Engineering protein folding and interactions: Towards potential novel therapeutic interventions

Callum F G Ormonde, BSc (Hons)

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia

School of Chemistry and Biochemistry

2017

THESIS DECLARATION

I, Callum Ormonde, certify that:

This thesis has been substantially accomplished during enrolment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of The University of Western Australia and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

The research involving animal data reported in this thesis was assessed and approved by The University of Western Australia Animal Ethics Committee. Approval #: RA/3/100/1336

The research involving animals reported in this thesis followed The University of Western Australia and national standards for the care and use of laboratory animals.

The work described in this thesis was funded by the Australian Research Council and National Health and Medical Research Council.

Callum F G Ormonde

Candidate

Contents

List of Figures

List of Tables

Abbreviations

Antp	Antennapedia peptide
Bcl-2	Apoptosis regulator Bcl-2
BG4	G4 binding antibody
BLBC	Basal-like breast cancer
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CAV	Caveolin
CDK	Cyclin-dependent kinases
ChIP	Chromatin immunoprecipitation
CI	Combination index
c-Kit	Receptor tyrosine kinase
c-Myb	Myb proto-oncogene protein
c-MYC	Myc proto-oncogene protein
CPP	Cell-penetrating peptide
DamID	DNA adenine methyltransferase identification
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOX	Doxorubicin
DTX	Docetaxel
dUTP	Deoxyuridine 5'-triphosphate nucleotidohydrolase
E.coli	*Escherichia coli*
EGFR	Endothelial growth factor receptor
EN1	Engrailed 1
ER	Estrogen receptors
FnBPs	Fibronectin-binding proteins

FRET	Förster resonance energy transfer
FT-IR	Fourier transform infrared spectroscopy
G	Guanine
G4	G-quadruplexes
GMP	Guanylic acid
GPCRs	G-protein couple receptors
GST	Glutathione s-transferase
HD	Homeodomain
HER2	Human epidermal growth factor receptor-2
HEWL	Hens egg white lysozyme
HIC	Hydrophobic interaction chromatography
HIF-1	Hypoxia-inducible factors
HOX	Homeobox gene
HPLC	High-performance liquid chromatography
hTERT	Human telomerase reverse transcriptase
htt	Ubiquitous huntingtin protein
IEX	Ion exchange chromatography
IMAC	Immobilized metal affinity chromatography
iPep	Interference peptide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KRAS	Kirsten rat sarcoma viral oncogene
mAbs	Monoclonal antibodies
MBP	Maltose binding protein
MIP	Maximum intensity projection
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NLS	Nuclear localisation sequence
NP	Nanoparticle
NusA	N-utilizing substance A
PAA	Poly(acrylic acid)
PARP	Poly (ADP-ribose) polymerase

PBS	Phosphate-buffered saline
PBX1	Pre-B-cell leukaemia transcription factor gene
pCR	Pathological complete response
PDB	Protein data bank
PDGF-A	Platelet-derived growth factor subunit A
PdI	Polydispersibility index
PEI	Polymer polyethylenimine
PG4	Potential G-quadruplex
PGMA	Poly(glycidyl methacrylate)
PgR	Progesterone receptors
PKA	Protein kinase A
PLK1	Polo-like kinase 1
Ret	Rearranged during transfection proto-oncogene
RhB	Rhodamine B
rhGH	Recombinant human growth hormone
RT	Room temperature
SD	Standard deviation
SEC	Size exclusion chromatography
SEM	Standard error of the mean
shRNA	Short hairpin ribonucleic acid
SMCC	Succinimidyl-4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate
SUMO	Small ubiquitin-like modifier
TAT	Trans-activating transcriptional activator
TdT	Terminal deoxynucleotidyl transferase
TEA	Triethylamine
TEM	Transmission electron microscopy
TF	Transcription factor
TNBC	Triple-negative breast cancer
TP	Transporin
TP53	Tumour protein 53

tRNA	Transfer ribonucleic acid
TUNEL	Terminal deoxynucleotidyl transferase dUTP nick end labelling
VEGF	Vascular endothelial growth factor
VFD	Vortex fluid device
W	Tryptophan

Abstract

Proteins and smaller peptides perform a wide variety of functions within living organisms making them an attractive field of study in biomedical research generally and cancer research in particular. Many proteins serve as cell cycle regulators and gene expression modulators through their interaction with other proteins; the overwhelming majority of molecular interactions in cell biology rely on protein-protein interaction for their mode of action. In the last couple of decades several powerful techniques have been developed that use protein or peptide based ligands to target cancer cells in increasingly more efficient and potent chemotherapies. Despite these advances the fundamental steps in protein engineering have remained largely the same and off-target effects are still common in so called 'gold standard' chemotherapies. This thesis is divided into five chapters. The first chapter will begin by compiling relevant literature covering aspects of current practices in protein engineering, cancer with a particular focus on triple negative breast cancer and DNA damage and the G-quadruplex secondary arrangement of DNA. Chapters Two, Three and Four report the body of my research where I describe a radically new approach to the way we produce proteins of interest that utilizes a shear stress induced refolding mechanism, new ways to deliver therapeutic proteins as well as addressing a need to focus on the physical structure of proteins and oncogene associated proteins. Finally, Chapter Five will offer concluding remarks and suggest possible future directions for the research presented in this thesis.

Currently, recombinant overexpression of proteins in bacterial hosts often results in misfolded or aggregated proteins that isolate into inclusion bodies. Bacterial recombinant overexpression of proteins is a fast way to produce large quantities of biologically active proteins and a crucial step to making biomedically useful proteins commercially viable. However, protein aggregation experienced during production drastically reduces yields and increase the time taken to produce the final active form of a desired protein. Conventional methods to correct misfolded proteins are time intensive and produce large volumes of chemical waste. We have addressed these problems in Chapter 2 with a novel method that implements shear stress to rapidly refold proteins from inclusions while simultaneously reducing bio- waste.

The use of proteins in anticancer treatments has been a steadily growing field of interest in laboratories around the world. Because of the very specific binding nature of proteins they have been suggested as a possible solution to systemic toxicity side effects observed in many chemotherapies. In Chapter 3 we describe the use of synthetic peptide that interferes with the functioning of the Engrailed 1 transcription factor to inhibit cancer cell proliferation in a potently specific manner. Furthermore, we show an enhanced effectiveness of our so called 'iPep' when delivered in a poly(glycidyl methacrylate) nanoparticle formulation and later explore this nanoplatform's ability to act as a co-delivery system in conjunction with other industry standard chemotherapeutic molecules.

The focus of the fouth chapter in this thesis is concerned with understanding the unique four-stranded G-quadruplex DNA structure and its

implications in tumourigenesis. When certain guanine (G) rich sequences occur in DNA they can form what is called a G-quadruplex, a planar assembly of non-Watson-Crick base pairing that are over-represented in regulator genome regions and oncogenes. G-quadruplexes reduce DNA stability which is believed to increase mutagenicity. Identifying these structures has led to a desire to design small binding molecules that could bind to G-quadruplex structures, stabilizing them and down regulate oncogene expression and procancer signalling. To do this we must understand the structure of a binding pocket to then create an analogous binding molecule. A monoclonal single chain antibody, BG4, has been shown to have a high affinity and specificity for the intramolecular G-quadruplex structures but its exact 3-dimensional structure has yet to be solved. We attempted to crystallise recombinant BG4 and in doing so have identified ways to improve yield and certain conditions more prone to sustaining a complete crystal structure of BG4.

The fifth and final chapter of this thesis will briefly summarise the major finding of each of the previous chapters. It will then go on to offer means to improve our understanding of the conclusions drawn from the research presented by way of experimental observations or proposed future work. Collectively, this thesis explores the many important roles proteins play in aberrant cell cycling, in particular, as it relates to cancer and how we can inhibit tumour formation. This research, combined with further studies could drastically reduce the costs associated with production of chemotherapeutics, offers a new, more efficient alternative anticancer drug while also guiding our comprehension of the next wave of protein-DNA based chemotherapeutic ligands.

Acknowledgements

Over the many years that have preceded this moment of completion there have been scores of people that have come together to help me to produce this thesis. I will attempt to acknowledge you all but no doubt I'm sure I will forget the contributions of some of you. For this I apologise in advance and extend an invitation to share a moment in the future to make it up to you.

These years have been my most trying and exciting to date due in no small part to the amazing people with which I have shared them. I must begin by acknowledging my debt of gratitude to my supervisor, **Iyer**. I returned from a year away to find my desk packed up and my laboratory had moved interstate. You immediately approached me with open arms and an offer to join your team. You gave me a home at UWA when I had none and managed to incorporate my skills into research that was interesting and challenging. We've shared tough times and great times as colleagues and friends. You encouraged me when I thought I couldn't go on and I express my deepest thanks to you. I wish you and Nicole a most amazing future together.

Colin Raston. You were where it all began. You saw something in me in my honours year and pushed me to continue study into this PhD. When results didn't make sense and criticism was high you forced me to find the answer why? You gave me the opportunity to travel with science and discover things that have defined my career. I am indebted to you Colin. I hope to carry your optimism with me always in what is to come.

Charlie Bond. This thesis could not have been finished without you. Your tutelage in a field of science that I was unfamiliar with has expanded my learning and instilled a willingness to take on the unknown. Always calm and encouraging, you were exactly what I needed in the closing years of this thesis. Thank you. I must also express my gratitude to all the members of the Bond lab for their patience and assistance while teaching me the science and art of crystallography. In particular **Amanda Blythe**, who graciously took the time to train me when I joined the Bond lab. You remain a fount of knowledge and made my transition into a new work environment seamless. **Kate Shearston**, your beaming energy made early starts and late nights at work a joy. Your editing and review of my writing was a gift to me. Please let me repay the favour to you as soon as possible. I can change nappies and babysit at the drop of hat. Go out and enjoy yourself on me.

Pilar Blancafort. I was fascinated by your work the moment I heard about it. Allowing me to work in your team alongside **Anabel** provided me a front row seat to developing technologies in science. The two of you drove me to become a more effective scientist. Thankyou.

Gregory Weiss. You opened up your laboratory to me and in short time created a home away from home for me. You introduced to me people I will never forget and

call lifelong friends, **Tom**, **Josh**, **Mariam**, **Kaitlin**, **Luz**, **Tivoli** and many more. The work we did will always be made even more special by my memories of the people of California. Thank you to everyone at UCI and the Weiss Lab.

To all the past and present member of Iyer's lab I would like to thank you for creating such a warm and diverse work place to come to each day. My extended time at the lab has meant I have seen many of you come and go throughout the years and rather than attempt to list you and risk forgetting someone I am forced to acknowledge you all at once. The lab has always provided me with witty back and forth, challenged me professionally and lifted me out of the lowest times. I thank you all.

