

Bohndiek, S. E. & BRINDLE, K. M. 2010. Imaging and 'omic' methods for the molecular diagnosis of cancer. *Expert Review of Molecular Diagnostics,* 10**,** 417-434.

bonvillian, B. 2011. New Frontiers: Opportunities & Challenges For Cooperation. *Building the 21st Century: U.S.-China Cooperation on Science, Technology, and Innovation: Summary of a Symposium.* Washington (DC): National Research Council (US) Committee on Comparative National Innovation Policies: Best Practice for the 21st Century.

Brack, S., Attinger-Toller, I., Schade, B., Mourlane, F., Klupsch, K., Woods, R., Hachemi, H., Von Der Bey, U., Koenig-Friedrich, S., Bertschinger, J. & Grabulovski, D. 2014. A Bispecific HER2-Targeting FynomAb with Superior Antitumor Activity and Novel Mode of Action. *Molecular Cancer Therapeutics,* 13**,** 2030-2039.

Bridoux, A., Mousa, S. A. & Samama, M. M. 2012. Pro- and anti-angiogenic agents. *Journal Des Maladies Vasculaires,* 37**,** 132-139.

Broadley, S. A., Vanags, D., Williams, B., Johnson, B., Feeney, D., Griffiths, L., Shakib, S., Brown, G., Coulthard, A., Mullins, P. & Kneebone, C. 2009. Results of a phase IIa clinical trial of an anti-inflammatory molecule, chaperonin 10, in multiple sclerosis. *Multiple Sclerosis,* 15**,** 329-336.

Buehler, P. W., Zhou, Y. P., Cabrales, P., Jia, Y. P., Sun, G. Y., Harris, D. R., Tsai, A. G., Intaglietta, M. & Palmer, A. F. 2010. Synthesis, biophysical properties and pharmacokinetics of ultrahigh molecular weight tense and relaxed state polymerized bovine hemoglobins. *Biomaterials,* 31**,** 3723-3735.

Burgess, R. R. 2009. Refolding solubilized inclusion body proteins. *In:* BURGESS, R. R. & DEUTSCHER, M. P. (eds.) *Guide to Protein Purification, Second Edition.* San Diego: Elsevier Academic Press Inc.

Burry, R. W. 2010. *Immunocytochemistry: A Practical Guide for Biomedical Research,* New York, Springer.

Butler, G. S. & Overall, C. M. 2009. OPINION Proteomic identification of multitasking proteins in unexpected locations complicates drug targeting. *Nature Reviews Drug Discovery,* 8**,** 935-948.

Campochiaro, P. A., Channa, R., Berger, B. B., Heier, J. S., Brown, D. M., Fiedler, U., Hepp, J. & Stumpp, M. T. 2013. Treatment of Diabetic Macular Edema With a Designed Ankyrin Repeat Protein That Binds Vascular Endothelial Growth Factor: A Phase I/II Study. *American Journal of Ophthalmology,* 155**,** 697-704.

Cappello, F., Bellafiore, M., David, S., Anzalone, R. & Zummo, G. 2003. Ten kilodalton heat shock protein (HSP10) is overexpressed during carcinogenesis of large bowel and uterine exocervix. *Cancer Letters,* 196**,** 35-41.

Carter, P. 2001. Improving the efficacy of antibody-based cancer therapies. *Nature Reviews Cancer,* 1**,** 118-129.

Carter, P. J. & Senter, P. D. 2008. Antibody-drug conjugates for cancer therapy. *Cancer Journal,* 14**,** 154-169.

Cavanagh, A. C. 1996. Identification of early pregnancy factor as chaperonin 10: implications for understanding its role. *Reviews of reproduction,* 1**,** 28-32.

CDER. 2014. *CDER Therapeutic Biologic Products* [Online]. USA: FDA. Available: http://www.fda.gov/downloads/forindustry/userfees/prescriptiondruguserfee/ucm164641.pdf [Accessed 24-04-2014].

Chen, J. H., Brooks, C. L. & Khandogin, J. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations. *Current Opinion in Structural Biology,* 18**,** 140-148.

Cohen, M. H., Gootenberg, J., Keegan, P. & Pazdur, R. 2007. FDA drug approval summary: Bevacizumab (Avastin (R)) plus carboplatin and paclitaxel as first-line treatment of advanced/metastatic recurrent nonsquamous non-small cell lung cancer. *Oncologist,* 12**,** 713-718.

Colas, P., Cohen, B., Jessen, T., Grishina, I., Mccoy, J. & Brent, R. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature,* 380**,** 548-550.

Coutts, J. C. & Gallagher, J. T. 1995. Receptors for fibroblast growth factors. *Immunology and Cell Biology,* 73**,** 584-589.

De Abreu, F. B., Wells, W. A. & Tsongalis, G. J. 2013. The Emerging Role of the Molecular Diagnostics Laboratory in Breast Cancer Personalized Medicine. *American Journal of Pathology,* 183**,** 1075-1083.

Dempke, W. C. M. & Zippel, R. 2010. Brivanib, A Novel Dual VEGF-R2/bFGF-R Inhibitor. *Anticancer Research,* 30**,** 4477-4483.

Dennis, M. S., Eigenbrot, C., Skelton, N. J., Ultsch, M. H., Santell, L., Dwyer, M. A., O'connell, M. P. & Lazarus, R. A. 2000. Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature,* 404**,** 465-470.

Dennis, M. S. & Lazarus, R. A. 1994a. Kunitz Domain Inhibitors of Tissue Factor-Factor VIIa .1. Potent and Inhibitors Selected from Libraries by Phage Display. *Journal of Biological Chemistry,* 269**,** 22129-22136.

Dennis, M. S. & Lazarus, R. A. 1994b. Kunitz Domain Inhibitors of Tissue Factor-Factor VIIa .2. Potent and Specific Inhibitors by Competitive Phage Selection. *Journal of Biological Chemistry,* 269**,** 22137-22144.

Dennis, M. S., Roberge, M., Quan, C. & Lazarus, R. A. 2001. Selection and characterization of a new class of peptide exosite inhibitors of coagulation factor VIIa. *Biochemistry,* 40**,** 9513-9521.

Devy, L., Rabbani, S. A., Stochl, M., Ruskowski, M., Mackie, I., Naa, L., Toews, M., Van Gool, R., Chen, J., Leyz, A., Ladner, R. C., Dransfield, D. T. & Henderikx, P. 2007. PEGylated DX-1000: Pharmacokinetics and antineoplastic activity of a specific plasmin inhibitor. *Neoplasia,* 9**,** 927-937.

Dhir, R., Gau, J. T., Krill, D., Bastacky, S., Bahnson, R. R., Cooper, D. L. & Becich, M. J. 1997. CD44 expression in benign and neoplastic human prostates. *Molecular Diagnosis,* 2**,** 197-204.

Diao, L. & Meibohm, B. 2013. Pharmacokinetics and Pharmacokinetic-Pharmacodynamic Correlations of Therapeutic Peptides. *Clinical Pharmacokinetics,* 52**,** 855-868.

Dineen, S. P., Sullivan, L. A., Beck, A. W., Miller, A. F., Carbon, J. G., Mamluk, R., Wong, H. & Brekken, R. A. 2008. The Adnectin CT-322 is a novel VEGF receptor 2 inhibitor that decreases tumor burden in an orthotopic mouse model of pancreatic cancer. *Bmc Cancer,* 8.

Donahue, S. M. & Otto, C. M. 2005. Thromboelastography: a tool for measuring hypercoagulability, hypocoagulability, and fibrinolysis. *Journal of Veterinary Emergency and Critical Care,* 15**,** 9-16.

Ds, D. 2012. Therapeutic Proteins. *In:* Vladimir Voynov & Caravella, J. A. (eds.) *Therapeutic Proteins Methods and Protocols.* 2 ed. USA: Humana Press.

Duan, J., Wu, J., Valencia, C. A. & Liu, R. 2007. Fibronectin type III domain based monobody with high avidity. *Biochemistry,* 46**,** 12656-12664.

Duan, L., Marvdashti, T., Lee, A., Tang, J. Y. & Ellerbee, A. K. 2014. Automated identification of basal cell carcinoma by polarization-sensitive optical coherence tomography. *Biomedical Optics Express,* 5**,** 3717-3729.

Eibl, R. H., Pietsch, T., Moll, J., Skrochangel, P., Heider, K. H., Vonammon, K., Wiestler, O. D., Ponta, H., Kleihues, P. & Herrlich, P. 1995. Expression of variant CD44 epitopes in human astrocytic brain tumors. *Journal of Neuro-Oncology,* 26**,** 165-170.

Enjyoji K, Miyata T, Kamikubo Y & H., K. 1995. Effect of heparin on the inhibition of factor Xa by tissue factor pathway inhibitor: a segment, Gly212-Phe243, of the third Kunitz domain is a heparin-binding site. *Biochemistry***,** 5725-35.

Expasy. 2014a. *DNA/Protein Translate Tool* [Online]. Swiss Institute of Bioinformatics. Available: http://web.expasy.org/translate/ [Accessed 01-09-2014].

Expasy. 2014b. *Proteomics, ProtParam* [Online]. Swiss Institute of Bioinformatics. Available: http://www.isb-sib.ch/ [Accessed 01-09-2014].

Eyles, S. J. & Gierasch, L. M. 2000. Multiple roles of prolyl residues in structure and folding. *Journal of Molecular Biology,* 301**,** 737-747.

Ferrara, N., Hillan, K. J. & Novotny, W. 2005. Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy. *Biochemical and Biophysical Research Communications,* 333**,** 328-335.

Ferrer, M., Hamilton, A. C. & Inglese, J. 2002. A PDZ domain-based detection system for enzymatic assays. *Analytical Biochemistry,* 301**,** 207-216.

Ferrer, M., Maiolo, J., Kratz, P., Jackowski, J. L., Murphy, D. J., Delagrave, S. & Inglese, J. 2005. Directed evolution of PDZ variants to generate high-affinity detection reagents. *Protein Engineering Design & Selection,* 18**,** 165-173.

Fischer, T. H., Nichols, T. C., Scull, C. M., Smith, C. J. & Demcheva, M. 2011. Poly-N-Acetylglucosamine Fibers Amplify the Effectiveness of Recombinant Factor VIIA on Clot Formation in Hemophilia B Canine Blood. *Journal of Trauma-Injury Infection and Critical Care,* 71**,** S171-S175.

Fogolari, F., Brigo, A. & Molinari, H. 2002. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition,* 15**,** 377-392.

Forrer, P., Stumpp, M. T., Binz, H. K. & Pluckthun, A. 2003. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *Febs Letters,* 539**,** 2-6.

Frenkel, D. & SMIT, B. 2002. *Understanding molecular simulation: from algorithms to applications,* Academic Press.

Friedman, M., Orlova, A., Johansson, E., Eriksson, T. L. J., Hoiden-Guthenberg, I., TOLMACHEV, V., NILSSON, F. Y. & STAHL, S. 2008. Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. *Journal of Molecular Biology,* 376**,** 1388-1402.

Gacche, R. N. & Meshram, R. J. 2014. Angiogenic factors as potential drug target: Efficacy and limitations of anti-angiogenic therapy. *Biochimica Et Biophysica Acta-Reviews on Cancer,* 1846**,** 161-179.

Galvez, K. & Cortes, C. 2012. Tromboelastografía: nuevos conceptos en la fisiología de la hemostasia y su correlación con la coagulopatía asociada al trauma correlation with trauma associated coagulopathy. *Revista Colombiana de Anestesiología,* 40**,** 224-230.

Gebauer, M. & Skerra, A. 2009. Engineered protein scaffolds as next-generation antibody therapeutics. *Current Opinion in Chemical Biology,* 13**,** 245-255.

Getmanova, E. V., Chen, Y., Bloom, L., Gokemeijer, J., Shamah, S., Warikoo, V., Wang, J., Ling, V. & Sun, L. 2006. Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro. *Chemistry & Biology,* 13**,** 549-556.

Gill, D. S. & Damle, N. K. 2006. Biopharmaceutical drug discovery using novel protein scaffolds. *Current Opinion in Biotechnology,* 17**,** 653-658.

Giri, D., Ropiquet, F. & Ittmann, M. 1999. Alterations in expression of basic fibroblast growth factor (FGF) 2 and its receptor FGFR-1 in human prostate cancer. *Clinical Cancer Research,* 5**,** 1063-1071.

Gray, E., Hogwood, J. & Mulloy, B. 2012. The anticoagulant and antithrombotic mechanisms of heparin. *Handbook of experimental pharmacology***,** 43-61.

Greenfield, N. J. 2006. Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nature Protocols,* 1**,** 2891-2899.

Greenwald, D. R., LI, H. L., Luger, S. M., Go, R. S., King, D., Patel, T., Gascoyne, R. D., Kolesar, J., Kahl, B. S. & Horning, S. 2013. A phase II study of sorafenib (BAY 43-9006) in recurrent diffuse large B cell lymphoma: an eastern cooperative oncology group study (E1404). *Journal of Hematology & Oncology,* 6.

Gronwall, C., Jonsson, A., Lindstrom, S., Gunneriusson, E., Stahl, S. & Herne, N. 2007. Selection and characterization of affibody ligands binding to Alzheimer amyloid beta peptides. *Journal of Biotechnology,* 128**,** 162-183.

Gronwall, C., Snelders, E., Palm, A. J., Eriksson, F., Herne, N. & Stahl, S. 2008. Generation of Affibody (R) ligands binding interieukin-2 receptor alpha/CD25. *Biotechnology and Applied Biochemistry,* 50**,** 97-112.

Grzyb, J., Xu, F., Weiner, L., Reijerse, E. J., Lubitz, W., Nanda, V. & Noy, D. 2010. De novo design of a non-natural fold for an iron-sulfur protein: Alpha-helical coiled-coil with a four-iron four-sulfur cluster binding site in its central core. *Biochimica Et Biophysica Acta-Bioenergetics,* 1797**,** 406-413.

Guidry, J. J., Moczygemba, C. K., Steede, N. K., Landry, S. J. & Wittung-Stafshede, P. 2000. Reversible denaturation of oligomeric human chaperonin 10: Denatured state depends on chemical denaturant. *Protein Science,* 9**,** 2109-2117.

Guidry, J. J., Shewmaker, F., Maskos, K., Landry, S. & Wittung-Stafshede, P. 2003. Probing the interface in a human co-chaperonin heptamer: residues disrupting oligomeric unfolded state identified. *BMC biochemistry,* 4**,** 14.

Guvench, O. & Alexander D. Mackerell, J. 2008. Comparison of Protein Force Fields for Molecular Dynamics Simulations. *In:* Kukol, A. (ed.) *Molecular Modeling of Proteins.* Totowa, NJ: Humana Press.

Hackel, B. J., Kapila, A. & Wittrup, K. D. 2008. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of Molecular Biology,* 381**,** 1238-1252.