I would like to thank all the funding bodies and organisations that supported me throughout my PhD and travels; the Australian Government for the Australian Postgraduate Award scholarship and The University of Western Australia for the top-up scholarship. I would also thank the funding bodies that provided me with the money to conduct the large body of work that made up my PhD; the Australian Research Council and National Health and Medical Research Council.

Finally, I cannot thank my friends and family enough for your unwavering support over the past years. You have listened to my gripes and given me nothing but encouragement to see it though. All the residents of Colbuirg, Thomas Street and Nanson Street who have watched me come and go at all hours only to share a drink and a sympathetic ear at the end of the day. **Mum and Dad**, you, more than anyone know how big this is to me as you've been there through it all. You've seen behind the curtain and yet still see me as the great and powerful Oz. When things were great it has been you I've wanted to share them with because it was you who put me there. When things were tough I could always look to you for inspiration but most of all you were simply there for me. There is no greater mark for parenting than that.

Statement of candidate contribution

This thesis contains published work and work prepared for publication, some of which has been co-authored. The bibliography details of work and author contributions are outlined below.

Chapter 1. Introdcution and literature review.

Callum Ormonde (CO) reviewed literature and wrote the chapter. Contribution to thesis chapter by CO: 100%

Chapter 2. Shear-stress-mediated refolding of proteins from aggregates and inclusion bodies.

CO wrote the chapter. CO and Tom Yuan performed all of the experiments. These include expression and purification of proteins, ELISA binding assays, VFD shear experimentation and circular dichroism. Stephan Kudlacek and Sameeran Kunche conducted fluid modelling. Joshua Smith and William Brown contributed to protein expression with help from Kaitlin Pugliese, Tivoli Olsen and Mariam Iftikhar. Colin Raston and Gregory Weiss supervised the work. Contribution to thesis chapter by CO: 70%

Chapter 3. Sensitising basal-like breast cancer to chemotherapy using nanoparticles conjugated with interference peptide.

CO wrote the chapter. CO prepared and characterised nanoparticles with contribution by Diwei Ho (DH). Anabel Sorolla (AS) and CO tested nanoparticles *in vitro*. AS and Edina Wang contributed to *in vivo* testing. DH and Cameron Evans contributed to discussion. Rabab Rashwan and Ruhani Singh provided technical assistance. Swaminathan Iyer and Pilar Blancafort supervised the work. Contribution to thesis chapter by CO: 60%

Chapter 4. Recombinant expression and crystallisation of G-quadruplex binding BG4 antibody.

CO wrote the chapter. All experiments performed by CO. Amanda Blythe and Kate Shearston contributed to recombinant expression of BG4 protein as well as chapter discussion. Swaminathan Iyer and Charlie Bond supervised the work. Contribution to thesis chapter by CO: 90%

Chapter 5. Conclusion and future work
CO wrote the chapter. Contribution to thesis chapter by CO: 100%

Publications, Presentations and Awards

- 2016 – Sorolla A, Ho D, Wang E, Evans CW, Ormonde CF, Rashwan R, Singh R, Iyer KS, Blancafort P. **Sensitizing basal-like breast cancer to chemotherapy using nanoparticles conjugated with interference peptide.** Nanoscale. 2016;8(17):9343-53.

- 2015 – Yuan TZ, Ormonde CF, Kudlacek ST, Kunche S, Smith JN, Brown WA, Pugliese KM, Olsen TJ, Iftikhar M, Raston CL, Weiss GA. **Shear-Stress-Mediated Refolding of Proteins from Aggregates and Inclusion Bodies.** ChemBioChem. 2015 Feb 9;16(3):393-6

- 2015 - "Unboiling an egg" at TEDxPeth 2015

- 2015 - Ig Nobel Prize for Chemistry for "Inventing a chemical recipe to partially un-boil and egg"

- 2014 – Poster Presentation " Anionic multifunctional polymer nanoparticle delivery of oncogene transcription factors" at ANN Early Career Researcher Symposium, UTS (Sydney)

Callum F G Ormonde — Candidate

K Swaminathan Iyer — Co-ordinating

CHAPTER 1

Introduction and literature review

When considering their chemical and physiological behaviour, proteins are by far the most structurally and functionally sophisticated of all biological molecules. This is not surprising when we realise proteins constitute most of a cell's dry mass. Proteins are necessary for a wide range of cellular functions, including both chemical catalysis and structural integrity. For example, enzymatic proteins provide chemical catalysis and membrane proteins form channels and pumps to regulate cellular import and export. Proteins may serve as signal integrators that relay signal cascades, or even act as moving parts for tiny molecular machinery, such as kinesin or topoisomerase. Proteins can even form highly specialised structures, like antibodies or hormones, which are biochemically active throughout entire body systems. It comes as no surprise then that the characterisation and use of proteins in modern science spans all disciplines of research.

The observed functional diversity among proteins is due to both innate protein structure, as well as environmental factors. The amino acid sequence of a protein molecule determines its three-dimensional conformation, as noncovalent interactions between different regions of a polypeptide chain stabilise its folded structure. Amino acids possessing hydrophobic side chains tend to cluster within the interior of the molecule, while localised hydrogen-bond interactions between

neighbouring peptide bonds give rise to secondary α helices and β sheets. Larger tertiary globular structures, known as domains, containing ~40-350 amino acids also direct behaviour; but this does not mean large proteins are rigid materials. They contain precisely engineered moving parts whose mechanical actions couple with a specific chemical event. The coupling of chemistry and movement is what gives proteins their extraordinary capacity to carry out cellular activities. Protein binding interactions display great specificity, in that a protein's active site often binds a very few number of molecules. The ability for a protein to selectively bind a ligand with high affinity is dependent on the formation of numerous weak forces, including noncovalent bonds, hydrogen-bonds, electrostatic attraction and van der Waals' forces, in addition to hydrophobic interactions. This type of bonding means a single protein can participate in multiple reactions without being consumed in the process.

For the above reasons, proteins are an incredibly attractive group of molecules to study. In this chapter I will review current literature surrounding the usefulness of proteins in modern biomedical sciences. First, the methodologies used to engineer and harvest proteins will be discussed. Secondly, I will discuss the application of proteins as therapeutic molecules against cancer cells and tumours, with particular focus placed on triple negative breast cancers. Thirdly, the important roles that protein-DNA interactions play in newly emerging cancer research will be addressed in the context of unique G quadruplex structures. Finally,

this chapter will conclude with a summary of the ideas and formulation of the research aims of this thesis.

1.1 Laboratory production of protein

The isolation and purification of a single desired protein from cells is a difficult process. Unlike genomics, there is no standard protocol for protein expression. A common approach to proteomic analysis incorporates five essential steps: 1) biological extraction; 2) separation, from non-protein elements like lipids and nucleic acids; 3) precipitation, to separate crude extracts from bulk protein; 4) chromatography, either by ion-exchange, hydrophobic interactions or size fractionation by which the target protein is removed from the bulk protein; 5) tailored use of a precise "affinity" matrix to yield large amounts of high purity target protein. This is achieved by exploiting properties unique to the target protein during each of the five phases. This has led to protein purification to be referred to as "more of an art than a science".

There are three common approaches to protein expression. These are recombinant overexpression, cell-free expression, and artificial bio-conjugation. As each technique possesses its own inherent advantages and disadvantages, the choice of expression system will affect purity, yield volume, and time costs, more than any other step in the process of harvesting proteins[1]. The most widely used technique is recombinant protein expression. This technique employs living cells as

factories to build protein constructs based on supplemented genetic templates. Synthetic DNA vectors are transfected into cells where they are transcribed and translated, producing large volumes of the desired protein. High yields, low time costs, and ease of scalability, make this methodology the most widely spread of all protein expression systems[2,3].

Cell-free protein expression is another type of *in vitro* protein synthesis. Here, translation-compatible whole cell extracts including RNA polymerase, regulatory protein factors, ribosomes, transcription factors, and transfer RNA (tRNA) are supplemented with specific gene nucleotides and cofactors to synthesise small amounts of proteins within a few hours[4-6]. Although yields are considerably smaller, this method allows rapid synthesis in conditions free of proteolytic degradation while enabling easy protein labelling without the need of cell cultures[7,8]. Finally, scientists can artificially synthesise proteins via chemical crosslinking reactions. Often described as bio-conjugates, proteins are covalently bonded to each other or modified via labelling at specific sites to produce small volumes of highly pure, short, and unnatural proteins[9,10]. This is a prohibitively expensive process that is usually restricted to peptides or proteins that are toxic to express in biological systems. Alternatively, bio-conjugation is also used for labelling reactions with significant degrees of difficulty associated with site specificity[11].

With low yields and high costs plaguing the cell-free and crosslinking methodologies, researchers routinely turn to recombinant protein expression

systems to express proteins of interest. Under individually optimised conditions that suit each protein of interest, the relative purity and rapid production of larger volumes created by recombinant protein expression make it the only viable solution to mass production of affordable protein therapies.

1.1.1 Recombinant protein expression

Industrial scalability is one of the major hurdles that drug designers must overcome if their drug is to become widely available. This is the major advantage of recombinant protein expression systems. The advent of recombinant DNA technology allowed scientists to effectively express desired genes inside foreign cells. Today, recombinant proteins produced for industrial, agricultural, environmental, and pharmaceutical purposes, represent a $160 billion global market[12]. Recombinant protein expression was hailed as the solution to sourcing rare and expensive proteins required for modern scientific research. As protein recombination became standardised, researchers began exploring the way in which the choice of host cell influenced both the quality and quantity of protein produced.

Recombinant protein expression in mammalian cells yields biologically active protein bestowed with all its necessary posttranslational modifications. It must be noted, however, that this method is synonymous with long cultivation times, low volumetric yields, and requires expensive bioreactors and consumables cost[12,13] . In contrast, bacterial cultivation is faster, less expensive, and yields high

cell densities associated with higher protein production rates. Unfortunately, the recombinant protein produced by bacteria such as *Escherichia coli,* frequently yields inactive protein that collects as aggregates called inclusion bodies, which must be further processed to be useful. All of these properties have a great impact on production costs. Each protein must go through thorough optimisation to determine which avenue is most tailored to the highest volumetric yield. With speed and volumetric yield being the biggest draw to recombinant expression and bacterial systems being the most effective host to achieve these features, it is the most frequently used expression system in laboratories.

1.1.1.1 Bacterial recombinant protein expression

Production of an inactive protein isolated in inclusion bodies can be a considerable drawback of bacterial recombinant expression. The recovery of proteins is laborious and expensive; however, when recovery of correctly folded protein is possible, it is often preferable over other means of harvesting proteins. Protein isolation into inclusion bodies does present some advantages, such as a high degree of purity of target protein aggregates and an increased protection against proteolytic degradation. Until recently, inclusion bodies were thought to be inert to *in vivo* dissolution. In 2001, Carrió and Villaverde described a process to successfully overturn inclusion body protein inactivity[14]. The ability to recover protein activity caused a shift towards *E.coli* bacterial systems, and away from expensive eukaryotic expression systems. This meant that isolated recombinant proteins could regain function through robust refolding techniques: direct dilution, dialysis, and size

exclusion chromatography, are just a few examples of such methods[15-17]. Despite a wide variety of techniques that exist to recover inclusion body protein, problems still persist. These methods commonly require large volumes of refolding media (typically 1-10 litres per mg of desired protein) which prevents up-scaling and non-specific separation (or poor separation) from different folding intermediates limiting sample purity[12].

1.1.1.2 Protein refolding: Industry versus nature

The recovery of inclusion bodies generally results in a soluble protein lacking its native conformation, which affects or even prevents action entirely. Once solubilised, proteins must then undergo further processing in conditions that favour formation of native conformation. However, this is not always possible as energy barriers may be too great to allow intermediate structures to return to native state. These limitations reflect the differences between laboratory and natural means of folding and refolding proteins. Natural protein folding is driven by intramolecular interactions between amino acids, while hydrophobic intermolecular interactions are responsible for aggregates. This difference in adhesive vs cohesive forces is what drives protein folding intermediates as amino acids have not been allowed to bury hydrophobic moieties as they would naturally[18]. Figure 1-1 provides a description of the comparative pathways that new translated proteins can follow while inside the cell as they transition through intermediates to their native conformation.

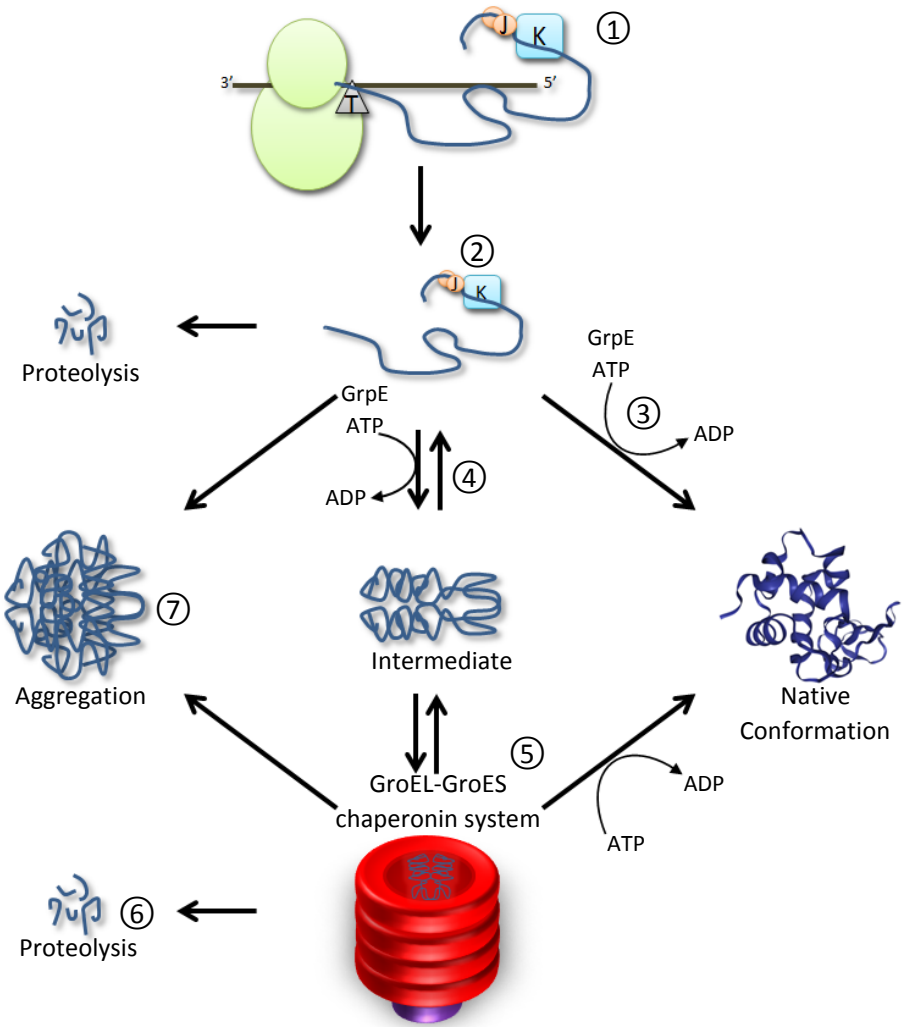


Figure 1-1. Schematic representation of protein refolding in E.coli. Displayed above are the various progressions a newly translated protein can cycle through during transition from an intermediate conformation to native conformation. (1) Newly translated nascent polypeptide emerging from the ribosome is met by local DnaJ (J) chaperone molecule. This binding encourages polypeptide chain recognition by DnaK (K) co-chaperone. (2) DnaK-DnaJ form a complex that shields hydrophobic sequences in the polypeptide preventing undesired intermediate interactions while also critically amplifying ATP conversion in co-chaperones[19]. This is important as binding and release of protein substrates occurs in ATP-dependent manner. (3) Further binding of GrpE nucleotide exchange factor may force structural changes into native conformation in one reaction. Alternately, multiple cycles of GrpE release and binding may be required (4). This is On Pathway protein refolding cascade. Should GrpE cycling fail correct folding a secondary Off Pathway exists through the GroEL-GroES chaperonin system (shown in red).(5) The barrel shaped GroEL-GroES complex further isolates the protein folding environment and facilitates isomerization of compacted proteins trapped in intermediate conformations. Protein released from the GroEL-GroES chaperonin may still be unable to achieve native conformation. In these cases the protein enters the proteolysis pathway (6) for cellular recycling or an aggregate(7) formation is released that will later undergo proteolysis[20].

In nature, cells have evolved molecular machinery to assist protein folding and prevent hydrophobic conformational pitfalls from forming: otherwise known as chaperone proteins. Chaperones are a conserved feature among all phylogenetic kingdoms. An example of one such class of chaperone molecules are chaperonins (particularly GroEL-GroES in *E. coli*). Chaperonins are able to reverse protein aggregation, and refold proteins by lowering the thermodynamic energy required[21]. Measurements of the large activation energy of unfolding actin protein models have recorded free energy change occurring at approximately 80 KJ mol^{-1}. This sort of *in vivo* chaperone assisted refolding is essential for any protein greater than 100 residues in length. When GroEL binds its substrate protein, GroES, ATP hydrolysis is triggered leading to unfolding of misfolded protein[20]. This is followed by ATP-dependent binding of the GroES complex and target protein refolding inside a chaperonin basin[21].

Two critically important features must be noticed in the mechanism of the GroEL-GroES chaperonin system – protein unfolding happens via mechanical means and the target protein is shielded during transitioning through partially folded intermediate conformations. These two steps are important because they are two primary processes that laboratory science has failed to recreate when devising ways to improve protein folding following recombinant expression. Laboratory science can direct folding in protein structures using thermal energy, chemical energy, and even physical energy when using pressure, but attempts at creating a feasible way to redirect folding with mechanical means is too expensive for routine applicability.

If refolding proves too difficult, it is generally recommended to return to expression in a mammalian cell system. Considering the potential advantages that bacterial cell lines offer with respect to total volumetric yield per dollar, it is no surprise that most methodologies would rather focus on optimising bacterial vectors and deal with recovery troubleshooting later. For this reason, only physical and chemical techniques that seek to improve protein yields in bacterial cell lines will be discussed from here on.

1.1.1.3 Important considerations during protein recovery

The most important physical variable shaping protein refolding yields is temperature[22,23]. Temperature effects protein folding in two ways – speed of folding and tendency to aggregate during intermediate folding states. Each protein has a limited temperature range that sustains a stable structure. Low temperatures are generally agreed to support productive folding and suppress hydrophobic aggregation. The slow total kinetic energy associated with low temperatures also decreases the folding rate, subsequently increasing the time needed for renaturation[23]. For this reason, studies focusing on altering temperatures have little universal application or scope for industrial up-scaling.

Altering pressure can also be used as an important variable to exploit larger yields. Robinson and colleagues showed that once disassembled, protein monomers can retain native-like structure when exposed to hydrostatic pressure up

to 5k bar. They also showed that high pressure, to a limit of 3k bar, were capable of disrupting oligomeric structures[24] and even dissolve inclusion body aggregates[25,26]. Together, these findings recognise critical pressure ranges that can be harnessed to specifically disrupt intermolecular forces without undesired insult to intramolecular forces.

The aforementioned techniques disrupt intermolecular forces to release monomers and restore protein nativity and solubility. However, there are alternate approaches to solving the same problem. The addition of solubilising tags to proteins is a common means to promote soluble expression when using bacterial vectors. Popular examples of solubilising tags include glutathione s-transferase (GST), maltose binding protein (MBP), small ubiquitin-like modifier (SUMO), and N-utilizing substance A (NusA)[27]. Expressions using these tags do improve solubilisation in fused protein targets, but can still result in reduced overall protein yield and the added complication of potentially needing to remove the tag afterwards.

Too often the benefits of the above recovery strategies occur at a cost to the inherent advantages bacterial over-expression offers. Additional post expression processing results in time or cost-efficiencies being nullified. Many investigators will simply ignore solubility issues altogether and instead purposefully drive the protein into the inclusion body in favour of maximising target protein yields. Subsequently, the inclusion body must be solubilised and refolded. Existing

in vitro strategies for protein refolding can do this by direct dilution into refolding buffers,[15,28] dialysis to slowly remove denaturants[16,22] and size exclusion chromatography (SEC) or hydrophobic interaction chromatography (HIC)[18,29,30]. Each of these methodologies share a common approach – the target protein is isolated as amorphous aggregates or as amyloid-like fibrils that can be unfolded from their current kinetically trapped state into a higher energy transition state. In doing so, the protein can potentially fold into its desired proper native state restoring functionality for future application (Figure 1-2)[31].