Hartl, F. U. & Hayer-Hartl, M. 2002. Protein folding - Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science,* 295**,** 1852-1858.

Hartman, D. J., Hoogenraad, N. J., Condron, R. & Hoj, P. B. 1992. Identification of a Mammalian 10-kDa Heat-shock Protein, a Mitochondarial Chaperonin-10 Homology Essential for Assisted Folding of Trimeric Ornithine Transcarbamolyase *in vitro*

Proceedings of the National Academy of Sciences of the United States of America, 89**,** 3394-3398.

Hassett, M. J. & Elkin, E. B. 2013. What Does Breast Cancer Treatment Cost and What Is It Worth? *Hematology-Oncology Clinics of North America,* 27**,** 829-+.

Hicks, D. G. & Tubbs, R. R. 2005. Assessment of the HER2 status in breast cancer by fluorescence in situ hybridization: a technical review with interpretive guidelines. *Human Pathology,* 36**,** 250-261.

Higurashi, T., Hiragi, Y., Ichimura, K., Seki, Y., Soda, K., Mizobata, T. & Kawata, Y. 2003. Structural stability and solution structure of chaperonin GroES heptamer studied by synchrotron small-angle X-ray scattering. *Journal of Molecular Biology,* 333**,** 605-620.

Hilpert, K., Hansen, G., Wessner, H., Kuttner, G., Welfle, K., Seifert, M. & Hohne, W. 2001. Anti-c-myc antibody 9E10: epitope key positions and variability characterized using peptide spot synthesis on cellulose. *Protein Engineering,* 14**,** 803-806.

Ho, B. K., Coutsias, E. A., Seok, C. & Dill, K. A. 2005. The flexibility in the proline ring couples to the protein backbone. *Protein Science,* 14**,** 1011-1018.

Holliger, P. & Hudson, P. J. 2005. Engineered antibody fragments and the rise of single domains. *Nature Biotechnology,* 23**,** 1126-1136.

Hong, S. K. 2014. Kallikreins as biomarkers for prostate cancer. *BioMed research international,* 2014**,** 526341.

Hori, A., Sasada, R., Matsutani, E., Naito, K., Sakura, Y., Fujita, T. & Kozai, Y. 1991. Suppression of solid tumor-growth by immunoneutralizing monoclonal-antibody against human basic fibroblast growth-factor *Cancer Research,* 51**,** 6180-6184.

Hortobagyi, G. N., De La Garza Salazar, J., Pritchard, K., Amadori, D., Haidinger, R., Hudis, C. A., Khaled, H., Liu, M.-C., Martin, M., Namer, M., O'shaughnessy, J. A., Shen, Z. Z., Albain, K. S. & Investigators, A. 2005. The global breast cancer burden: variations in epidemiology and survival. *Clinical breast cancer,* 6**,** 391-401.

Huang, J., Koide, A., Nettle, K. W., Greene, G. L. & Koide, S. 2006. Conformation-specific affinity purification of proteins using engineered binding proteins: Application to the estrogen receptor. *Protein Expression and Purification,* 47**,** 348-354.

Hummers, L. K., Hall, A., Wigley, F. M. & Simons, M. 2009. Abnormalities in the Regulators of Angiogenesis in Patients with Scleroderma. *Journal of Rheumatology,* 36**,** 576-582.

Humphrey, W., Dalke, A. & Schulten, K. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling,* 14**,** 33-38.

Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. & Deisenhofer, J. 1996. The crystal structure of the GroES co-chaperonin at 2.8 angstrom resolution. *Nature,* 379**,** 37-45.

Hurt, E. M., Kawasaki, B. T., Klarmann, G. J., Thomas, S. B. & Farrar, W. L. 2008. CD44(+) CD24(-) prostate cells are early cancer progenitor/stem cells that provide a model for patients with poor prognosis. *British Journal of Cancer,* 98**,** 756-765.

Ishimoto, T., Oshima, H., Oshima, M., Kai, K., Torii, R., Masuko, T., Baba, H., Saya, H. & Nagano, O. 2010. CD44(+) slow-cycling tumor cell expansion is triggered by cooperative actions of Wnt and prostaglandin E-2 in gastric tumorigenesis. *Cancer Science,* 101**,** 673-678.

Jaggupilli, A. & Elkord, E. 2012. Significance of CD44 and CD24 as Cancer Stem Cell Markers: An Enduring Ambiguity. *Clinical & Developmental Immunology*.

Janeway, C. A., Travers, P., Walport, M., & Shlomchik, M. J. 2001. *Immunobiology: The Immune System in Health and Disease*, New York, Garland Science.

Jayamohan, H., Sant, H. J. & Gale, B. K. 2013. Applications of Microfluidics for Molecular Diagnostics. *In:* Jenkins, G. & Mansfield, C. D. (eds.) *Microfluidic Diagnostic Methods and Protocols.* Hatfield, UK: Humana Press, Springer

Jemal, A., Ward, E. & Thun, M. 2010. Declining Death Rates Reflect Progress against Cancer. *Plos One,* 5.

Jonsson, A., Wallberg, H., Herne, N., Stahl, S. & Frejd, F. Y. 2009. Generation of tumour-necrosis-factor-alpha-specific affibody molecules capable of blocking receptor binding in vitro. *Biotechnology and Applied Biochemistry,* 54**,** 93-103.

Kaaijk, P., Troost, D., Morsink, F., Keehnen, R. M. J., Leenstra, S., Bosch, D. A. & Pals, S. T. 1995. Expression of CD44 splice variants in human primary brain tumors. *Journal of Neuro-Oncology,* 26**,** 185-190.

Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. & Schulten, K. 1999. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics,* 151**,** 283-312.

Kaluna-Czaplinska, J. & Jozwik, J. 2014. Current applications of chromatographic methods for diagnosis and identification of potential biomarkers in cancer. *Trac-Trends in Analytical Chemistry,* 56**,** 1-12.

Kamal, A. H., Tefferi, A. & Pruthi, R. K. 2007. How to interpret and pursue an abnormal prothrombin time, activated partial thromboplastin time, and bleeding time in adults. *Mayo Clinic Proceedings,* 82**,** 864-873.

Karatan, E., Merguerian, M., Han, Z. H., Scholle, M. D., Koide, S. & Kay, B. K. 2004. Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chemistry & Biology,* 11**,** 835-844.

Karlstrom, A. & Nygren, P. A. 2001. Dual labeling of a binding protein allows for specific fluorescence detection of native protein. *Analytical Biochemistry,* 295**,** 22-30.

Kawe, M., FORRER, P., AMSTUTZ, P. & PLUCKTHUN, A. 2006. Isolation of intracellular proteinase inhibitors derived from designed ankyrin repeat proteins by genetic screening. *Journal of Biological Chemistry,* 281**,** 40252-40263.

Kelley, R. F., Refino, C. J., Oconnell, M. P., Modi, N., Sehl, P., Lowe, D., Pater, C. & Bunting, S. 1997. A soluble tissue factor mutant is a selective anticoagulant and antithrombotic agent. *Blood,* 89**,** 3219-3227.

Khoursheed, M., Mathew, T. C., Makar, R. R., Sonia, L., Abul, H., Asfar, S., Al-Sayer, H., Dashti, H. M. & Al-Bader, A. 2002. Expression of CD44s in human colorectal cancer. *Pathology & Oncology Research,* 8**,** 170-174.

Kim, H. J., Eichinger, A. & Skerra, A. 2009. High-Affinity Recognition of Lanthanide(III) Chelate Complexes by a Reprogrammed Human Lipocalin 2. *Journal of the American Chemical Society,* 131**,** 3565-3576.

Klement, P. & Rak, J. 2006. Emerging anticoagulants: mechanism of action and future potential. *Vnitr Lek,* 52 Suppl 1**,** 119-22.

Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. 2002. Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 1253-1258.

Koide, A., Bailey, C. W., Huang, X. L. & Koide, S. 1998. The fibronectin type III domain as a scaffold for novel binding proteins. *Journal of Molecular Biology,* 284**,** 1141-1151.

Koide, A. & Koide, S. 2007. Monobodies - Antibody mimics based on the scaffold of the fibronectin type III domain. *Protein Engineering Protocols,* 352**,** 95-109.

Kontermann, R. E. 2010. Alternative antibody formats. *Current Opinion in Molecular Therapeutics,* 12**,** 176-183.

Korndorfer, I. P., Schlehuber, S. & Skerra, A. 2003. Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin. *Journal of Molecular Biology,* 330**,** 385-396.

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science,* 302**,** 1364-1368.

Lagadec, C., Vlashi, E., Della Donna, L., Meng, Y., Dekmezian, C., Kim, K. & Pajonk, F. 2010. Survival and self-renewing capacity of breast cancer initiating cells during fractionated radiation treatment. *Breast Cancer Research,* 12.

Landry, S. J., Steede, N. K. & Maskos, K. 1997. Temperature dependence of backbone dynamics in loops of human mitochondrial heat shock protein 10. *Biochemistry,* 36**,** 10975-10986.

Lazar, G. A., Desjarlais, J. R. & Handel, T. M. 1997. De novo design of the hydrophobic core of ubiquitin. *Protein Science,* 6**,** 1167-1178.

Leach, A. R. 2001. *Molecular modelling: principles and applications,* Harlow, England, Prentice Hall.

Lefkowitz, J. B. 2013. Coagulation Pathway and Physiology. *Hemostasis Physiology*

Lehmann, A. 2008. Ecallantide (DX-88), a plasma kallikrein inhibitor for the treatment of hereditary angioedema and the prevention of blood loss in on-pump cardiothoracic surgery. *Expert Opinion on Biological Therapy,* 8**,** 1187-1199.

Leung, D., Abbenante, G. & Fairlie, D. P. 2000. Protease inhibitors: Current status and future prospects. *Journal of Medicinal Chemistry,* 43**,** 305-341.

Li, J. & Zhu, Z. P. 2010. Research and development of next generation of antibody-based therapeutics. *Acta Pharmacologica Sinica,* 31**,** 1198-1207.

Li, Q. C., Gao, S. S., Yu, Y. L., Wang, W. H., Chen, X. L., Wang, R. X., LI, T., Wang, C., Li, X. K. & Wu, X. P. 2012. A novel bFGF antagonist peptide inhibits breast cancer cell growth. *Molecular Medicine Reports,* 6**,** 210-214.

Li, X., Jiang, L., Wang, Y., Xiao, Y., Huang, Y., Yao, Q., Yang, Y. & Wu, X. 2007. Inhibition of angiogenesis by a novel small peptide consisting of the active fragments of platelet factor-4 and vasostatin. *Cancer Letters,* 256**,** 29-32.

Lin, K. M., Hollander, J. M., Kao, V. Y., Lin, B., Macpherson, L. & Dillmann, W. H. 2004. Myocyte protection by 10 kD heat shock protein (Hsp10) involves the mobile loop and attenuation of the Ras GTP-ase pathway. *Faseb Journal,* 18**,** 1004-+.

LINDAHL, E. R. 2008. Molecular Dynamics Simulations. *In:* KUKOL, A. (ed.) *Molecular Modeling of Proteins.* Totowa, NJ: Humana Press.

Liu, J. Q., Ning, B., Liu, M., Sun, Y. A., Sun, Z. Y., Zhang, Y. H., Fan, X. J., Zhou, Z. J. & Gao, Z. X. 2012. Construction of ribosome display library based on lipocalin scaffold and screening anticalins with specificity for estradiol. *Analyst,* 137**,** 2470-2479.

Lofblom, J., Feldwisch, J., Tolmachev, V., Carlsson, J., Stahl, S. & Frejd, F. Y. 2010. Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *Febs Letters,* 584**,** 2670-2680.

Lofblom, J., Frejd, F. Y. & Stahl, S. 2011. Non-immunoglobulin based protein scaffolds. *Current Opinion in Biotechnology,* 22**,** 843-848.

Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature,* 423**,** 185-190.

Lu, Z. J., Murray, K. S., Vancleave, V., Lavallie, E. R., Stahl, M. L. & Mccoy, J. M. 1995. Expression of Thioredoxin Random Peptide Libraries on the Escherichia-Coli cell-surface as Functional Fusions to Flagellin - A System Designed for Exploring Protein-protein Interactions. *Bio-Technology,* 13**,** 366-372.

Luke, K., Apiyo, D. & Wittung-Stafshede, P. 2005a. Dissecting homo-heptamer thermodynamics by isothermal titration calorimetry: Entropy-driven assembly of co-chaperonin protein 10. *Biophysical Journal,* 89**,** 3332-3336.

Luke, K., Apiyo, D. & Wittung-Stafshede, P. 2005b. Role of the unique peptide tail in hyperthermostable Aquifex aeolicus cochaperonin protein 10. *Biochemistry,* 44**,** 14385-14395.

Luke, K., Perham, M. & Wittung-Stafshede, P. 2006. Kinetic folding and assembly mechanisms differ for two homologous heptamers. *Journal of Molecular Biology,* 363**,** 729-742.

Lunn, M. & Banta, E. 2011. Ecallantide for the treatment of hereditary angiodema in adults. *Clinical Medicine Insights. Cardiology,* 5**,** 49-54.

Macario, A. J. L. & De Macario, E. C. 2007. Molecular chaperones: Multiple functions, pathologies, and potential applications. *Frontiers in Bioscience,* 12**,** 2588-2600.

Macdonald, S. G. & Luddington, R. J. 2010. Critical Factors Contributing to the Thromboelastography Trace. *Seminars in Thrombosis and Hemostasis,* 36**,** 712-722.

Mackerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B,* 102**,** 3586-3616.

Mamluk, R., Carvajal, I. M., Morse, B. A., Wong, H., Abramowitz, J., Aslanian, S., Lim, A. C., Gokemeijer, J., Storek, M. J., Lee, J., Gosselin, M., Wright, M. C., Camphausen, R. T., Wang, J., Chen, Y., Miller, K., Sanders, K., Short, S., Sperinde, J., Prasad, G., Williams, S., Kerbel, R., Ebos, J., Mutsaers, A., Mendlein, J. D., Harris, A. S. & Furfine, E. S. 2010. Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. *Mabs,* 2**,** 199-208.

Marks, J. D. 2009. Molecular Engineering of Antibodies. *In:* WALKER, J. M. & RAPLY, R. (eds.) *Molecular Biology and Biotechnology.* Cambridge, UK: Royal Society of Chemistry.

Martin, J., Geromanos, S., Tempst, P. & Hartl, F. U. 1993. Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES *Nature,* 366**,** 279-282.

Martinez-Madrid, B., Donnez, J., Van Eyck, A. S., Veiga-Lopez, A., Dolmans, M. M. & Van Langendonckt, A. 2009. Chick embryo chorioallantoic membrane (CAM) model: a useful tool to study short-term transplantation of cryopreserved human ovarian tissue. *Fertility and Sterility,* 91**,** 285-292.