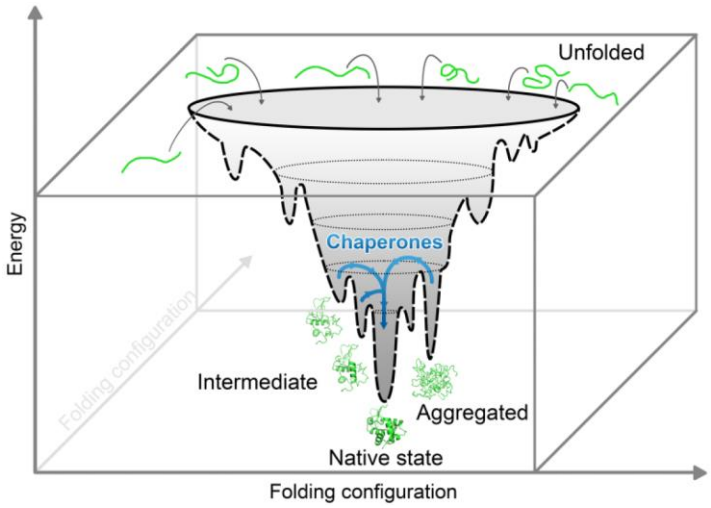


Figure 1-2. Protein folding in the folding funnel model. At the top of the schematic are unfolded proteins in an energy landscape rich in free-energy. Unfolded proteins follow the free energy surface in the irregular funnel shape. When partially folded proteins fall into the lower energy crevasses they become kinetically trapped and form protein aggregates[31,32]. Molecular chaperones in prokaryotic and eukaryotic cells can return misfolded and aggregated molecules to a higher free-energy state. This allows the protein to refold to reach the native state conformation[33-35]. Figure adapted from[31].

1.1.2 Protein unfolding by applied shear stresses

Earlier I discussed how cells use physical energy to unfold protein aggregates with GroEL chaperone proteins, and how high hydrostatic pressure can be used to alter protein secondary structures *in vitro*. Of particular importance to this thesis is the capacity for high shear stresses to denature or alter protein secondary and tertiary structures. In chapter two of this thesis I investigate the controlled application of these forces by a unique table top device to solubilise inclusion bodies created during recombinant protein expression. Shearing forces are commonly observed as a by-product during many biotechnologies such as centrifugation, fractionation, and cell lysis[36-38]. Shear forces exist within a liquid medium as fluids move along a stationary boundary with respect to the liquid. This is represented simplistically in Figure 1-3.

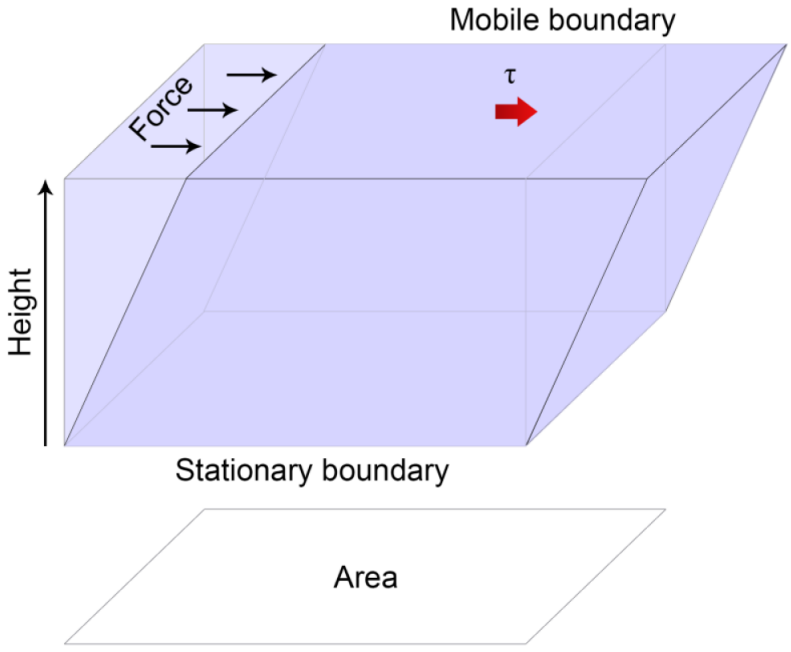


Figure 1-3. Three dimensional representation of a section of fluid under shear stress (τ). The liquid in this system is moving to the right, relative to the stationary boundary (bottom of image). When the velocity gradient from the stationary boundary to the mobile boundary is linear, shear stress is defined by $\tau = \frac{F}{A}$, where A is the area that force F is applied.

Previous shear force investigations using insulin[39] and recombinant human growth hormone (rhGH)[40] have been able to quantify the shear stress levels required to unfolded and aggregate these proteins. Hill *et al*. have additionally reported high shear stresses capable of inducing amyloid formation in β-lactoglobulin[41]. Likewise, the controlled use of high hydrostatic pressure has been demonstrated to both denature proteins and refold misfolded substrates. When this force is fine-tuned with respect to the amount and duration of shear stress, it can encourage unfolding and proper refolding to native state without the need for complicated pressurized systems.

1.1.2.1 Application of the Vortex Fluid Device

While at The University of Western Australia, Professor Colin Raston developed a low-cost benchtop device called the Vortex Fluidic Device (VFD). VFD uses micro-thin film fluids to apply controlled levels of shear stress in both a confined and continuous flow environment (Figure 1-4). The VFD was initially developed to exfoliate graphite[42] and hexagonal boron-nitride[43], however, the applied shear forces within the VFD have been shown to overcome the van der Waals interactions between carbon layers, without damaging the monolayers produced by processing. Subsequent studies have demonstrated the VFDs ability to disassemble hydrogen-bonded *p*-phosphonic acid calix[5]arene via reversible dissociation of non-covalent bonds[44]. In Chapter 2 of this thesis we characterise the VFDs application of shear

stress to successfully unfold protein aggregates as a means to address recovery issues and improve bacterial expression of large proteins.

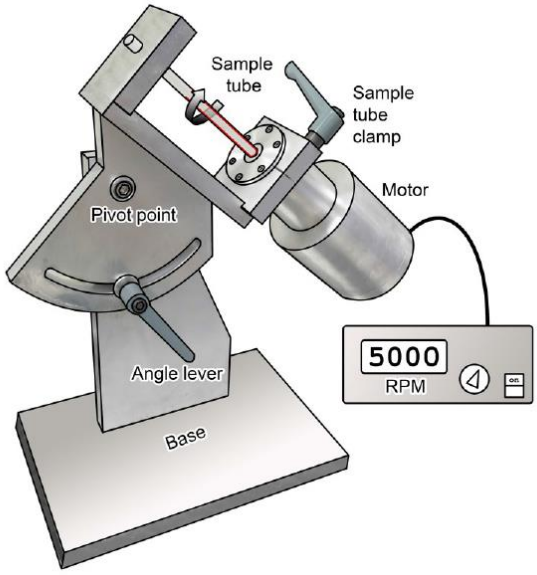


Figure 1-4. Illustrative representation of the Vortex Fluid Device (VFD). A 16 cm long glass tube with diameter 10 mm is inclined at a 45° angle. The fluid filled tube is spun at high speeds, forcing the sample the spread out along the walls of the glass into microfluidic films (0.2 – 1.2 mm for volumes of 1 ml) indicated by the red lines. Films thickness is dependent upon rotation speed and angle of incline[45].

1.2 Protein-protein interactions and human disease therapies

Alongside examining the ways in which we produce important proteins of interest, this thesis also seeks to explore novel protein therapies for the treatment of cancer. Before delving into this discussion, it would be pertinent to review some of the many roles proteins in particular dysfunctional proteins errors can contribute to human disease.

Aberrant protein interactions are frequently the root of human diseases. Specific protein interactions have been identified across nearly all disease systems,

including the papilloma virus, cervical cancer, bacterial infections, and neurodegenerative disease. By identifying where and how protein failures produce disease, we can design therapeutic approaches for correcting, and hopefully curing, human diseases as they arise.

For example, papilloma virus infections initially cause benign epithelial lesions that can result in cervical cancer[46,47]. The virus relies on viral E2 proteins to successfully segregate and replicate their episome into daughter cell nuclei. This is achieved by a C-terminal viral DNA binding region[48] and an N-terminal transactivation domain[49]. E2 proteins target cellular Brd4 proteins in host cells to attach to mitotic chromosomes and ensure the virus plasmid is passed onto daughter cells. Brannon *et al*. realised this and overexpressed the C-terminal E2 binding region of the Brd4 protein, blocking E2 binding and viral transmission to daughter cells[50]. Through the characterisation of protein-protein interactions, a promising target for therapeutics against papilloma viral infections and cervical cancer has been developed.

Protein-protein interactions can also be responsible for bacterial pathogen infections. For instance, Gram-positive bacteria hijack a host cell's internalisation machinery via surface fibronectin-binding proteins (FnBPs). This occurs when intrinsically disordered repeat sequences in the FnBPs bind fibronectin and adopt a β-sheet conformation within two of the fibronectin domains forming a 'tandem β-zipper' that facilitates bacterial invasion of the host cell[51]. The tandem β-zipper

protein interaction has been observed in many bacterial species including those that lead to endocarditis, scarlet fever and Lyme disease[51-53]. This suggests a common mechanism of bacterial invasion and offers a broad application for therapeutics that target or inhibit zipper formation.

Huntington's disease is one of several neurodegenerative disorders whose precise etiology is not entirely known. The disease is characterised by N-terminal expansion of ubiquitous huntingtin (htt) protein. This expansion contains additional glutamine residues in polyQ regions of proteins, prompting it to be referred to as a polyQ disease. This aberrant expansion causes insoluble aggregates to deposit in the brain and htt to self-assemble into β-sheet conformations similar to amyloid[54]. While the protein aggregates themselves are not cytotoxic, they can sequester healthy htt proteins depriving cells of their critical function relating to transcription regulation, anti-apoptotic effects and neuronal development[55]. This sequestering effect was thought to be the cause of the Huntington's disease, until the mid-1990s when the specific polyQ-htt associated protein (HAP-1) was identified as a binding partner for aggregated htt. This finding suggested that neuropathology may be due to interactions between polyQ-htt and other proteins and not simply protein deficiency[56].

These instances of protein involvement in human disease highlight the importance of protein functioning to the cell. Studying protein interactions is important not only for our understanding of the cause of disease, but also when designing therapies to combat them in the future. High binding affinity for very

select targets make proteins widely effective drugs. These inherent properties address precise problems that are currently commonplace in many cancer therapies and may offer solutions to many of the side effects that continue to trouble cancer treatment.

1.2.1 Protein's role in cell cycling

When considering the many facets of disease and cancer progression, no characteristic is more defining than the state of unregulated cell cycling and division. No single error in the cell cycling process will be wholly responsible for cancer development, however. Normal cell division is controlled by proteins at many separate levels, and a sum of errors are accrued over time that lead to a cell becoming cancerous. There are many different instances in which protein failure can occur during the cell cycle that can lead to disease. This is why treating cancers is so difficult.

A simple model of cell division is to divide it into two stages of activity: mitosis and interphase. Mitosis is where the cell is actively engaged in nuclear division, and can be further broken down into stages of prophase, metaphase, anaphase, and telophase. The cell spends most of its time in interphase, which includes the DNA replication required before cell division. Interphase can also be viewed in sequential sub-phases namely, G_1, S, and G_2 phases. The S phase of interphase importantly includes the process of DNA replication which if unsuccessful or incomplete can lead to disease including cancer. The transition

through each phase of the cell cycle occurs in orderly manner under the strict regulation of cellular proteins. This process, and the important proteins involved, is described in Figure 1-5. In this regard, complex interrelated protein functions ensure cell viability and mortality, and are often the cause of diseased states.

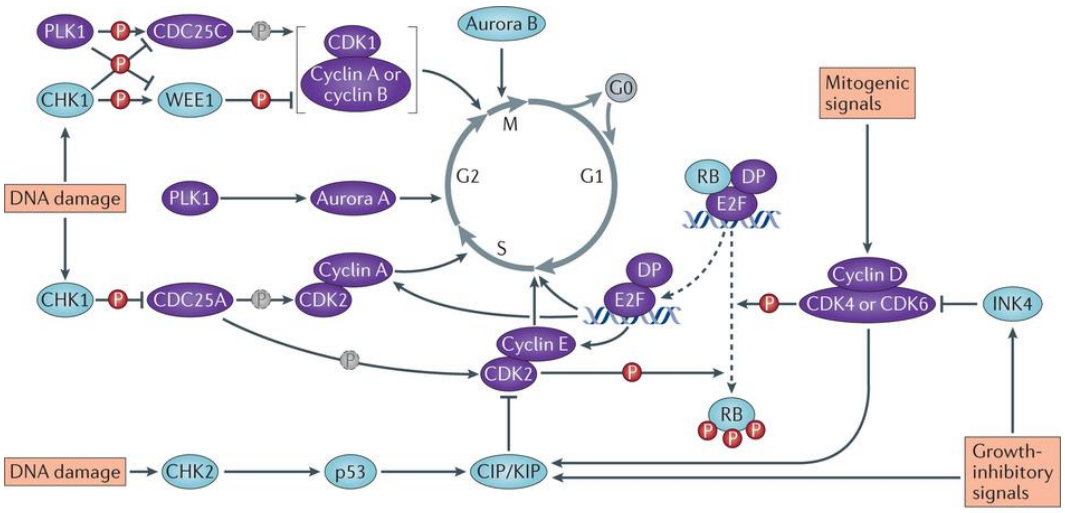


Figure 1-5. Cell cycle progression and major regulatory proteins. Mitogenic signals activate complexes of cyclins and cyclin-dependent kinases (CDKs) promoting progression from G1 phase into S phase. Progression occurs by targeted phosphorylation (P), shown here in the retinoblastoma protein (RB). Hyperphosphorylation of targets attenuates growth-suppressive properties and leads to activation of transcription factors. Upregulating CDK inhibitors (INK4 and CIP/KIP families) cause growth-inhibitory signals to antagonize G1–S progression. Cyclin–CDK complexes also control progression through S phase and from G2 phase into mitosis (M phase). Various other proteins, such as Polo-like kinase 1 (PLK1) and Aurora kinases (Aurora A and Aurora B) also take part in regulating cell cycle progression. Some cells exit their cell cycle and enter a temporary or permanent cell cycle arrest (G0 phase). When DNA damage is sensed specialized proteins trigger cell cycle arrest. If cell is in G1 phase this occurs via checkpoint kinase 2 (CHK2) and p53. Cells in S or G2 phase trigger arrest via CHK1. Positive regulators denoted by purple ovals. Blue ovals denote negative regulators. P in dashed circle indicates dephosphorylation[57]. Reprinted from Nature Reviews Cancer, 17, Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. 2017 Feb;17(2):93., with permission from Nature Publishing Group.

1.2.2 Understanding cancer

To understand where protein therapies can intervene in cancer cell survival, we must first understand the many factors that contribute to a cell becoming

cancerous. The term cancer describes a group of diseases that derive from single aberrant cells that are capable of affecting any region of the body. At its core, cancer is the result of dynamic changes in the genome of a cell and its progeny. This group of diseases can affect parenchymal and stromal cells alike and is characterized by six distinct biological 'hallmarks of cancer'. These distinct and complimentary hallmarks include: sustained proliferative signalling, growth suppression evasion, cell death resistance, replicative immortality, self-induced angiogenesis, metastasis and invasion.

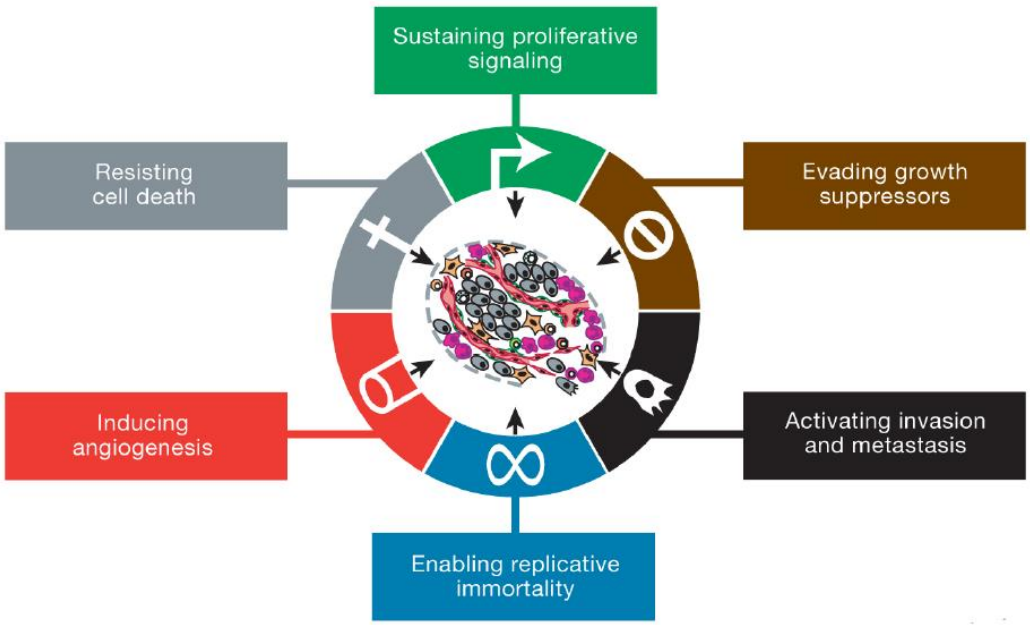


Figure 1-6. The Hallmarks of Cancer as described by Hanahan and Weinberg. The illustration presents the original six hallmark capabilities that cancer cells must acquire to successfully continue their pathogenesis[58]. Reprinted from Cell, 144/5, Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. 2011 Mar 4;144(5):646-74., with permission from Elsevier

When these capabilities are obtained collectively in a well-recognised multistep development, the resulting pathology is called cancer. Imperative to this definition is the notion that a normal cell can evolve progressively towards a neoplastic state. Through successive gains of these cell capabilities, human tumour

pathogenesis occurs leading to malignant cancers. It is important to note that the neoplastic tumours that arise from sustained cell proliferation are more than circumscribed cell masses. They are complex heterogeneous tissue with multiple marked cell types including tumour-associated stroma and importantly cancer stem cells contributing to the clouding of a clear target.

To understand what distinguishes cancer cells from normal cells is to recognise where disease begins and healthy tissue ends. With the seemingly incalculable number of cancer types, it could be argued that searching for the answer to treatment through the diseases' aetiology would only serve to add further layers of complexity. Instead, by recognising the many ways in which cancer cells correct regular cell behaviour to display the above hallmarks, we devise a bottom-up approach. This way, therapeutics could be applied to a multitude of cancers rather than one strain.

1.2.2.1 Breast cancer and triple-negative breast cancer

Breast cancer is the most common cancer affecting women worldwide, and the second most common cancer overall. Breast cancers represent 25% of all cancers in women and 12% of cancers worldwide. Breast cancer is a heterogeneous disease, divisible into variable numbers of subtypes. The complexity in categorising these diseases begins with whether to recognise disease by clinical subtypes or by etiological subtypes. Clinical taxonomy represents grouping breast cancer subtypes with similar presentation, prognosis, and prediction, while etiological subtyping

reflects grouping diseases that share a common set of causes[59]. The clinical system translates to a 'personalised' or 'precision medicine' view to optimise therapeutic management, whereas an etiological approach takes a more parsimonious view of breast cancer. This method of subdivision begins with all breast cancer being derived from one of two main cell types of origin and then creates a hierarchal ordering from disease progression[60,61]. This reporting is more concerned with designing and testing of new tailored drugs than understanding precise origins of disease and so will proceed with the clinical taxonomic system of subtyping breast cancer.

With the incidence of breast cancer so high, a concerted global effort has been mounted to fight the disease, with particular focus directed towards self-examination and early recognition. These highly successful public awareness campaigns have meant most of the estimated 15,270 Australian women annually diagnosed with breast cancer will survive their diagnosis. Early identification and disease education programs have meant 89% of women will go on to meet the important five-year survival milestone if given access to treatment.

Triple-negative breast cancers (TNBC) are a subtype of hyper aggressive malignancies found in ~15% of breast cancer cases. TNBC is said to be a diagnosis by exclusion characterised by a lacking of surface estrogen receptors (ER), progesterone receptors (PgR), and human epidermal growth factor receptor-2 (HER2). This morphology means TNBC fail to respond to conventional endocrine

and anti-HER2 therapies that are routinely the first course of treatment in breast cancer chemotherapy[62,63]. Typically TNBCs are high-grade tumours[64] with approximately 70% of molecular profiling identifying a basal-like profile that is synonymous with malignancy and targeting problems in many cancers[65].

The large proportion of TNBC that belong to the basal-like breast cancer (BLBC) subtype are characterised by p53 and BRCA1 mutations[66]. These are tumour suppressing genes that control cell cycling and DNA repair. When considered together these characteristics of TNBCs create a very diverse and complex disease profile. This complexity has led to some to consider TNBC to be a heterogeneous entity, not a single disease but rather a heading for additional tumour types[67]. The simple lack of consensus when it comes to defining TNBCs has no doubt contributed to the maze of solutions currently being proposed to fight the disease's varied features.

1.2.2.2 Feature of triple-negative and basal-like breast cancer

From a practical perspective three of the five prominent features needed to refer to a tumour as a TNBC are in reality observable absences in cells. This has directed most research to focus on the two gene mutations most synonymous with TNBC – BRCA1 and p53 mutations - as the starting point when designing discrete and objective drugs.