Martini, W. Z., Cortez, D. S., Dubick, M. A., Park, M. S. & Holcomb, J. B. 2008. Thrombelastography is better than PT, aPTT, and activated clotting time in detecting clinically relevant clotting abnormalities after hypothermia, hemorrhagic shock and resuscitation in pigs. *Journal of Trauma-Injury Infection and Critical Care,* 65**,** 535-543.

Mathonet, P. & Fastrez, J. 2004. Engineering of non-natural receptors. *Current Opinion in Structural Biology,* 14**,** 505-511.

Maun, H. R., Eigenbrot, C. & Lazarus, R. A. 2003. Engineering exosite peptides for complete inhibition of factor VIIa using a protease switch with substrate phage. *Journal of Biological Chemistry,* 278**,** 21823-21830.

Mitchell, F. L., Frank, F., Marks, G. E., Suzuki, M., Douglas, K. T. & Bryce, R. A. 2009. Molecular dynamics study of chemically engineered green fluorescent protein mutants: Comparison of intramolecular fluorescence resonance energy transfer rate. *Proteins-Structure Function and Bioinformatics,* 75**,** 28-39.

Moadel, R. M. 2011. Breast Cancer Imaging Devices. *Seminars in Nuclear Medicine,* 41**,** 229-241.

Moreira, I. S., Fernandes, P. A. & Ramos, M. J. 2005. Accuracy of the numerical solution of the Poisson-Boltzmann equation. *Journal of Molecular Structure-Theochem,* 729**,** 11-18.

Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein Science,* 13**,** 1435-1448.

Nakamura, H., Takamori, S., Fujii, T., Ono, M., Yamana, H., Kuwano, M. & SHIROUZU, K. 2005. Cooperative cell-growth inhibition by combination treatment with ZD1839 (Iressa) and trastuzumab (Herceptin) in non-small-cell lung cancer. *Cancer Letters,* 230**,** 33-46.

Nandi, T. K., Bairagya, H. R., Mishra, D. K., Mukhopadhyay, B. P. & Banerjee, A. 2012. Structural and Putative Functional Role of Conserved Water Molecular Cluster in the X-ray Structures of Plant Thiol Proteases: A Molecular Dynamics Simulation Study. *Journal of Chemical Crystallography,* 42**,** 1105-1118.

Naor, D., Nedvetzki, S., Golan, I., Melnik, L. & Faitelson, Y. 2002. CD44 in cancer. *Critical Reviews in Clinical Laboratory Sciences,* 39**,** 527-579.

Nawa, M., Osada, S., Morimitsu, K., Nonaka, K., Futamura, M., Kawaguchi, Y. & Yoshida, K. 2012. Growth Effect of Neutrophil Elastase on Breast Cancer: Favorable Action of Sivelestat and Application to Anti-HER2 Therapy. *Anticancer Research,* 32**,** 13-19.

NCBI. 2014. *Basic Local Alignment Search Tool* [Online]. Align Sequences Nucleotide BLAST. Available: http://blast.ncbi.nlm.nih.gov/Blast.cgi [Accessed 01-07-2014].

Negi, L. M., Talegaonkar, S., Jaggi, M., Ahmad, F. J., Iqbal, Z. & Khar, R. K. 2012. Role of CD44 in tumour progression and strategies for targeting. *Journal of Drug Targeting,* 20**,** 561-573.

Nelson, A. L. & Reichert, J. M. 2009. Development trends for therapeutic antibody fragments. *Nature Biotechnology,* 27**,** 331-337.

Nicaise, M., Valerio-Lepiniec, M., Minard, P. & Desmadril, M. 2004. Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold. *Protein Science,* 13**,** 1882-1891.

Nilsson, F. Y. & Tolmachev, V. 2007. Affibody (R) molecules: New protein domains for molecular imaging and targeted tumor therapy. *Current Opinion in Drug Discovery & Development,* 10**,** 167-175.

Nixon, A. E. & Wood, C. R. 2006. Engineered protein inhibitors of proteases. *Current Opinion in Drug Discovery & Development.*

Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M. & Nygren, P. A. 1997. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nature Biotechnology,* 15**,** 772-777.

Nord, K., Gunneriusson, E., Uhlen, M. & Nygren, P. A. 2000. Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1(M) and Taq DNA polymerase. *Journal of Biotechnology,* 80**,** 45-54.

Nord, K., Nord, O., Uhlen, M., Kelley, B., Ljungqvist, C. & Nygren, P. A. 2001. Recombinant human factor VIII-specific affinity ligands selected from phage-displayed combinatorial libraries of protein A. *European Journal of Biochemistry,* 268**,** 4269-4277.

Numoto, N., Kita, A. & Miki, K. 2005. Crystal structure of the co-chaperonin cpn10 from Thermus thermophilus HB8. *Proteins-Structure Function and Bioinformatics,* 58**,** 498-500.

Nuttall, S. D. & Walsh, R. B. 2008. Display scaffolds: protein engineering for novel therapeutics. *Current Opinion in Pharmacology,* 8**,** 609-615.

Olsson, E., Honeth, G., Bendahl, P. O., SAAL, L. H., GRUVBERGER-SAAL, S., Ringner, M., Vallon-Christersson, J., Jonsson, G., Holm, K., Lovgren, K., Ferno, M., Grabau, D., Borg, A. & Hegardt, C. 2011. CD44 isoforms are heterogeneously expressed in breast cancer and correlate with tumor subtypes and cancer stem cell markers. *Bmc Cancer,* 11.

Onufriev, A. 2010. Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. *Annual Reports in Computational Chemistry, Vol 4.*

Orlova, A., Magnusson, M., Eriksson, T. L. J., Nilsson, M., Larsson, B., Hoiden-Guthenherg, I., Widstrom, C., Carlsson, J., Tolmachev, V., Stahl, S. & Nilsson, F. Y. 2006. Tumor Imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer Research,* 66**,** 4339-4348.

Oshikiri, T., Miyamoto, M., Morita, T., Fujita, M., Miyasaka, Y., Senmaru, N., Yamada, H., Takahashi, T., Horita, S. & Kondo, S. 2006. Tumor-associated antigen recognized by the 22-1-1 monoclonal antibody encourages colorectal cancer progression under the scanty CD8(+) T cells. *Clinical Cancer Research,* 12**,** 411-416.

Paborsky, L. R., Law, V. S., Mao, C. T., Leung, L. L. K. & Gibbs, C. S. 1995. A peptide derived from a tissue factor loop region functions as a tissue factor-factor VIIa antagonist. *Biochemistry,* 34**,** 15328-15333.

Park, H.-Y., Lee, K.-J., Lee, S.-J. & Yoon, M.-Y. 2012. Screening of Peptides Bound to Breast Cancer Stem Cell Specific Surface Marker CD44 by Phage Display. *Molecular Biotechnology,* 51**,** 212-220.

Perham, M. & Wittung-Stafshede, P. 2007. Folding and assembly of co-chaperonin heptamer probed by forster resonance energy transfer. *Archives of Biochemistry and Biophysics,* 464**,** 306-313.

Petrenko, R. & Meller, J. 2010. *Molecular Dynamics* [Online]. eLS. John Wiley & Sons Ltd, Chichester. Available: http://www.els.net [Accessed 15-01-2014].

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. 2005. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry,* 26**,** 1781-1802.

Phillips, T. M., Mcbride, W. H. & Pajonk, F. 2006. The response of CD24(-/low)/CD44(+) breast cancer-initiating cells to radiation. *Journal of the National Cancer Institute,* 98**,** 1777-1785.

Pilkington, S. J. & WALKER, J. E. 1993. Complementary DNA sequence of bovine cpn10 (Hsp10), a chaperone protein from mitochondria. *DNA sequence : the journal of DNA sequencing and mapping,* 3**,** 291-5.

Ponta, H., Sherman, L. & Herrlich, P. A. 2003. CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology,* 4**,** 33-45.

Rahman, M. A., Amin, R. M. R., Wang, D. S., Koenig, L., Nannapaneni, S., Chen, Z. J., Wang, Z. B., Sica, G., Deng, X. M., Chen, Z. & Shin, D. M. 2013. RRM2 Regulates Bcl-2 in Head and Neck and Lung Cancers: A Potential Target for Cancer Therapy. *Clinical Cancer Research,* 19**,** 3416-3428.

Ranford, J. C., Coates, A. R. & Henderson, B. 2000. Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones. *Expert reviews in molecular medicine,* 2**,** 1-17.

Ranson, N. A., White, H. E. & Saibil, H. R. 1998. Chaperonins. *Biochemical Journal,* 333**,** 233-242.

RCSB-PDB. 2010. *Protein Data Bank* [Online]. RCSB-PDB. Available: http://www.rcsb.org/pdb/home/home.do [Accessed 14-06-2014].

Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H. & Pedersen, J. T. 1994. Antibody design - Beyond the natural limits. *Trends in Biotechnology,* 12**,** 199-206.

Renberg, B., Nordin, J., Merca, A., Uhlen, M., Feldwisch, J., Nygren, P. A. & Karlstrom, A. E. 2007. Affibody molecules in protein capture microarrays: Evaluation of multidomain ligands and different detection formats. *Journal of Proteome Research,* 6**,** 171-179.

Ricci-Vitiani, L., Lombardi, D. G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C. & De Maria, R. 2007. Identification and expansion of human colon-cancer-initiating cells. *Nature,* 445**,** 111-115.

Richardson, A., Schwager, F., Landry, S. J. & Georgopoulos, C. 2001. The importance of a mobile loop in regulating chaperonin/co-chaperonin interaction - Humans versus Escherichia coli. *Journal of Biological Chemistry,* 276**,** 4981-4987.

Roberts, M. M., Coker, A. R., Fossati, C., Mascagni, P., Coates, A. R. M. & Wood, S. P. 1999. Crystallization, X-ray diffraction and preliminary structure analysis of Mycobacterium tuberculosis chaperonin 10. *Acta Crystallographica Section D-Biological Crystallography,* 55**,** 910-914.

Ronnmark, J., Gronlund, H., Uhlen, M. & Nygren, P. A. 2002. Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A. *European Journal of Biochemistry,* 269**,** 2647-2655.

Ryu, J. H., Shin, M., Kim, S. A., Lee, S., Kim, H., Koo, H., Kim, B. S., Song, H. K., Kim, S. H., Choi, K., Kwon, I. C., Jeon, H. & Kim, K. 2013. In vivo fluorescence imaging for cancer diagnosis using receptor-targeted epidermal growth factor-based nanoprobe. *Biomaterials,* 34**,** 9149-9159.

Sakane, I., Hongo, K., Motojima, F., Murayama, S., Mizobata, T. & Kawata, Y. 2007. Structural stability of covalently linked GroES heptamer: Advantages in the formation of oligomeric structure. *Journal of Molecular Biology,* 367**,** 1171-1185.

Sakane, I., Ikeda, M., Matsumoto, C., Higurashi, T., Inoue, K., Hongo, K., Mizobata, T. & Kawata, Y. 2004. Structural stability of oligomeric chaperonin 10: the role of two beta-strands at the N and C termini in structural stabilization. *Journal of Molecular Biology,* 344**,** 1123-1133.

Samaranayake, H., Wirth, T., Schenkwein, D., Raty, J. K. & Yla-Herttuala, S. 2009. Challenges in monoclonal antibody-based therapies. *Annals of Medicine,* 41**,** 322-331.

Sato, A. K., Viswanathan, M., Kent, R. B. & Wood, C. R. 2006. Therapeutic peptides: technological advances driving peptides into development. *Current Opinion in Biotechnology,* 17**,** 638-642.

Schlehuber, S., Beste, G. & Skerra, A. 2000. A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *Journal of Molecular Biology,* 297**,** 1105-1120.

Schlehuber, S. & Skerra, A. 2002. Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant - a so-called 'anticalin' - using a molecular random approach. *Biophysical Chemistry,* 96**,** 213-228.

Schlehuber, S. & Skerra, A. 2005. Anticalins as an alternative to antibody technology. *Expert Opinion on Biological Therapy,* 5**,** 1453-1462.

Schmetzer, O., Moldenhauer, G., Nicolaou, A., Schlag, P., Riesenberg, R. & Pezzutto, A. 2012. Detection of circulating tumor-associated antigen depends on the domains recognized by the monoclonal antibodies used: N-terminal trimmed EpCAM-levels are much higher than untrimmed forms. *Immunology Letters,* 143**,** 184-192.

Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S. A., Baechi, T., Moelling, K. & Hovens, C. M. 1999. Mutagenesis and selection of PDZ domains that bind new protein targets. *Nature Biotechnology,* 17**,** 170-175.

Schonfeld, D., Matschiner, G., Chatwell, L., Trentmann, S., Gille, H., Hulsmeyer, M., Brown, N., Kaye, P. M., Schlehuber, S., Hohlbaum, A. M. & Skerra, A. 2009. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and

conformational features similar to antibodies. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 8198-8203.

Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E., Binz, H. K., Capitani, G., Baici, A., Pluckthun, A. & Grutter, M. G. 2007. Inhibition of caspase-2 by a designed ankyrin repeat protein: Specificity, structure, and inhibition mechanism. *Structure,* 15**,** 625-636.

Schweizer, A., Rusert, P., Berlinger, L., Ruprecht, C. R., Mann, A., Corthesy, S., Turville, S. G., Aravantinou, M., Fischer, M., Robbiani, M., Amstutz, P. & Trkola, A. 2008. CD4-specific Designed Ankyrin Repeat Proteins are novel potent HIV entry inhibitors with unique characteristics. *Plos Pathogens,* 4.

Sen, P., Komissarov, A. A., Florova, G., Idell, S., Pendurthi, U. R. & Rao, L. V. M. 2011. Plasminogen activator inhibitor-1 inhibits factor VIIa bound to tissue factor. *Journal of Thrombosis and Haemostasis,* 9**,** 531-539.

Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grutter, M. G. 2007. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *Plos Biology,* 5**,** 106-113.

Seoane Leston, J. & Diz Dios, P. 2010. Diagnostic clinical aids in oral cancer. *Oral Oncology,* 46**,** 418-422.

Shi, Z. S., Woody, R. W. & Kallenbach, N. R. 2002. Is polyproline II a major backbone conformation in unfolded proteins? *Unfolded Proteins,* 62**,** 163-240.

Shimamura, T., Koike-Takeshita, A., Yokoyama, K., Masui, R., Murai, N., Yoshida, M., aguchi, H. & Iwata, S. 2004. Crystal structure of the native chaperonin complex from Thermus thermophilus revealed unexpected asymmetry at the cis-cavity. *Structure,* 12**,** 1471-1480.

Silverman, J., Lu, Q., Bakker, A., To, W., Duguay, A., Alba, B. M., Smith, R., Rivas, A., LI, P., Le, H., Whitehorn, E., Moore, K. W., Swimmer, C., Perlroth, V., Vogt, M., Kolkman, J. & Stemmer, W. P. C. 2005. Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nature Biotechnology,* 23**,** 1556-1561.