Cancers with BRCA1 germline mutations will often simultaneously present as triple-negative breast cancers as well as basal-like breast cancers. The term

basal-like describes a cancer cell subtype expressing genes associated with the normal basal cell; a non-luminal myoepithelial layer of the breast duct and lobules[68,69]. Gene mutations that produce basal-like phenotypes include basal cytokeratins 5, 14 and 17, endothelial growth factor receptor (EGFR), HER1 along with common myoepithelial markers like caveolins (CAV) 1 and 2, c-kit, and P-cadherin, and are frequently associated with poor-prognosis in breast cancers[70-72]. Most BLBCs are TNBCs in that they are routinely hormone-receptor negative yet, additional overlapping features are also observed. These include high grade (3+), ductal carcinomas, high mitotic count, medullary-like histology, EGFR expression, TP53 mutations, and HER2 negativity. This pattern of development is similarly found in tumours possessing BRCA1 mutations without being wholly TNBC or BLCBs leading to the pattern being termed 'BRCAness'[73]. The BRAC1 gene encodes at least three known proteins: p220, Δ11b, and IRIS[74]. Little is known about the function of Δ11b and IRIS other than an implied involvement in DNA replication noticeable in certain oncoprotein behaviours[75,76]. The import functions of p220 are more widely recognised as tumour suppression proteins in breast and ovarian cancers[77-79]. Several studies of mutant variants of p220 have identified a multitude of DNA repair functions within basal-like and TNBC associated genes. Cells expressing mutant p220 fail to repair collapsed replication forks[80,81], retain cross-linking errors[82], and promote new mutagenicity in bases[81].

Shared patterns of gene expression suggests BRCA1 dysfunction most likely plays a role in BLBC and TNBC which has drawn significant research attention within the TNBC community. The intimate association between BLBCs and TNBCs has led

to the two terms being used interchangeably (and a growing acceptance of BRCA1ness by proxy). A schematic of the many features shared by these separate definitions is shown in Figure 1-7.

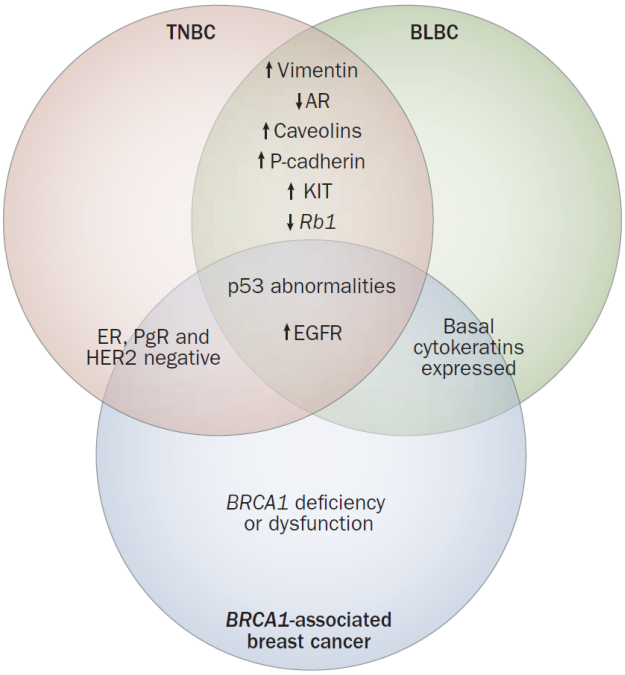


Figure 1-7. Shared features of triple-negative, basal-like and *BRCA1*-associated breast cancers. Abbreviations: AR, androgen receptor, BLBC, basal-like breast cancer; ER, estrogen receptor; PgR, progesterone receptor; TNBC, triple-negative breast cancer[67]. Reprinted from Nature reviews Clinical oncology, 7/12, Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience?. 2010 Dec;7(12):683, with permission from Nature Publication Group.

The multiple cellular pathways affected by the BRCA1 gene has meant substantial efforts to link sporadic BLBCs and BRCA1 dysfunction have been made to create a cohesive disease profile for TNBCs. The therapeutic implications of such a connection would mean these 'undruggable' tumours becoming susceptible to chemotherapy and other DNA repair targeting agents like poly (ADP-ribose)

polymerase (PARP) inhibition[83,84]. Unfortunately, despite considerable efforts, little success has been made towards a deliverable BRCA1 targeted drug.

The tumour suppressor protein p53 would have to be the most studied tumour associated protein. A mutation in p53 been observed in some form in almost every cancer type[85]. An initial error in early reporting led to p53 to be considered the product of oncogene expression for almost 10 years. Now, p53 is recognised as a naturally functioning wild-type p53. This shift occurred through incremental identification of p53 mutations that cause both loss in wild-type function and neomorphic gain of functions that result in tumour development. The reoccurring prevalence of p53 errors across cancers made it a broadly attractive target in many early therapeutic strategies. p53's consistent association with aggressive and metastatic breast cancers has maintained steady interest in the protein despite disappointing results.

TP53 is the most commonly mutated gene among human cancers[86]. Phenotypes vary with differing mutations. Nonsense or frameshift mutations usually produce loss of function[87] leading to reduced tumour suppression and early onset tumours. More common missense mutations, in contrast, frequently result in a single substituted amino acid in DNA-binding regions and produce higher functioning mutant proteins. These gain of function mutations create a distinct mutant protein that functions as a dominant negative inhibitor over any remaining wild-type p53[88]. Mouse models of this kind of p53 mutant create added treatment complications by dampening drug responses designed to restore wild-type p53. Other p53 mutations will prevent DNA binding in target gene promoters only to

have indirect effects on genes through transcription factors interactions occurring with mutants and not wild-type p53[89-91].

The hallmarks of TNBC – enhanced aggressive invasion and metastases coupled with broad drug resistance – are certainly (at least in part) a consequence of gain of function p53 mutant expression. Therapeutic targeting of p53 mutants have yielded varied degrees of success. Since p53 mutants interfere with multiple cellular pathways in such varied ways, their initial attractiveness is being foreshadowed by complexity. Interesting evidence for p53 mutants dampening restorative therapies[88] is only made more puzzling by recent studies that point to retention of wild-type p53 can be deleterious to therapeutic responses in breast cancers[92]. The clear role of p53 in TNBC makes it a tremendously appealing target when designing treatment strategies, however, with new pathway integrations still being discovered it would seem to be another strategy troubled by downstream side effects.

1.2.2.3 Current treatment strategies for TNBC

Patients currently being diagnosed with TNBC face an overall poor prognosis owing primarily to their tendency to relapse with distant aggressive metastases. When compared to other subtypes of breast cancers, TNBCs have a higher prevalence towards diagnosis as an 'interval cancer,' because detection occurs 12 months after an initial negative mammographic screening. TNBC patients initially present with large primary tumours with low lymph-node association that are primarily responsive to non-specific anticancer treatments. A short disease free period is

then followed by rapid progression of distance recurrences. This signature sequence of events describes the common pattern of TNBC disease progression[93]. Furthermore, despite a strong prevalence for distant metastasis, local tumour recurrences are only marginally increased compared to other tumour type progressions. This suggests TNBCs retain radiation sensitivity, a theory consistent with the observation that basal-like and BRCA1-associated tumours are responsive to DNA-damaging agents[84].

The poor patient prognosis currently associated with TNBCs come primarily from their tendency to relapse with distant metastasis. This indicates a distinct need for new effective therapies tailored for the TNBC phenotype. Target receptor deficiency has meant nonspecific chemotherapy is the mainstay of present patient treatments. These chemotherapies are effective in disease remission and trends in developments of these chemotherapies will most likely yield preferential benefits for this subtype of breast cancers[94].

Nonspecific cancer treatments focus on DNA repair and rapid proliferation targeting for a drug's mode of actions. Neoadjuvant studies show TNBC/BLBCs to be particularly responsive to taxane based chemotherapies. It is worth noting that in TNBCs and BLBCs there is higher instance of pathological complete response (pCR) to such chemotherapies when compared with other breast cancer subtypes. Unfortunately, overall survival remains worse in patients who fail to achieve a full pCR[94]. Additionally, TNBCs do not show increase sensitivity to taxanes when observed in a metastatic setting despite high proliferation rates and TP53

mutations. This example of learned drug resistance in TNBC is not an isolated one, contributing to the reputation of being an 'undruggable' breast cancer.

Other studies have suggested cyclophosphamide, methotrexate and 5-fluorouracil - so called CMF regimens to be more effective means of drugging TNBCs[95]. These antimetabolite chemotherapies produce greatest potency in TNBCs. One study found TNBCs to be 14% more responsive to CMFs than HER2+ tumours and 25% more than luminal tumours[96]. These improved CMF responses, however, have been isolated to basal-like TNBCs only, with no benefit observed in non-BLBCs and the data is restricted to neoadjuvant studies[97].

In vitro and *in vivo* data analysis of tumours with BRCA1 pathway dysfunction have shown an increased sensitivity to DNA crosslinking agents such as platinum compounds. In one retrospective study of initial treatments for BRCA1 mutations across breast cancers 83% of patients who received a cisplatin only therapy had pCR while fewer than 25% displayed the same on other regimens[98]. Another study compared therapies across neoadjuvant and advanced disease settings in TNBCs. Their findings showed response rates in TNBCs to platinum-based neoadjuvant chemotherapy were 32% greater than other tumour types (88% versus 51%) and 10% greater in advanced settings (41% versus 31%)[99]. The complexity and uncertainty of TNBCs has meant promising new developments in drug efficiency are often confined to isolated niches within the TNBC family. Drugs potent to one subtype will then yield results equal to, or only marginally better than, nonspecific

therapies for other TNBC types. A gold standard umbrella treatment for all TNBC/BLBCs still eludes scientists today.

Nanoparticle therapeutics are a new treatment modality gaining popularity in cancer as a whole. The unique chemical properties associated with nanoparticles can show enhanced potency, improve targeting and increase cellular uptake via avenues inaccessible by larger conventional drugs[100]. The small scale has meant a shift to short peptide sequences rather than whole proteins and may potentially produce similar advancements in TNBC therapies.

1.2.3 Peptides in nanomedicine

Historically, anticancer pharmaceuticals made from peptides have mostly been used as construct polyvalent vaccines or hormones directed against G-protein couple receptors (GPCRs)[101,102]. This is predominantly owing to their faster clearance and lower affinity when compared to antibodies and or larger ligands. Ongoing research in multiple therapeutic lines of monoclonal antibodies (mAbs) and larger protein ligands continued to return to two major limitations: poor tumour delivery – a product of size hindering passive diffusion across capillary membranes, and dose-limiting toxicity – where nonspecific uptake in the reticuloendothelial system and by the liver results in hepatic and bone marrow cytotoxicity[103,104]. Today these limitations are still challenging progress and mAbs are largely relegated to antiangiogenic family of drugs with luminal targets for mode of action[105,106]. A new wave of thinking now focuses on peptides over larger

macromolecules. By their very nature, being smaller than larger protein macromolecules, peptides do not encounter the two aforementioned blockades in their path to becoming a viable anticancer drug. Noticing this, modern cancer researchers are preferring to address the previously perceived limitations of larger proteins and exploit peptide advantages rather than have to once again answer the inevitable size and toxicity issues of macro-proteins.