Skerra, A. 2000a. Engineered protein scaffolds for molecular recognition. *Journal of Molecular Recognition,* 13**,** 167-187.

Skerra, A. 2000b. Lipocalins as a scaffold. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology,* 1482**,** 337-350.

Skerra, A. 2001. 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. *Journal of Biotechnology,* 74**,** 257-75.

Skerra, A. 2007a. Alternative non-antibody scaffolds for molecular recognition. *Current Opinion in Biotechnology,* 18**,** 295-304.

Skerra, A. 2007b. Anticalins as alternative binding proteins for therapeutic use. *Current Opinion in Molecular Therapeutics,* 9**,** 336-344.

Skerra, A. 2008. Alternative binding proteins: Anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *Febs Journal,* 275**,** 2677-2683.

Slavotinek, A. M. & Biesecker, L. G. 2001. Unfolding the role of chaperones and chaperonins in human disease. *Trends in Genetics,* 17**,** 528-535.

Song, S. & Naeim, F. 2004. New applications of flow cytometry in cancer diagnosis and therapy. *Cancer Diagnostics: Current and Future Trends***,** 199-232.

Steffen, A. C., Wikman, M., Tolmachev, V., Adams, G. P., Nilsson, F. Y., Stahl, S. & Carlsson, J. 2005. In vitro characterization of a bivalent anti-HER-2 affibody with potential for radionuclide-based diagnostics. *Cancer Biotherapy and Radiopharmaceuticals,* 20**,** 239-248.

Steinmeyer, D. E. & Mccormick, E. L. 2008. The art of antibody process development. *Drug Discovery Today,* 13**,** 613-618.

Street, A. G. & Mayo, S. L. 1999. Computational protein design. *Structure with Folding & Design,* 7**,** R105-R109.

Stumpp, M. T., Binz, H. K. & Amstutz, P. 2008. DARPins: A new generation of protein therapeutics. *Drug Discovery Today,* 13**,** 695-701.

Su, Z. Y. & Wang, Y. T. 2009. A Molecular Dynamics Simulation of the Human Lysozyme Camelid VHH HL6 Antibody System. *International Journal of Molecular Sciences,* 10**,** 1719-1727.

Takaishi, S., Okumura, T., Tu, S., Wang, S. W., Marrache, F., Smirnova, I., Dubeikovskiy, A., Dubeikovskaya, Z., Ai, W., Penz-Oesterreicher, M., Osawa, E., Baumgartner, J., Betz, K. S. & Wang, T. C. 2007. Isolation of gastric cancer-initiating cells using cell surface marker Cd44. *Gastroenterology,* 132**,** A632-A633.

Taneja, B. & Mande, S. C. 1999. Conserved structural features and sequence patterns in the GroES fold family. *Protein Engineering,* 12**,** 815-818.

Tanner, D. E., Chan, K. Y., Phillips, J. C. & Schulten, K. 2011. Parallel Generalized Born Implicit Solvent Calculations with NAMD. *Journal of Chemical Theory and Computation,* 7**,** 3635-3642.

Tarrega, C., Nunes-Xavier, C., Cejudo-Marin, R., Martin-Perez, J. & Pulido, R. 2010. Studying the Regulation of MAP Kinase by MAP Kinase Phosphatases In Vitro and in Cell Systems. *In:* SEGER, R. (ed.) *Map Kinase Signaling Protocols, Second Edition.* Totowa: Humana Press Inc.

Tawara, K., Oxford, J. T. & Jorcyk, C. L. 2011. Clinical significance of interleukin (IL)-6 in cancer metastasis to bone: potential of anti-IL-6 therapies. *Cancer management and research,* 3**,** 177-89.

Teicher, B. A. 2010. *Secreted Growth Factors as Therapeutic Targets*, Totowa, Humana Press Inc.

Tereshko, V., Uysal, S., Koide, A., Margalef, K., Koide, S. & Kossiakoff, A. A. 2008. Toward chaperone-assisted crystallography: Protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Science,* 17**,** 1175-1187.

To, K., Fotovati, A., Reipas, K. M., Law, J. H., Hu, K., Wang, J., Astanehe, A., Davies, A. H., Lee, L., Stratford, A. L., Raouf, A., Johnson, P., Berquin, I. M., Royer, H.-D., Eaves, C. J. & Dunn, S. E. 2010. Y-Box Binding Protein-1 Induces the Expression of CD44 and CD49f Leading to Enhanced Self-Renewal, Mammosphere Growth, and Drug Resistance. *Cancer Research,* 70**,** 2840-2851.

Tolmachev, V., Orlova, A., Pehrson, R., Galli, J., Baastrup, B., Andersson, K., Sandstrom, M., Rosik, D., Carlsson, J., Lundqvist, H., Wennborg, A. & Nilsson, F. Y. 2007.

Radionuclide therapy of HER2-positive microxenografts using a Lu-177-labeled HER2-specific affibody molecule. *Cancer Research,* 67**,** 2773-2782.

Tramontano, A. & Lesk, A. M. 1992. Common features of the conformations of antigen-binding loops in immunoglobulins and application to modelinh loop conformations *Proteins-Structure Function and Genetics,* 13**,** 231-245.

Trapasso, S. & Allegra, E. 2012. Role of CD44 as a marker of cancer stem cells in head and neck cancer. *Biologics : targets & therapy,* 6**,** 379-83.

Tresaugues, L., Collinet, B., Minard, P., Henckes, G., Aufrere, R., Blondeau, K., Liger, D., Zhou, C.-Z., Janin, J., Van Tilbeurgh, H. & Quevillon-Cheruel, S. 2004. Refolding strategies from inclusion bodies in a structural genomics project. *Journal of Structural and Functional Genomics,* 5**,** 195-204.

UniProt. 2010. *Sequence Alignment* [Online], UniProt Database, EMBL-EBI, Available: http://www.uniprot.org/ [Accessed 15-10-2010].

Van Eden, W. 2008. XToll, a recombinant chaperonin 10 as an anti-inflammatory immunomodulator. *Current Opinion in Investigational Drugs,* 9**,** 523-533.

Vasudev, N. S. & Reynolds, A. R. 2014. Anti-angiogenic therapy for cancer: current progress, unresolved questions and future directions. *Angiogenesis,* 17**,** 471-494.

Vicari, D., Foy, K. C., Liotta, E. M. & Kaumaya, P. T. P. 2011. Engineered Conformation-dependent VEGF Peptide Mimics Are Effective in Inhibiting VEGF Signaling Pathways. *Journal of Biological Chemistry,* 286**,** 13612-13625.

Vincent, T. & Mechti, N. 2004. IL-6 regulates CD44 cell surface expression on human myeloma cells. *Leukemia,* 18**,** 967-975.

Vogel, M., Keller-Gautschi, E., Baumann, M. J., Amstutz, P., Ruf, C., Kricek, F. & Stadler, B. M. 2007. Designed ankyrin repeat proteins as anti-idiotypic-binding molecules. *Autoimmunity, Part A: Basic Principles and New Diagnostic Tools.*

Vugmeyster, Y., Xu, X., Theil, F.-P., Khawli, L. A. & Leach, M. W. 2012. Pharmacokinetics and toxicology of therapeutic proteins: Advances and challenges. *World journal of biological chemistry,* 3**,** 73-92.

Walters, S., Maringe, C., Butler, J., Rachet, B., Barrett-Lee, P., Bergh, J., Boyages, J., Christiansen, P., Lee, M., Waernberg, F., Allemani, C., Engholm, G., Fornander, T., Gjerstorff, M. L., Johannesen, T. B., Lawrence, G., Mcgahan, C. E., Middleton, R., Steward, J., Tracey, E., Turner, D., Richards, M. A., Coleman, M. P. & Grp, I. M. W. 2013. Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: a population-based study. *British Journal of Cancer,* 108**,** 1195-1208.

Wang, N. F., Shi, L., Li, H. Y., Hu, Y. J., Du, W., Liu, W., Zheng, J. E., Huang, S. & Qu, X. C. 2012. Detection of circulating tumor cells and tumor stem cells in patients with breast cancer by using flow cytometry. *Tumor Biology,* 33**,** 561-569.

Weigel, M. T. & Dowsett, M. 2010. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocrine-Related Cancer,* 17**,** R245-R262.

WHO. 2002. *National Cancer Control Programmes: Policies and managerial guidelines,* Geneva, World Health Organization.

Wikman, M., Rowcliffe, E., Friedman, M., Henning, P., Lindholm, L., Olofsson, S. & Stahl, S. 2006. Selection and characterization of an HIV-1 gp120-binding affibody ligand. *Biotechnology and Applied Biochemistry,* 45**,** 93-105.

Wikman, M., Steffen, A. C., Gunneriusson, E., Tolmachev, V., Adams, G. P., Carlsson, J. & Stahl, S. 2004. Selection and characterization of HER2/neu-binding affibody ligands. *Protein Engineering Design & Selection,* 17**,** 455-462.

Wise, J. 2013. UK cancer death rates drop by more than a fifth over 20 years. *Bmj-British Medical Journal,* 347.

Wu, B. & Sun, Y. N. 2014. Pharmacokinetics of Peptide-Fc Fusion Proteins. *Journal of Pharmaceutical Sciences,* 103**,** 53-64.

Wu, X., Yan, Q., Huang, Y., Huang, H., Su, Z., Xiao, J., Zeng, Y., Wang, Y., Nie, C., Yang, Y. & Li, X. 2010. Isolation of a novel basic FGF-binding peptide with potent antiangiogenetic activity. *Journal of Cellular and Molecular Medicine,* 14**,** 351-356.

Wu, Y. B., Punta, M., Xiao, R., Acton, T. B., Sathyamoorthy, B., Dey, F., Fischer, M., Skerra, A., Rost, B., Montelione, G. T. & Szyperski, T. 2012. NMR Structure of Lipoprotein YxeF from Bacillus subtilis Reveals a Calycin Fold and Distant Homology with the Lipocalin Blc from Escherichia coli. *Plos One,* 7.

Wurch, T., Pierre, A. & Depil, S. 2012. Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept. *Trends in Biotechnology,* 30**,** 575-582.

Xu, L. H., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H. X., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W. & Lipovsek, D. 2002. Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & Biology,* 9**,** 933-942.

Xu, Z. H., Horwich, A. L. & Sigler, P. B. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)(7) chaperonin complex. *Nature,* 388**,** 741-750.

Yamazaki, K., Nagao, T., Yamaguchi, T., Saisho, H. & Kondo, Y. 1997. Expression of basic fibroblast growth factor (FGF-2)-associated with tumour proliferation in human pancreatic carcinoma. *Virchows Archiv-an International Journal of Pathology,* 431**,** 95-101.

Zahnd, C., Kawe, M., Stumpp, M. T., de Pasquale, C., Tamaskovic, R., Nagy-Davidescu, G., Dreier, B., Schibli, R., Binz, H. K., Waibel, R. & Plueckthun, A. 2010. Efficient Tumor Targeting with High-Affinity Designed Ankyrin Repeat Proteins: Effects of Affinity and Molecular Size. *Cancer Research,* 70**,** 1595-1605.

Zahnd, C., Wyler, E., Schwenk, J. M., Steiner, D., Lawrence, M. C., Mckern, N. M., Pecorari, F., Ward, C. W., Joos, T. O. & Pluckthun, A. 2007. A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *Journal of Molecular Biology,* 369**,** 1015-1028.

Zhang, J. B., Tanha, J., Hirama, T., Khieu, N. H., To, R., Hong, T. S., Stone, E., Brisson, J. R. & Mackenzie, C. R. 2004. Pentamerization of single-domain antibodies from phage libraries: A novel strategy for the rapid generation of high-avidity antibody reagents. *Journal of Molecular Biology,* 335**,** 49-56.

Zoeller, M. 2011. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nature Reviews Cancer,* 11**,** 254-267.

Zoon, C. K., Starker, E. Q., Wilson, A. M., Emmert-Buck, M. R., Libutti, S. K. & Tangrea, M. A. 2009. Current molecular diagnostics of breast cancer and the potential incorporation of microRNA. *Expert Review of Molecular Diagnostics,* 9**,** 455-467.

Appendix 1

Amino acid sequences and theoretical pI/Mwt for all protein candidates used in this study:

1. hCpn10

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAETVTKGG IMLPEKSQGK VLQATVVAVG SGSKGKGGEI

         70          80          90         100
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VD
```

Theoretical pI/Mwt: 8.89 / 10931 Da (per monomer)

2. ΔL-Cpn10

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAGKVLQATV VAVGSGSKGK GGEIQPVSVK VGDKVLLPEY

         70          80
GGTKVVLDDK DYFLFRDGDI LGKYVD
```

Theoretical pI/Mwt: 8.96 / 9260 Da (per monomer)

3. ΔRf-Cpn10

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAETVTKGG IMLPEKSQGK VLQATVVAVG GGVKVGDKVL

         70          80          90
LPEYGGTKVV LDDKDYFLFR DGDILGKYVD
```

Theoretical pI/Mwt: 7.94 / 9733 Da (per monomer)

4. β-barrel (ΔRf-ΔL-Cpn10)

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAGGKVLQAT VVAVGGGPVS VKVGDKVLLP EYGGTKVVLD

         70
DKDYFLFRDG DILGKYVD
```

Theoretical pI/Mwt: 7.98 / 8402 Da (per monomer)

5. CE76

```
          10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SALCDDPRVD RWYCQFVEGK VLQATVVAVG SGSKGKGGEI

          70          80          90         100
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VD
```

Theoretical pI/Mwt: 6.21 / 11287 Da (per monomer)

6. CE76$_{P1}$

```
          10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAAPLCDDP RVDRWYCQFV EPAAGKVLQA TVVAVGSGSK

          70          80          90         100
GKGGEIQPVS VKVGDKVLLP EYGGTKVVLD DKDYFLFRDG DILGKYVD
```

Theoretical pI/Mwt: 6.21 / 11765 Da (per monomer)

7. CE76$_{P2}$

```
          10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAPPLCDDPR VDRWYCQFVE PPGKVLQATV VAVGSGSKGK

          70          80          90         100
GGEIQPVSVK VGDKVLLPEY GGTKVVLDDK DYFLFRDGDI LGKYVD
```

Theoretical pI/Mwt: 6.21 / 11675 Da (per monomer)

8. CE76$_{E5}$

```
          10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAETVTLCD DPRVDRWYCQ FVEPEKSQGK VLQATVVAVG

          70          80          90         100         110
SGSKGKGGEI QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VD
```

Theoretical pI/Mwt: 5.43 / 12358 Da (per monomer)