The movement to smaller peptide based therapies quickly revealed advantages outweighing their failings. In one instance, researchers used an antibody-mimicking peptide (3kDa) that demonstrated greater tumour penetration and targeting capacity (despite a binding affinity best reported to be one tenth of the parent antibody) which also participated in rapid renal clearance in *in vivo* mice modelling[107]. Together these results could circumvent the liver and bone marrow toxicity associated with traditional mAb-anticancer therapies. The excitement attributed to the advantages of peptide, including potency, low toxicity and specificity (not to be confused with affinity), must come humbly as we also consider their short-comings. Poor stability, digestive susceptibility, and short half-lives, are all hurdles that researchers routinely meet when working with peptides.

In August 2012, the FDA approved Linaclotide (sold as Linzess or Constella) a peptide drug delivered orally. Linaclotide's cysteine rich sequence allow for three disulphide bridges which bestow enough stability to allow it to be taken orally. Despite its intended use in the relief of constipation and irritable bowel syndrome, and not as a cancer antagonist, this does serve as proof that peptide drugs can

complete all phase trials and become viable drugs for human consumption. Moreover, Linaclotide is not a single occurrence. With advances in phage display libraries peptide substrates are commonly incorporated into design strategies for a host of drugs including anticancer.

Developments in the recent decade have made nanoparticles a powerful tool in drug delivery. Their multifunctional nature offers answers to barriers that traditional pharmaceuticals have failed to overcome. The combination of peptide specificity with nanoparticle deliverability as emerging drug platforms has served to strengthen their inherent advantages and alleviates their disadvantages. This new technology is very diverse and this review will focus on peptides as drug carriers and peptides as targeting ligands.

1.2.4 Peptides as nanodrug carriers

Translocation through the plasma membrane of cells is one of the foremost hurdles that many drugs must conquer. Larger drugs and ionized drugs struggle particularly with plasma protein adherence issues[108]. One solution to this problem has been found in the cell-penetrating peptide (CPP). The earliest description of a CPP came in 1988 from the appreciation of the trans-activating transcriptional activator (TAT) of the human immunodeficiency virus 1. Interestingly, the discovery was reported independently by two laboratories in the same year[109,110]. Over half a decade later the third helix of a homeodomain transcription factor isolated from *Drosophila* revealed the antennapedia peptide (Antp) to be a CPP and another example of CPPs

derived from naturally occurring proteins[111]. The usefulness of CPPs meant synthetic or chimeric CPPs soon began to appear. 1998 marked the arrival of transporin (TP), a chimera produced by fusing the galanin neuropeptide with a sequence of 14 amino acids isolated from wasp venom mastoparan[112]. A third group of CPPs exist, the wholly synthetic CPPs of which the polyarginines are the most abundant[113]. Whichever of the three categories one belongs to, CPPs have proven passage across the plasma membrane of various cargos to occur with relative ease. This is true even for cargos with molecular weights considerably larger than their own[114]. Many anti-cancer drugs seek to correct a damaged gene, and must first be up taken by cells unaltered. Covalently decorating nanoparticles with CPPs is a relatively easy process that can ensure internal delivery of a nanoparticle's payload and an attractive reason to include peptide conjugation in TNBC nanodrugs.

1.2.5 Engrailed 1 targeting for peptide technology in TNBC

Promising response by TNBCs to conventional cancer chemotherapies are often hindered by rapidly acquired resistance to treatment, followed quickly by recurrence and metastatic disease[115]. Recent studies have attributed highly selective overexpression of the Engrailed 1 (EN1) transcription factor (TF) to play a significant role in the drug resistance acquired by basal-like TNBCs[116]. Large-scale studies of breast carcinomas have also implicated TFs to be a primary force behind oncogenesis in BLBCs. In particular, aberrant expression of many developmental

homeodomain (HD) containing TFs have been identified as drivers of cancer initiation, recurrence, and resistance.

The EN1 TF is particularly overexpressed in basal-like and triple negative tumours. In human embryology, the two paralogs (members of the homeobox gene family) EN1 and EN2 control central nervous system development, and their expression is largely silenced in mature cells[117]. EN1 is expressed by neuronal progenitor cells to grow and maintain the pool of dopaminergic neurons. Such prosurvival activity is thought to promote resistance to apoptotic insults experienced during cell adulthood[118]. This belief has prompted knockdown study of EN1 using short hairpin RNA (shRNA) which triggered potent and selective cell death (mediated by caspase-3 activated apoptosis) while atypical overexpression in normal cells stimulated survival pathways and conferred resistance to chemotherapeutic agents. Furthermore exogenous overexpression via EN1 cDNA prompted reprogramming in breast epithelial cells towards a long-lived stem cell-like phenotype[116].

Recognising this function and conserved pattern of overexpression has encouraged researchers to find new ways to exploit the EN1 TF as a possible target for treatment of BL and TNBCs. Targeting TFs is not a novel idea but due to their relatively small molecular binding pockets (when compared to other cancer targets like tyrosine kinase receptors) this has proved difficult in practice, leading to a belief that they are “undruggable”. Agents have be developed that target various levels in

the transcriptional pathway which include DNA binding by TFs, protein-protein interaction interrupters and epigenetic alterations[119]. Examples of transcription pathway interrupting drugs have been summarised in Table 1-1.

Table 1-1. Different transcriptional targeting agents and their level of action[119].

Action	*Targeting Agent*	*Example*
Targeting binding of transcription factors to gene-specific promoters	• DNA-binding small molecules • Polyamides • Transcription factor (TF) decoys	• Mithramycin • Antracyclines • N-methylpyrroles • N-methylimidazole • Double strand or hairpin single strand oligodeoxynucleotides
Targeting protein-protein interactions involved in transcription regulation	• Peptide mimetics and stapling peptides • Small molecules targeting protein-protein interactions	• S3-M2001 (STAT3 mediated transcription inhibitor) • SAH-p53-8 (p53 activator) • Myc-MAX • P53-MDM2
Epigenetic interventions in transcription	• DNMT inhibitors • HDAC inhibitors • HMT inhibitors	• 5-azactidine & Decitabine • SAHA & VPA • BIX-01294 & Dot1
Artificial transcription factors for gene-specific transcription regulation	• Zinc finger proteins	• MASPIN • VEGFA

DNMT-DNA methyltransferases: HDAC-Histone deacetylases: HMT-Histone methyltransferases: VEGFA-vascular endothelial growth factor-A

Peptides pose a promising answer to the size restricted binding problem synonymous with TFs. Now that scientists can chemically synthesise peptides for high sequence specificity on a large scale, peptide drugs development is a staple of modern drug development especially in cancer chemotherapy. Past uses for peptides have mainly focused on improving drug delivery efficiency in an effort to reduce side effects rather than using them as a primary mode of action in drug therapy[120]. Peptides used as drug carriers, cancer associated protein-targets and ligand-targeting peptide are just a few examples of the many and varied ways

peptides have been incorporated into new drug designs. Another emerging application for peptides is as so-called interfering peptides that work via competitive-inhibition. Interference peptides disrupt complex formation or dimerisation of oncogene TFs. By displacing an oncogenes native TF, these interference peptides search out binding partners to influence transcription and alter expression. In 2013, the Blancafort laboratory proved that it was possible to design an interference peptide which they called iPep that would effectively inhibit the EN1 oncogene expression in BLBC cells[116]. In chapter three of this thesis we explore the use of this novel small molecule iPep as a potent and highly selective potential therapeutic agent for use against TNBCs.

1.3 Protein-DNA interactions

The heart of this thesis is reporting the myriad roles protein interactions can play in improving cancer disease treatment. Improvements in protein engineering and drug specificity by highly selective protein-protein interactions have been addressed (or more correctly peptide-protein interactions). Now protein-DNA interactions, and how they can be incorporated into new disease therapies, will be investigated. Pertinent to the fourth chapter of this thesis is the interaction of proteins with DNA and the ways these interactions can be altered to affect cell viability on a pre-transcriptional level.

Cancers arise when genomic DNA is damaged and the processes that ensure DNA maintenance by proteins fail. Identifying regulatory sequences in DNA and deciphering the mechanisms by which they act has to date proved a great challenge. Knowledge of the locations within the genome that allow DNA-binding for proteins in the nucleus should identify and define the nature of transcriptional regulatory systems in cells. Simply cataloguing the many oncogenic promoter or enhancer regions of genes does not address the reason behind the inherent mutagenicity of these sequences. Unusual DNA structures occurring in guanine rich sequences, called G-quadruplexes (G4s), have recently been identified to concentrate in gene regulatory regions of the genome associated with oncogenic potential[121]. The non-canonical structure of G4s 'weakens' the gene and is thought to underlie mutagenicity. By targeted stabilisation of these structures using DNA-binding proteins, new chemotherapies can disrupt oncogenes and suppress tumour growth[122-124].

1.3.1 Characterising protein-DNA interactions

Having identified G4s as a target, the next step is to identify site specific features that will allow a DNA-protein binding. Identifying these features can lead to proteins selectively interacting with quadruplex DNA while ignoring native DNA. There are many methodologies that aid in our understanding of protein-DNA interactions. Recently, chromatin immunoprecipitation (ChIP) has become a particularly powerful tool for researchers. Generally, the first step in ChIP is to treat living cells with a cross-linking agent (commonly formaldehyde) to fix proteins to

their DNA substrate inside cells. Chromosomes are then extracted, fragmented, specific sequences purified (usually with aid from specific antibodies), and the protein-DNA association can be determined by fluorescent labelling[125]. Conventional ChIP examines a limited number of binding sites that are subsequently amplified to detect protein-DNA sites throughout millions of gene sequences when combined with microarray technologies[126,127]. ChIP is particularly useful because it directly examines protein-DNA interaction in living cells with the added benefit of results having direct physiological relevance. However, these results are limited somewhat by a reliance on specific antibodies to detect proteins and in many cases such antibodies do not exist.

Alternate protein-DNA characterisation methods using naked DNA fragments *in vitro* fail to observe proteins interacting with compact DNA as it would exist as chromatin with all its distinct properties. Other techniques, such as DNA adenine methyltransferase identification (DamID), induce localised DNA methylation at binding sites inside cells to report location in lieu of highly specific antibodies. Although, Dam-fusion proteins currently identified for use often have distinct properties similar to those of the endogenous protein and critically lack the physiologically relevant data of ChIP[128]. Therefore, to accurately characterise specific locations for protein-DNA binding targets as they truly exist *in vivo* we must concede that specific antibodies currently represent the best avenue when designing novel drug strategies.