9. FH-Cpn10

```
        10         20         30         40         50         60
MAGQAFRKFL PLFDRVLVER SAAETVTKGG IMLPEKSQGK VLQATVVAVG SGSKGKGGEI

        70         80         90        100        110        120
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VDGGGAGQAF RKFLPLFDRV

       130        140        150        160        170        180
LVERSAAETV TKGGIMLPEK SQGKVLQATV VAVGSGSKGK GGEIQPVSVK VGDKVLLPEY

       190        200        210        220        230        240
GGTKVVLDDK DYFLFRDGDI LGKYVDGGGA GQAFRKFLPL FDRVLVERSA AETVTKGGIM

       250        260        270        280        290        300
LPEKSQGKVL QATVVAVGSG SKGKGGEIQP VSVKVGDKVL LPEYGGTKVV LDDKDYFLFR

       310        320        330        340        350        360
DGDILGKYVD GGGAGQAFRK FLPLFDRVLV ERSAAETVTK GGIMLPEKSQ GKVLQATVVA

       370        380        390        400        410        420
VGSGSKGKGG EIQPVSVKVG DKVLLPEYGG TKVVLDDKDY FLFRDGDILG KYVDGGGAGQ

       430        440        450        460        470        480
AFRKFLPLFD RVLVERSAAE TVTKGGIMLP EKSQGKVLQA TVVAVGSGSK GKGGEIQPVS

       490        500        510        520        530        540
VKVGDKVLLP EYGGTKVVLD DKDYFLFRDG DILGKYVDGG GAGQAFRKFL PLFDRVLVER

       550        560        570        580        590        600
SAAETVTKGG IMLPEKSQGK VLQATVVAVG SGSKGKGGEI QPVSVKVGDK VLLPEYGGTK

       610        620        630        640        650        660
VVLDDKDYFL FRDGDILGKY VDGGGAGQAF RKFLPLFDRV LVERSAAETV TKGGIMLPEK

       670        680        690        700        710        720
SQGKVLQATV VAVGSGSKGK GGEIQPVSVK VGDKVLLPEY GGTKVVLDDK DYFLFRDGDI
```

LGKYVD

Theoretical pI/Mwt: 9.18 / 76653 Da (per heptamer)

10. FH-CE76

```
        10         20         30         40         50         60
MAGQAFRKFL PLFDRVLVER SALCDDPRVD RWYCQFVEGK VLQATVVAVG SGSKGKGGEI

        70         80         90        100        110        120
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VDGGGAGQAF RKFLPLFDRV

       130        140        150        160        170        180
LVERSALCDD PRVDRWYCQF VEGKVLQATV VAVGSGSKGK GGEIQPVSVK VGDKVLLPEY

       190        200        210        220        230        240
GGTKVVLDDK DYFLFRDGDI LGKYVDGGGA GQAFRKFLPL FDRVLVERSA LCDDPRVDRW

       250        260        270        280        290        300
YCQFVEGKVL QATVVAVGSG SKGKGGEIQP VSVKVGDKVL LPEYGGTKVV LDDKDYFLFR

       310        320        330        340        350        360
DGDILGKYVD GGGAGQAFRK FLPLFDRVLV ERSALCDDPR VDRWYCQFVE GKVLQATVVA

       370        380        390        400        410        420
VGSGSKGKGG EIQPVSVKVG DKVLLPEYGG TKVVLDDKDY FLFRDGDILG KYVDGGGAGQ

       430        440        450        460        470        480
AFRKFLPLFD RVLVERSALC DDPRVDRWYC QFVEGKVLQA TVVAVGSGSK GKGGEIQPVS

       490        500        510        520        530        540
VKVGDKVLLP EYGGTKVVLD DKDYFLFRDG DILGKYVDGG GAGQAFRKFL PLFDRVLVER

       550        560        570        580        590        600
SALCDDPRVD RWYCQFVEGK VLQATVVAVG SGSKGKGGEI QPVSVKVGDK VLLPEYGGTK

       610        620        630        640        650        660
VVLDDKDYFL FRDGDILGKY VDGGGAGQAF RKFLPLFDRV LVERSALCDD PRVDRWYCQF

       670        680        690        700        710        720
VEGKVLQATV VAVGSGSKGK GGEIQPVSVK VGDKVLLPEY GGTKVVLDDK DYFLFRDGDI
```

LGKYVD

Theoretical pI/Mwt: 6.62 / 79140 Da (per heptamer)

11. CPMyc

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAEQKLISEE DLGKVLQATV VAVGSGSKGK GGEIQPVSVK

         70          80          90
VGDKVLLPEY GGTKVVLDDK DYFLFRDGDI LGKYVD
```

Theoretical pI/Mwt: 5.39 / 10446 Da (per monomer)

12. CPMyc$_{P1}$

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAAPEQKLI SEEDLPAAGK VLQATVVAVG SGSKGKGGEI

         70          80          90         100
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VD
```

Theoretical pI/Mwt: 5.39 / 10924 Da (per monomer)

13. CP7$_{CD44-P1}$

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAAPFNLPL PSRPLLRPAA GKVLQATVVA VGSGSKGKGG

         70          80          90         100
EIQPVSVKVG DKVLLPEYGG TKVVLDDKDY FLFRDGDILG KYVD
```

Theoretical pI/Mwt: 9.52 / 11144 Da (per monomer)

14. CP11$_{bFGF-P1}$

```
         10          20          30          40          50          60
MAGQAFRKFL PLFDRVLVER SAAAPPLLQT LGGGSPAAGK VLQATVVAVG SGSKGKGGEI

         70          80          90         100
QPVSVKVGDK VLLPEYGGTK VVLDDKDYFL FRDGDILGKY VD
```

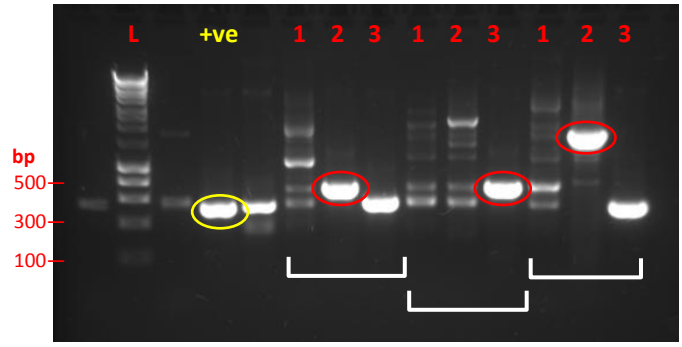
Theoretical pI/Mwt: 8.96 / 10663 Da (per monomer)

15. CR2$_{\text{-VEGF1-P2}}$

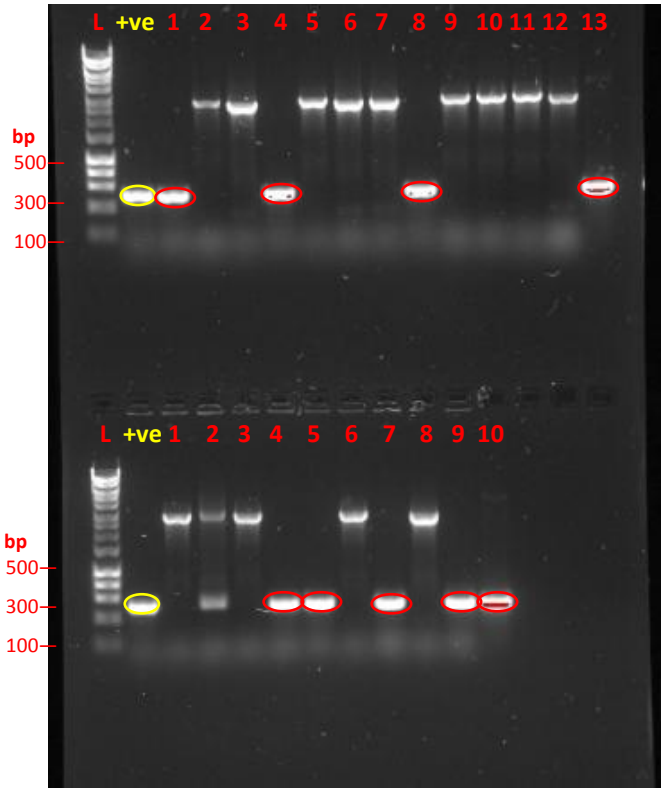
```
         10         20         30         40         50         60
MAGQAFRKFL PLFDRVLVER SAPPITMQIM RIKPHQGQHI GEMSFPPGKV LQATVVAVGS

         70         80         90        100        110
GSKGKGGEIQ PVSVKVGDKV LLPEYGGTKV VLDDKDYFLF RDGDILGKYV D
```

Theoretical pI/Mwt: 9.26 / 12114 Da (per monomer)

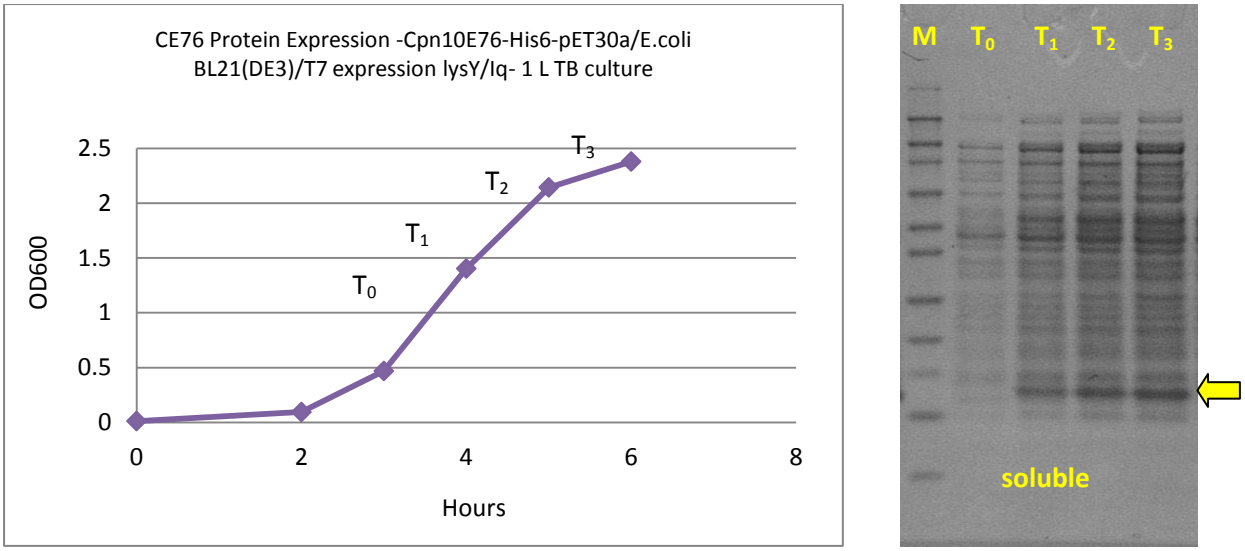
Appendix 2



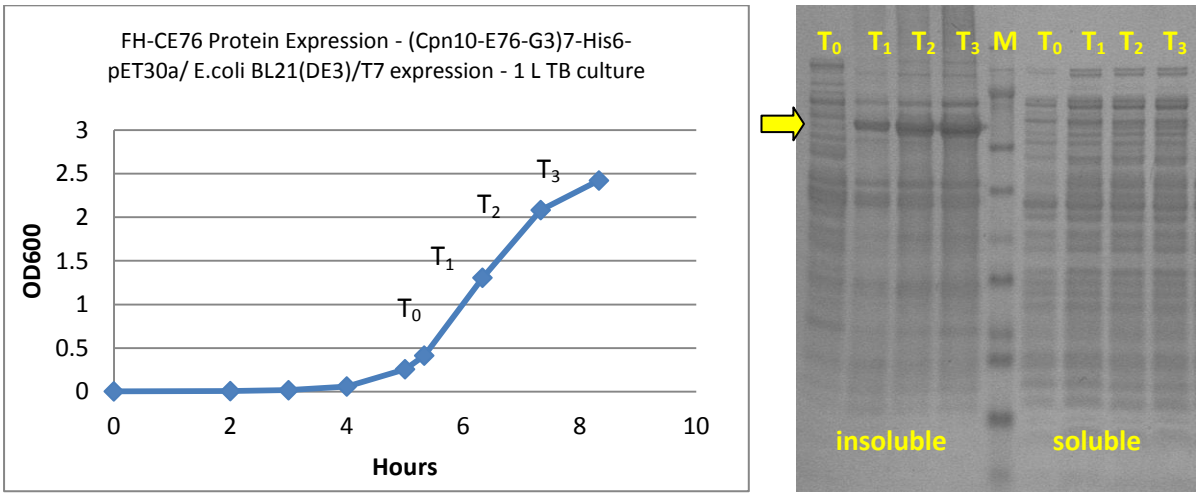
Representation example shows PCR product of hCpn10 (+ve, control), CE76, CE76$_{P1}$, and FH-CE76, respectively, using HyperLadder™ 100 bp standard. Red ovals illustrate a correct bp size.



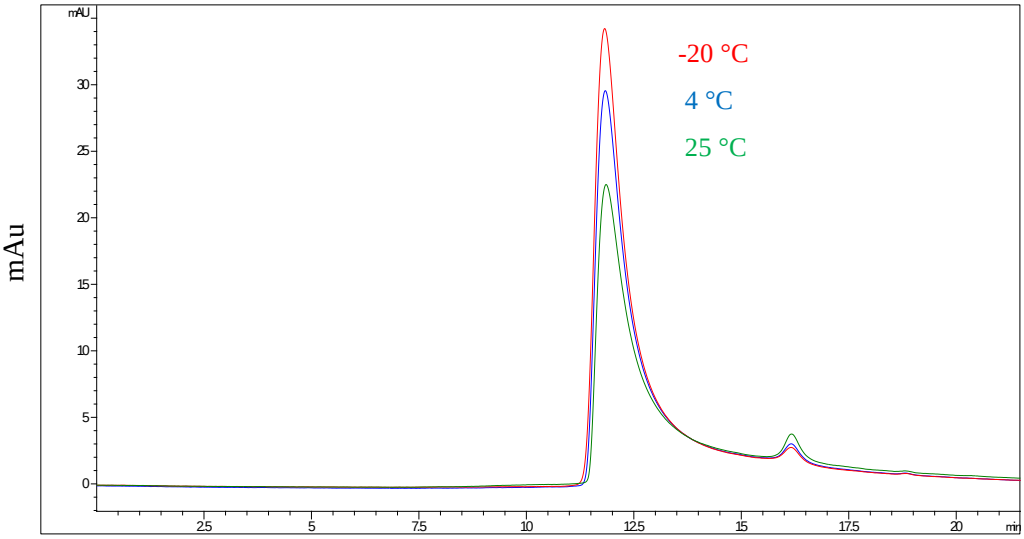
Representation example shows cPCR of transformed cells with ligated product. **(A)** cPCR product of CE76$_{P1}$ (from 1-13 colony), and **(B)** cPCR product of CPMyc$_{P1}$ (from 1-10 colony). Using HyperLadder™ 100 bp standard. Yellow ellipses were (+ve) controls and red ellipses illustrated colony with positive oligonucleotide sizes.



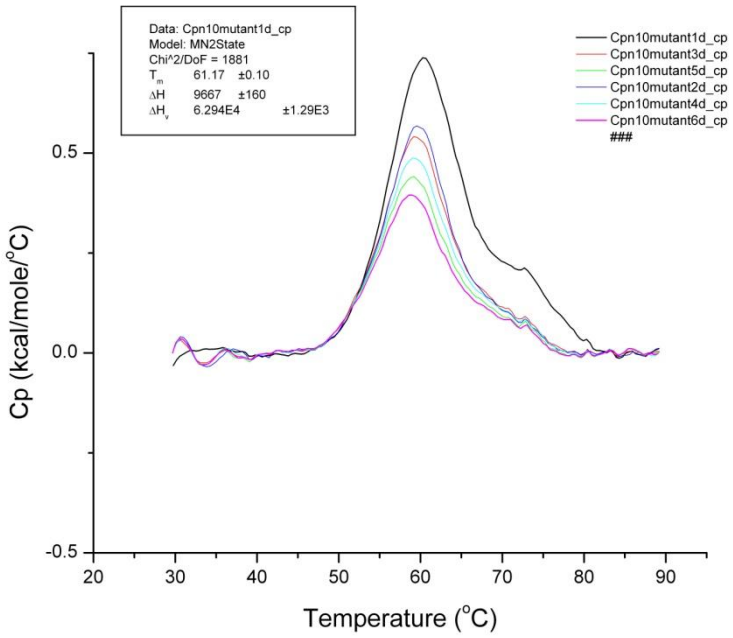
Example of protein expression of CE76 induced by 0.04 mM IPTG during log-phase and confirmed by SDS-PAGE. M: SeeBlue® Plus2 Mwt standard, T_0: pre-induction, T_1-T_3: post-induction.



Example of protein expression of FH-CE76 induced by 0.04 mM IPTG during log-phase and confirmed by SDS-PAGE. M: SeeBlue® Plus2 Mwt standard, T_0: pre-induction, T_1-T_3: post-induction.

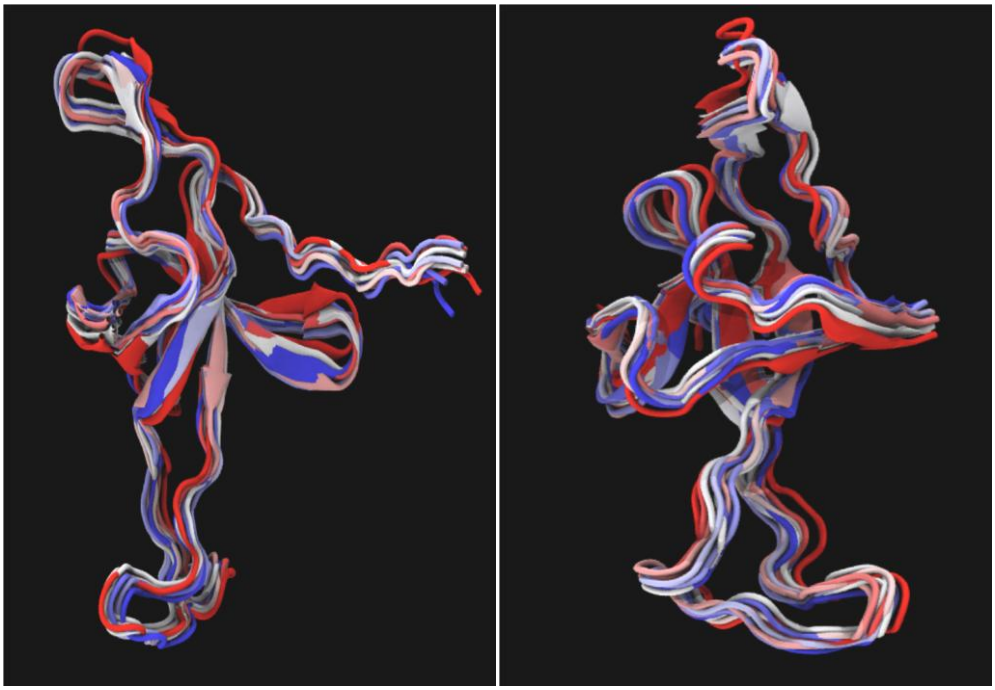


Example of SE-HPLC stack chromatograms of $CE76_{P2}$ stored for 30 days at -20, 4, and 25 °C, respectively. By using Agilent™ 1200 HPLC system and Tosoh™ TSK column and chromatograms were stacked to illustrated protein $CE76_{P2}$ solubilities, stabilities, and degradations.

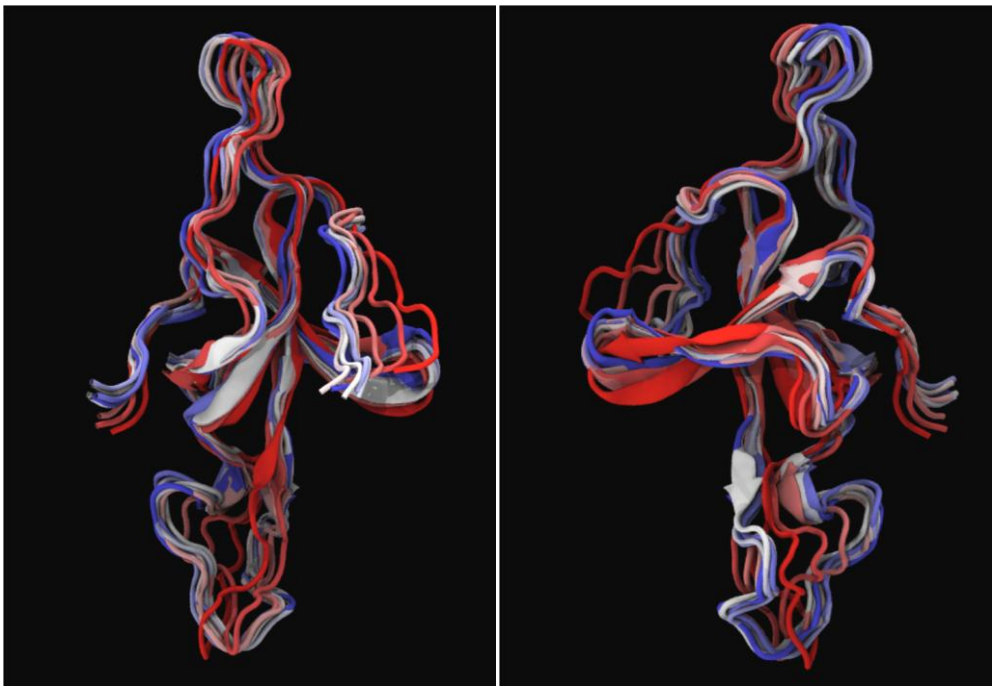


Respresination of $CE76_{P1}$ denaturing and refolding study by DSC. This is a preliminary result from a single run and not verified.

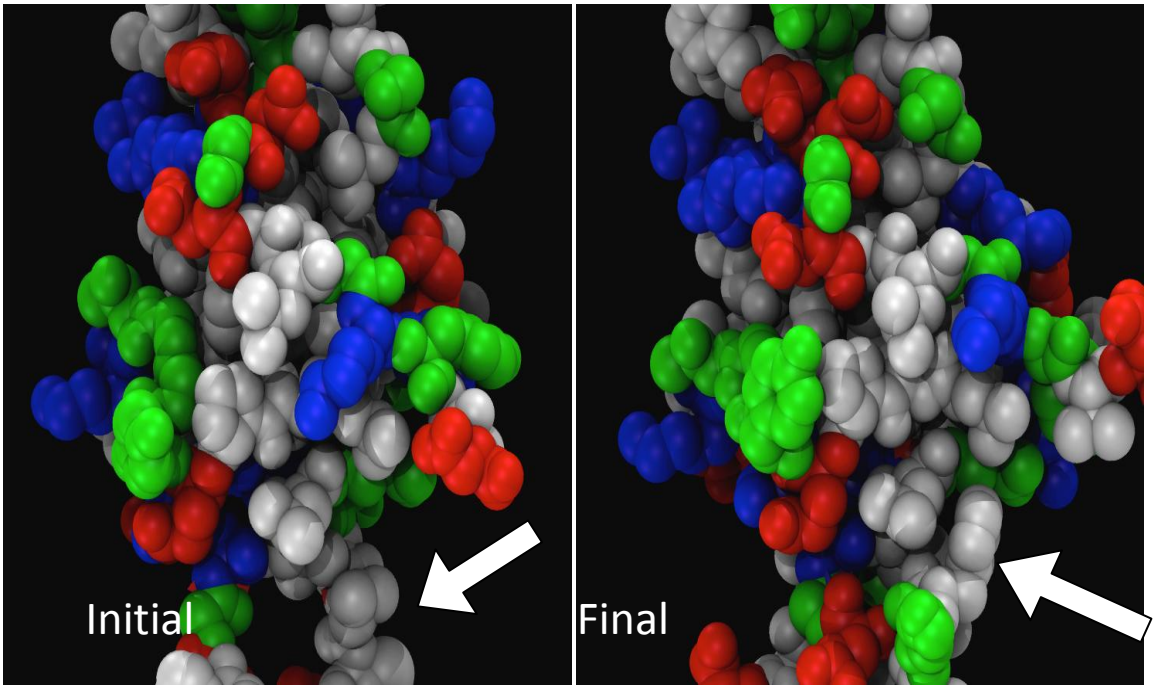
Appendix 3



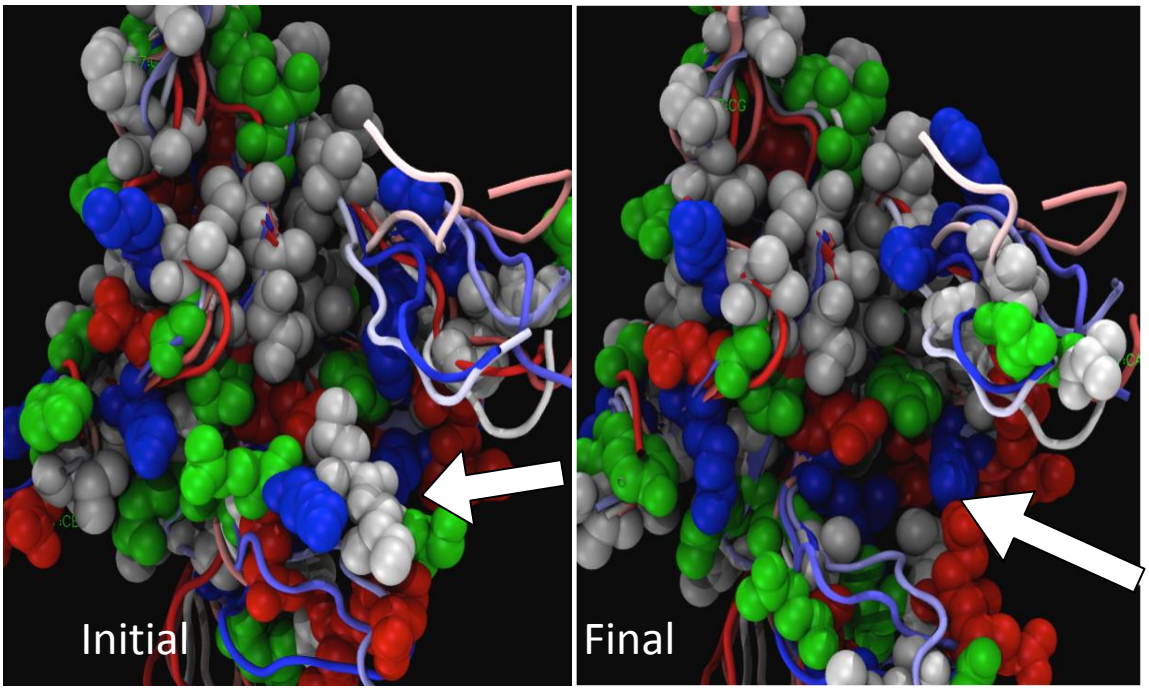
Snapshots of hCpn10 MD simulations over 1 ns duriation. Show overlaid monomer protein backbone during last 700 ps simulation taken every 50 ps. The monomer structures range from initial (red) to the final structure (blue), showing in two different views.



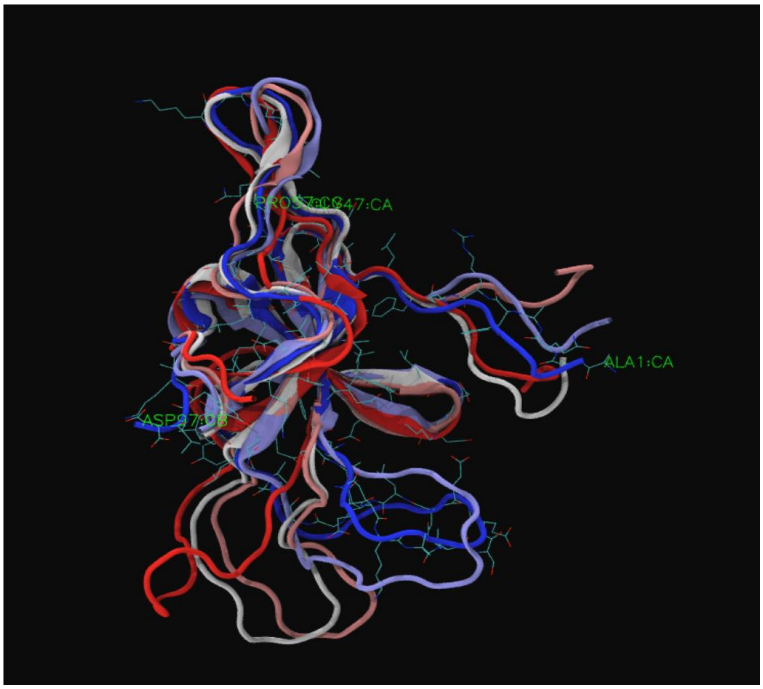
Snapshots of CE76 MD simulations over 1 ns duriation. Show overlaid monomer protein backbone during last 700 ps simulation taken every 50 ps. The monomer structures range from initial (red) to the final structure (blue), showing in two different views.



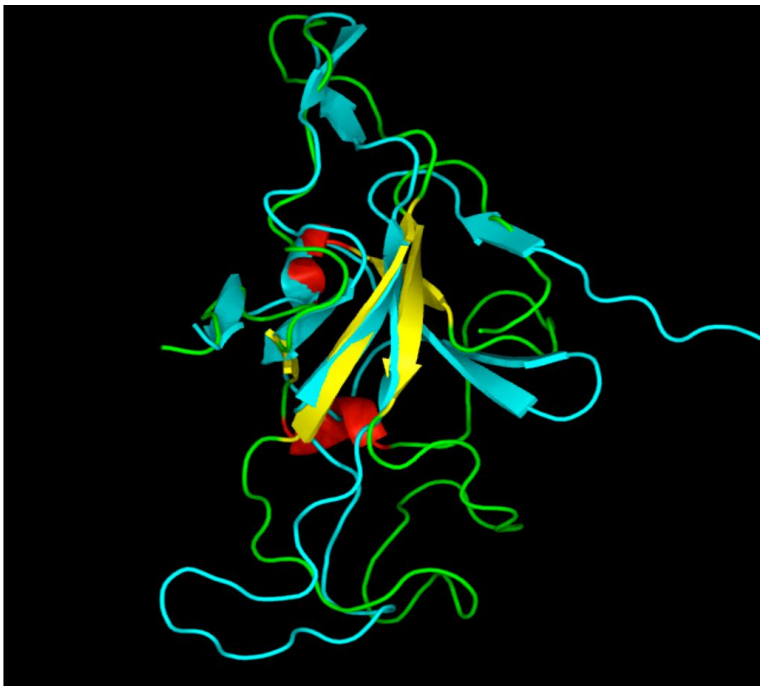
Snapshots of CE76 over 1 ns duriation MD simulations. Show only initial and final of CE76 monomer in space-filling during last 700 ps simulation.



Snapshots of CPMyc over 1 ns duriation MD simulations. Show only initial and final of CPMyc monomer in space-filling during last 700 ps simulation.

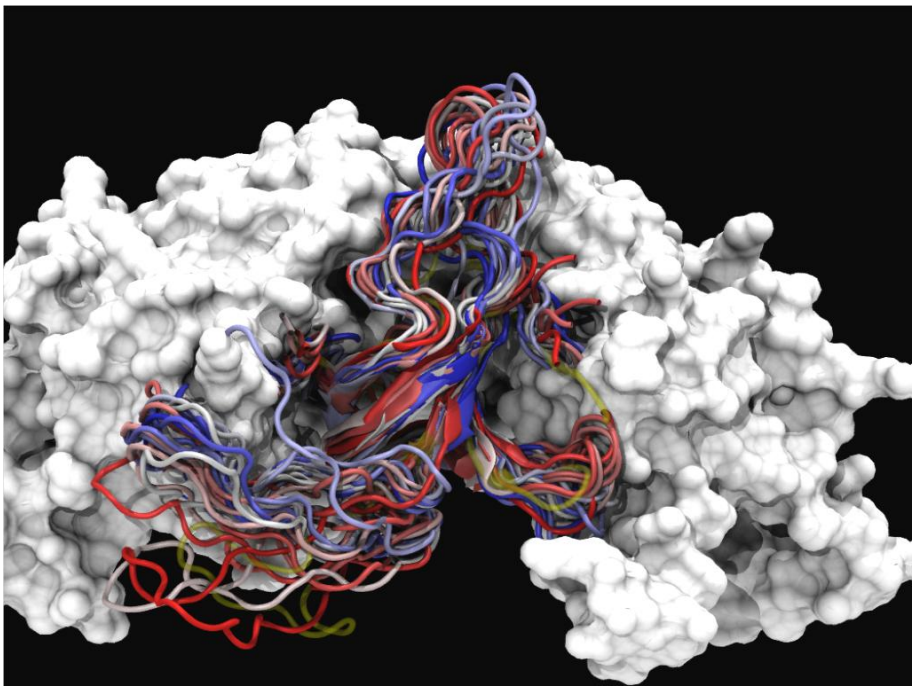


Sanapshots of CMyc over 1 ns duration MD simulation. Shows overlaid monomer backbone of the last half of 700 ps simulation taken every 50 ps. The monomer structures range from initial (red) to the final structure (blue).

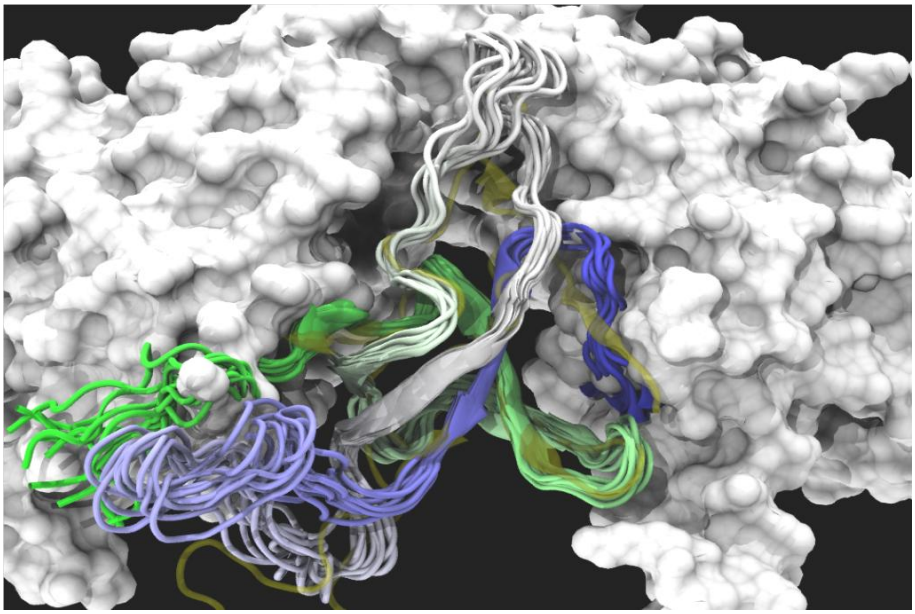


Overlaied wt Cpn10 X-ray monomer "U" from PDB ID. 1WE3 and the final structure of CE76 form 1 ns MD simulations. CE76 is coloured according to secondary structure (red: helix, yellow: strands, and green: loops), and wt Cpn10 (1WE3, cyan).

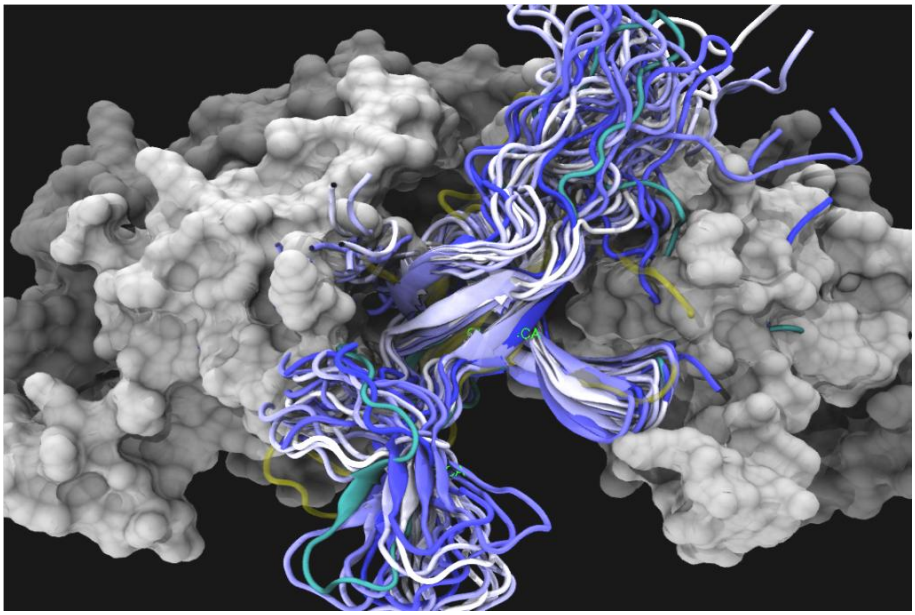
MD simulation of $CE76_{P1}$ full length in ε=78.5 over 1 ns duration. The $CE76_{-P1}$ monomer simulated is shown in a coloured ribbon representation: the β-sheet of the protein core is show in yellow, turns are shown in cyan and random conformations are show as a white ribbon. A wt Cpn10 (1WE3) monomer is shown in red and is located in an orientation and position necessary for a wt Cpn10 or it variant to form heptamer. The mobile loops are located towards the bottom of the figure. The red protein was not included in the simulation.



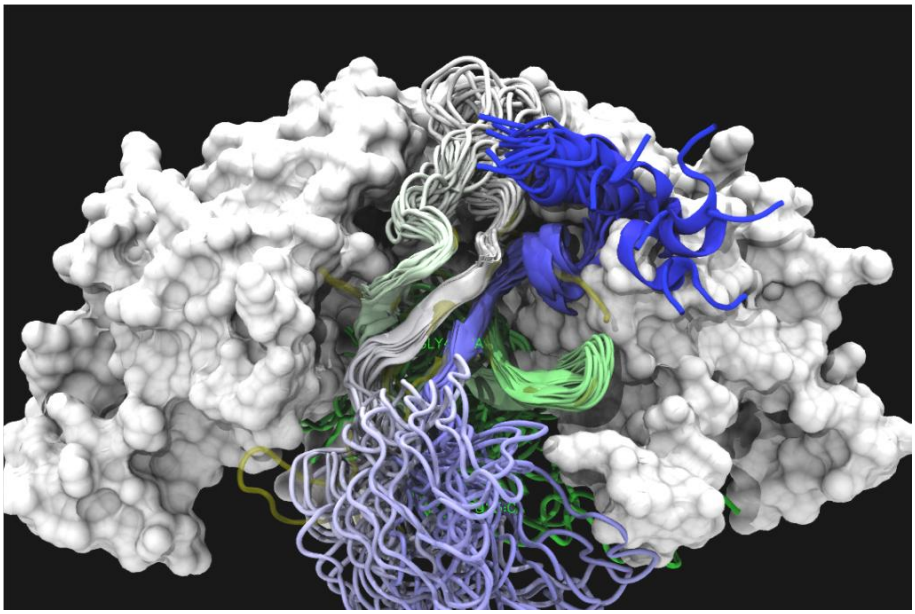
MD simulation of $CE76_{P1}$ with the *His*-tag removed in ε=78.5 over 1 ns duration. The $CE76_{P1}$ simulated monomer is shown in ribbon representation as a series of snapshots from the simulation in 40 ps intervals. The snapshots are coloured according to simulation time red (start) to blue (end). $CE76_{P1}$ was overlaid onto a hypothetical heptameric wt Cpn10 (1WE3) structure which was not included in simulation. $CE76_{P1}$ monomer is located in a cut-out of the heptamer to estimate the ability to form heptamers.



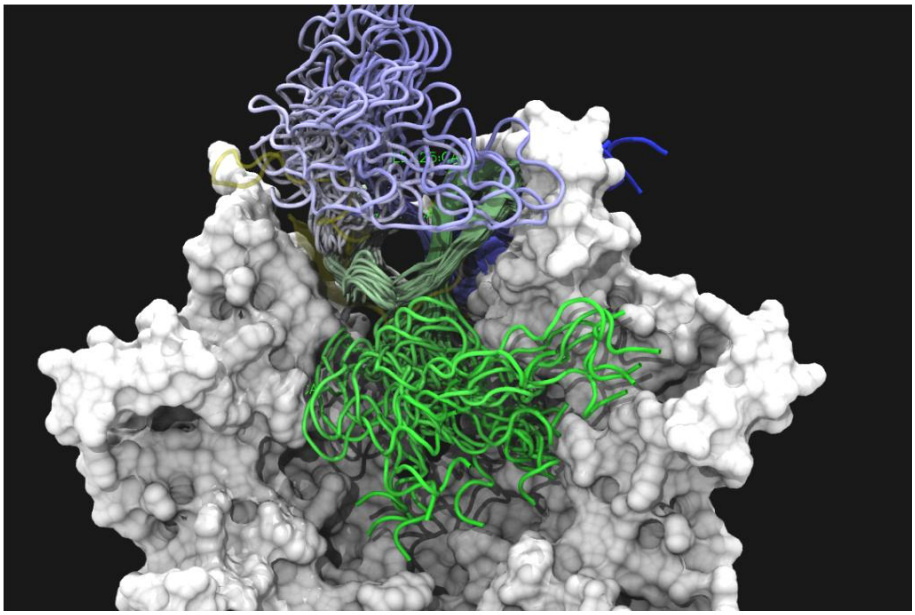
MD simulation of $CE76_{P1}$ in ε=40 over 1 ns duration. The $CE76_{P1}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according residue number: green (*N*-terminus) to blue (*C*-terminus).



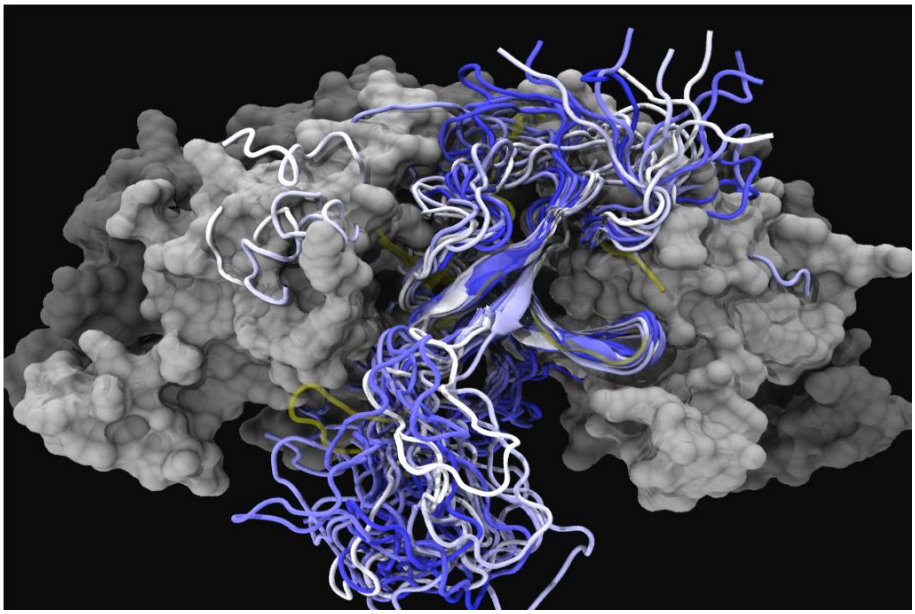
MD simulation of $CE76_{P2}$ in ε=78.5 with *His*-tag removed over 1 ns duration. The $CE76_{P2}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from light to dark blue.



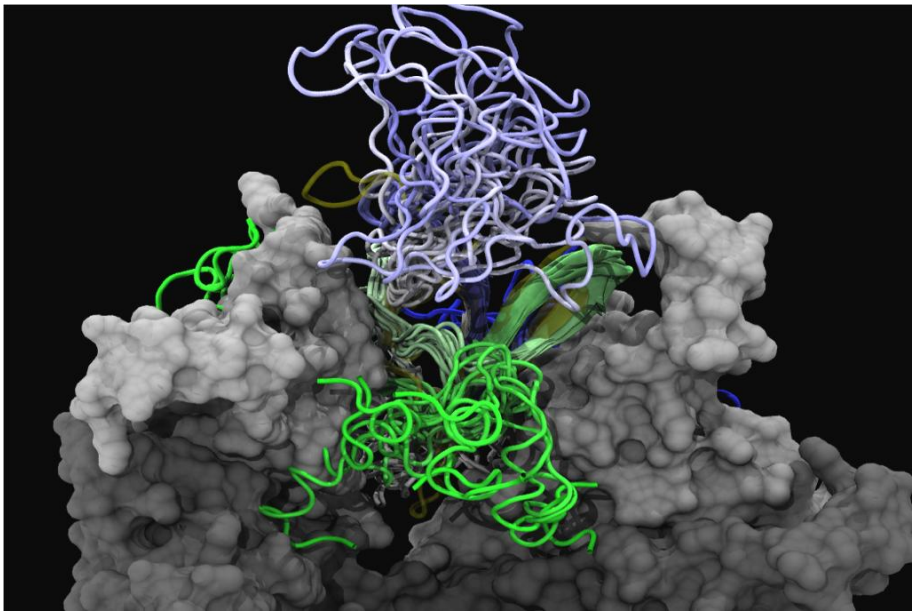
MD simulation of CE76$_{P2}$ full length in ε=40 over 1 ns duration. The CE76$_{P2}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from light to dark blue.



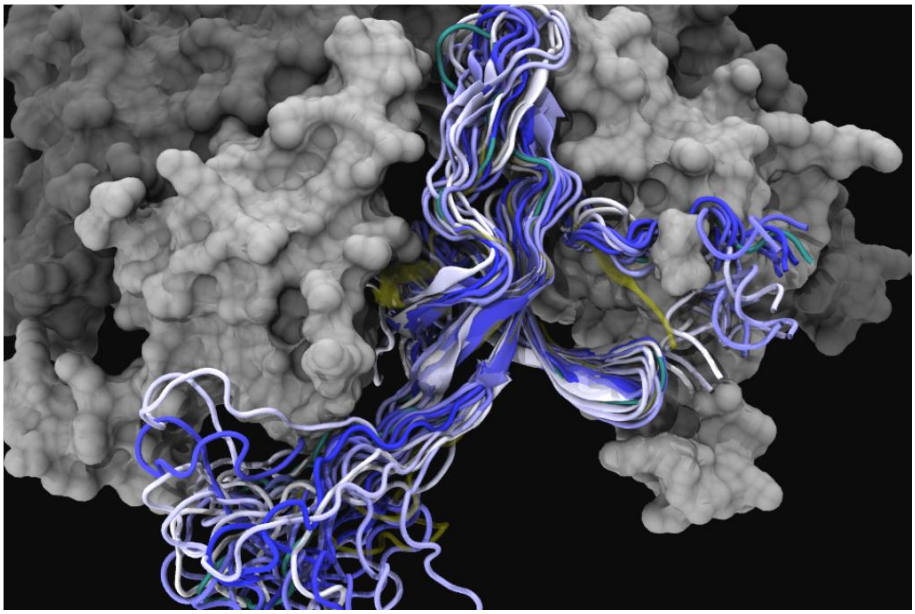
MD simulation of CE76$_{P2}$ in ε=40 over 1 ns duration. Same as above Fig., but the protein is rotated by 90° around the horizontal axis to allow a view from "bottom" the potential hexameric structure.



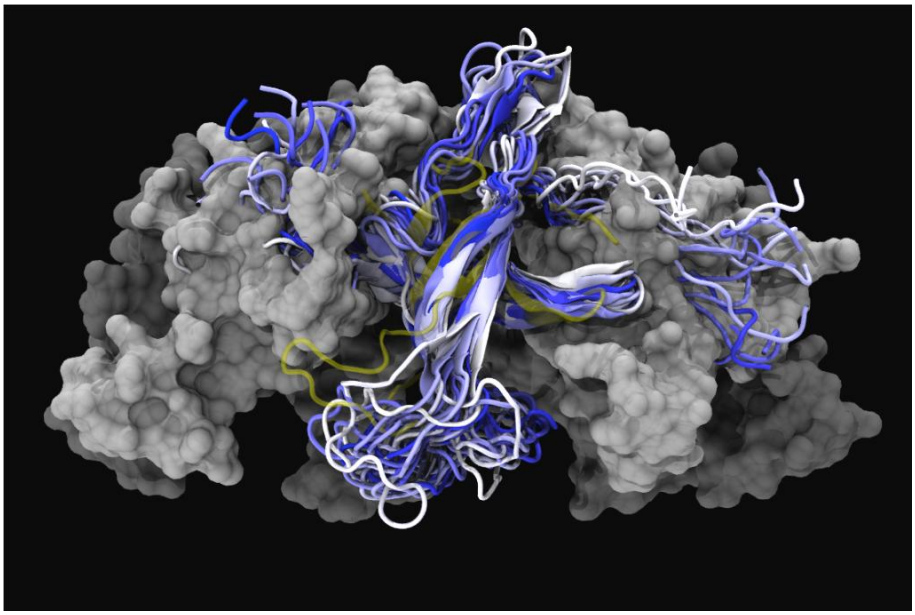
MD simulation of $CE76_{E5}$ in ε=78.5 over 1 ns duration. The $CE76_{E5}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from white to dark blue.



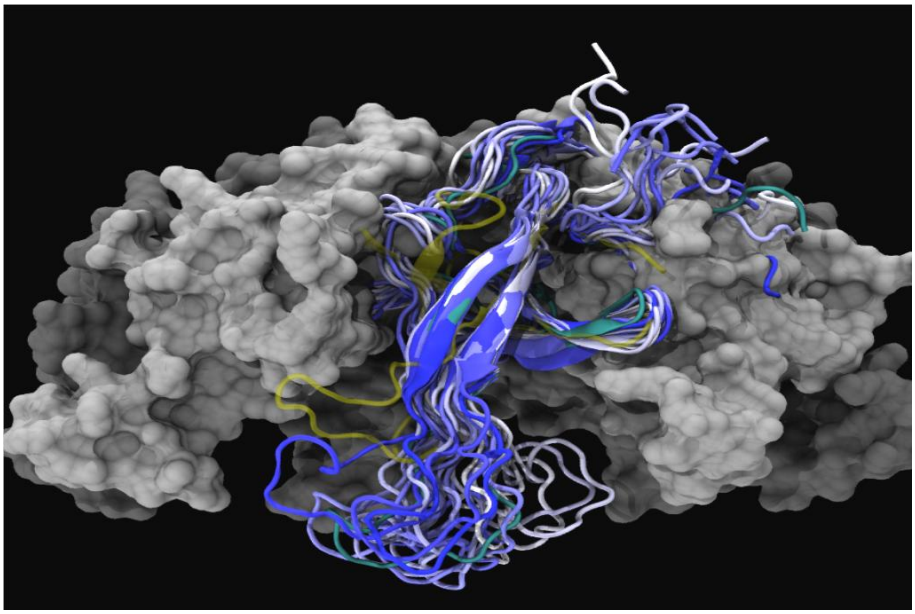
MD simulation of $CE76_{E5}$ in ε=78.5 over 1 ns duration. Similar to above Fig., but the protein is rotated by 90° around the horizontal axis to allow a view from "underneath" the potential hexameric structure.



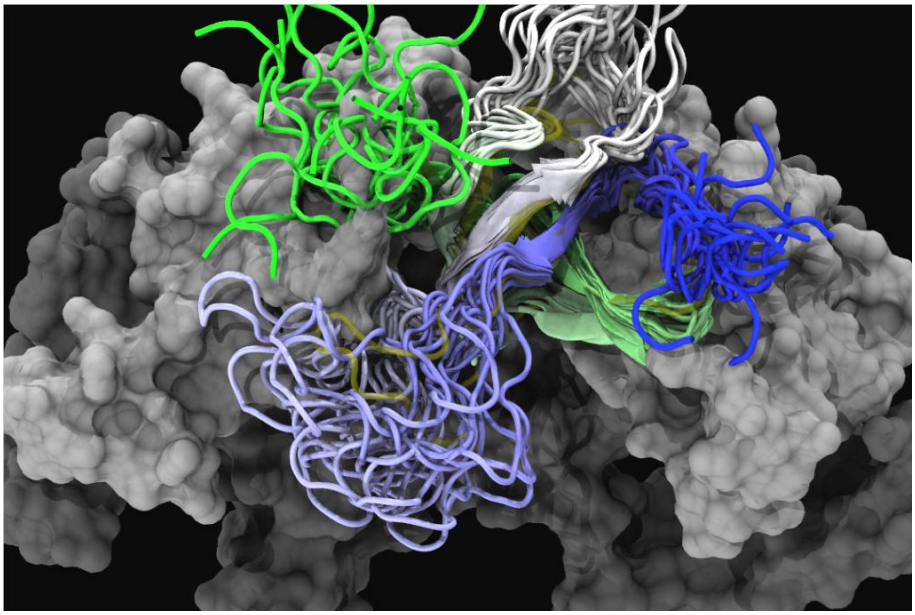
MD simulation of $CE76_{E5}$ in ε=78.5 without *His*-tag over 1 ns duration. The $CE76_{E5}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from light to dark blue.



MD simulation of $CE76_{E5}$ in ε=40 over 1 ns duration. The $CE76_{E5}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from light to dark blue.

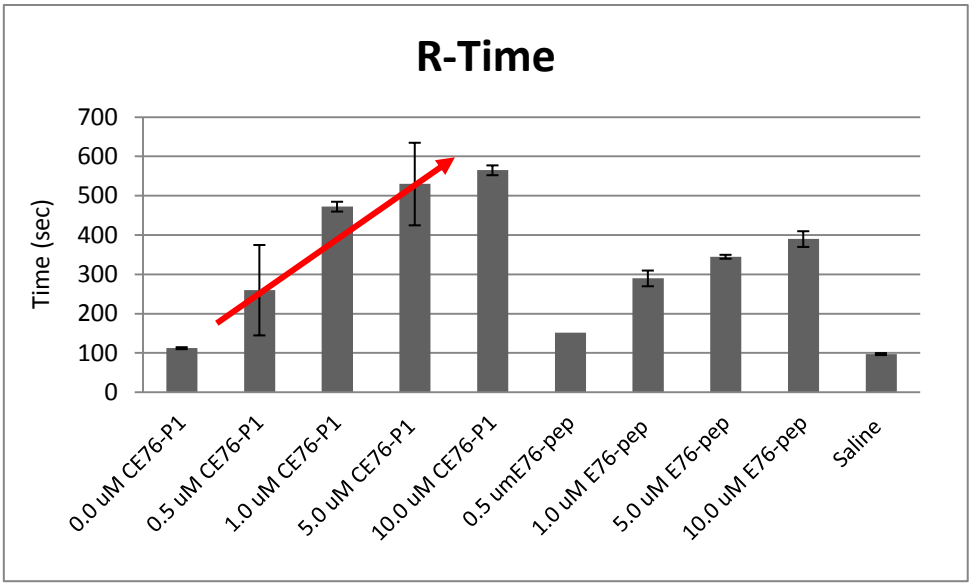


MD simulation of CPMyc$_{P1}$ in ε=78.5 without *His*-tag over 1 ns duration. The CPMyc$_{P1}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to simulation time from light to dark blue.

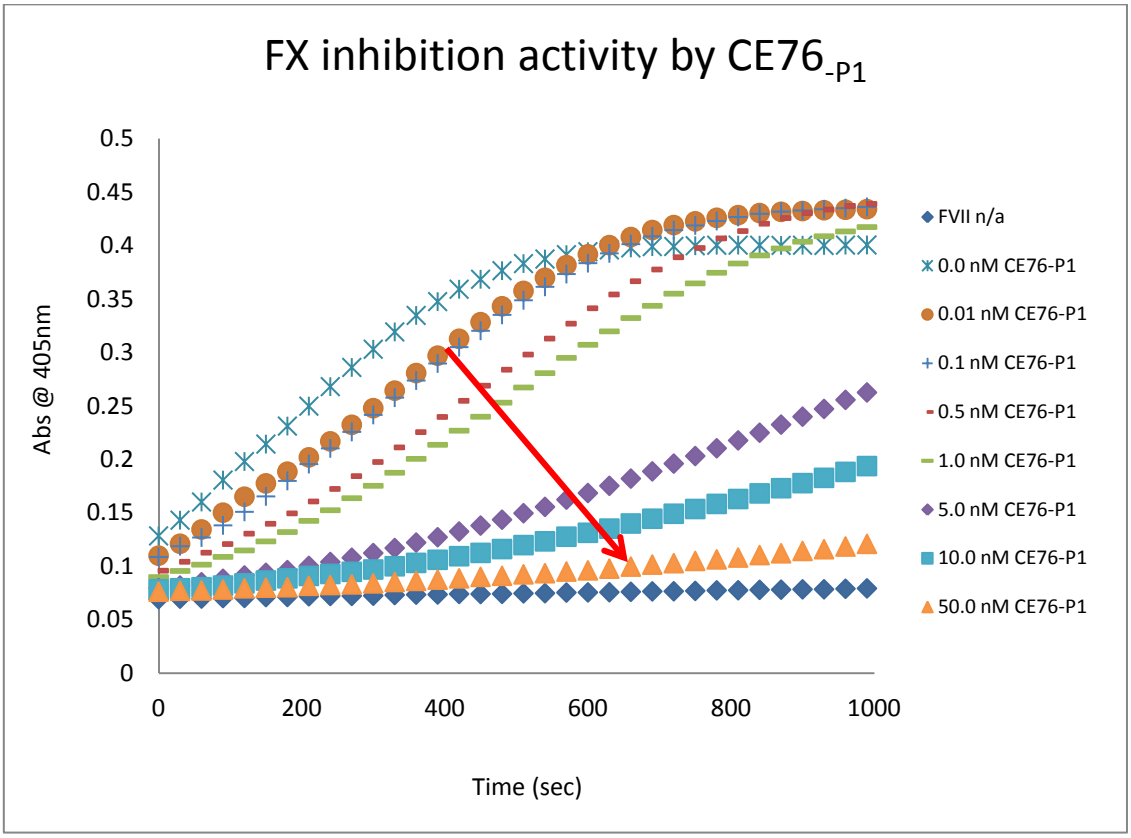


MD simulation of CPMyc$_{P1}$ in ε=40 over 1 ns. The CPMyc$_{P1}$ simulated monomer is shown in ribbon representation as a series of snapshots from the last 100 ps simulation in 10 ps intervals. The snapshots are coloured according to residue sequence number time from green (*N*-terminus) to blue (*C*-terminus).

Appendix 4



Anticoagulant Effect of $CE76_{P1}$ vs E-76 peptide (control) on the clot generation time (R-time) by TEG using human blood. Blood clotting initiated by Thromborel® S and Ca^{2+}. Uninhibited clotting times were 100 sec.



Chromogenic assays for determination the $CE76_{P1}$ median inhibition concentration (IC50) for FX. $CE76_{P1}$ IC_{50} = 1.4994 ±0.4763 nM.