1.3.1.1 Structural importance of protein-DNA binding sites

Oncogene mutations occur frequently in promoter and enhancer regions of genes but another overrepresented site for oncogenic mutagenicity is in regions coding for transcription factors. This is unsurprising when considering the required activation of transcription to convert DNA to RNA in the process of expressing a gene. To understand the molecular basis of genomic-binding site for therapeutic antibodies we can look to another class of DNA-binding protein; transcription factors. An ever-growing number of transcription factors are being recognised as oncoproteins when their direct DNA-binding sites undergo mutant alterations. Examples include MYC/MAX, E2F/p130/p107, WNT/TCF4 and p53[129-133]. In characterising the precise binding site structure with the protein-DNA interaction of these transcription factors, researchers can reverse engineer complimentary drugs with high specificity for *in vivo* DNA targeted therapies.

A crucial aspect of gene expression regulation is the recognition of specific DNA sequences by specific proteins. Structural examinations have clearly shown no one-to-one correspondence between amino acids and DNA bases, rather their interactions are broadly distributed in space. Simply put, the same pair may interact via a wealth of different geometries. This makes understanding the precise manner by which proteins recognise target sequences within a genome very unpredictable. Data has shown some amino acids to have a higher preference for certain bases. For example, Asn is frequently found to interact with A and Lys with G[134,135] nitrogen bases, which does suggest a framework for predicting interaction could

exist. However, we also know transcription factor proteins usually bind a family of related sequences within a genome and regulate multiple genes[136,137] which seems to be in contradiction of such a rule of protein-DNA interface. The ambiguity of recognition is expanded upon further by protein's ability to tolerate various base mutations within protein-DNA interactions.

Protein-DNA interactions are not limited to amino acid-DNA base interfaces but also include indirect recognition. This is observable when base mutations occurring in sequence regions that are not in direct contact with amino acids affect binding affinity[138-140]. These indirect interactions are thought to be the result of proteins recognising DNA sequences via water-mediated contacts, sequence dependent geometries and/or binding-induced DNA distortion (flexibility). Such interactions have been determined experimentally by observed changes in conformation energy readouts in both protein and DNA[141] following binding as well as altered specificity amongst distinct major and minor groove conformations[142]. With so many different forces driving direct and indirect binding during DNA-protein interactions scientists cannot rely on simple amino acid and DNA base sequences alone when beginning to design therapeutic molecules. Fluorescent labelling is an easy way determine protein binding - once binding is confirmed, the protein's physical structure should be determined by X-ray crystallography to wholly characterise the binding pocket. The structural information gained from this method can allow subsequent alteration to potentially improve binding affinities or

identify analogous molecules for conjugation to deliver the best possible chemotherapy.

1.3.2 Understanding G-quadruplex DNA

The structure of DNA is most commonly thought of as a duplex molecule, with two complementary strands of polynucleotides held together by Watson-Crick base pairing. Certain guanine (G) rich DNA sequences can form four-stranded structures called G-quadruplexes (G4s). G4s are stacked associations of G-quartets that arrange into planar assemblies of four Hoogsteen-bonded guanines. G-quartets arise through the interaction of guanines on single (intramolecular) or multiple (intermolecular) nucleic acid strands[143]. Knowledge of G-quartets and quadruplex structures actually precedes the DNA double-helix structure by 50 years. Yet, the nature of these G aggregates was only elucidated in the 1960s when fibre diffraction studies at a polynucleotide helix level became possible[144].

The somewhat serendipitous nature of discovery of the G-quartet by Gellert et al in 1962 hinged on a brief observational notes included in a 1910 publication by Ivar Bang[145] where he reported that concentrated solutions of guanylic acid formed a gel. Gellert and colleagues had also observed that “concentrated solutions (25.0mg/ml) of guanylic acid (GMP) at pH 5 are extremely viscous” and, if cooled form a “clear gel”. Importantly, this paper revealed the structural basis for the G4 phenomenon lay in G-tetrad formations driving the assembly of four-stranded helixes by guanine-rich oligonucleotides. The next major breakthrough in the

recognition of G4s came in 1982 when Sen and Gilbert showed single stranded DNA containing short guanine-rich motifs were capable of self-associating at physiological conditions[146]. In that publication, Sen and Gilbert recognised that the thermodynamic stability of G4s was cation-dependent, and correctly proposed that complexes were held together by Hoogsteen base-pairing between guanine residues. They identified guanine-rich sequences in immunoglobulin switch regions, gene promoters, and chromosomal telomeres. Additionally G4 structures were implicated in cancer development as possible reasons for precancerous cell behaviour including non-allelic homologous recombination. This prediction would later be confirmed at the beginning of the new millennium when G4 structural instability was linked to endogenous and induced mutagenicity within the genome and their tendency to concentrate in replication promoter and enhancer regions, as well as telomere caps[147,148].

1.3.2.1 Diversity amongst G-quadruplexes

The stacking of tetrads produces structures with an appearance reminiscent of duplex DNA with regular rises and twist between planes that form a right-handed helical structure[143] (Figure 1-8.)

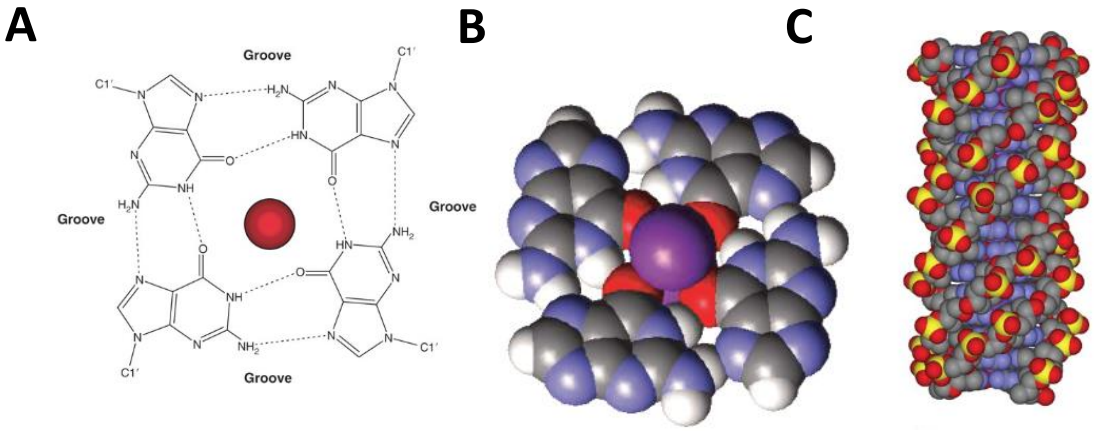


Figure 1-8. Schematic representation of guanine base structures. (**A**) The arrangement of guanine bases in the G-quartet, shown together with a centrally placed metal ion. Hydrogen bonds are shown as dotted lines, and the positions of the grooves are indicated. (**B**) Three dimensional space filling model of G-quartet. (**C**) The poly(dG) four-fold, right-handed helix.

There is a great diversity in G4 topology, which arises from various permutations in strand number, direction, loop size, and loop sequence[143] (Figure 1-9). When the four bases of a G-quartet belong to the same DNA strand, the resulting G4 structure is called an intramolecular G-quadruplex or monomer. Stacking between quartets belonging to two separate strands are then called intermolecular G-quadruplexes, dimers and four stands are logically tetramers[149]. Intramolecular G4s are more stable, rapidly forming structures. This is contrasted by slower forming tetramer structures[150]. The quick monomer formation times leads to topology variations occurring both within, and between, distinct G-rich sequences[151-154]. It remains unclear to what degree different conformations affect binding and targeting of G4s, although, it is only logical that it must play some role. This would be particularly important when accounting for steric hindrance in G4 targeting ligands that rely predominantly on pi stacking interactions for binding. What can be concluded is that regions containing four G bases can spontaneously

form G-quartets with ease and can then form G4s. These potential quadruplex-forming sequences are especially abundant in genomic regions synonymous with cancer progression.

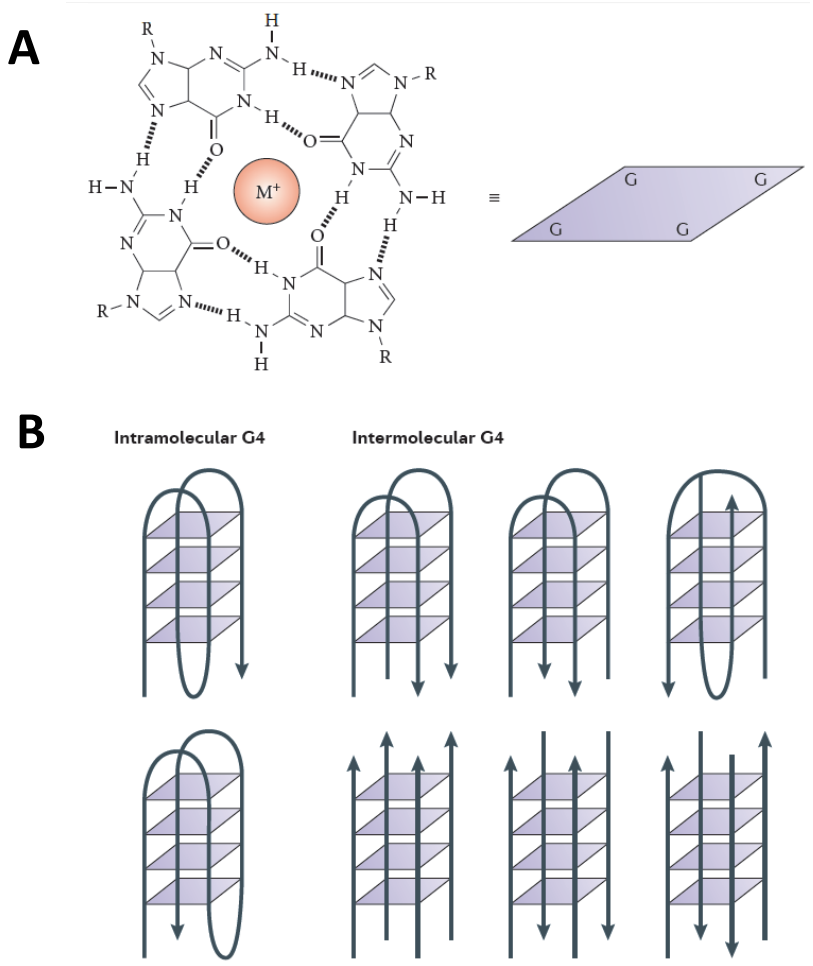


Figure 1-9. Arrangement of G-quartets and the varied G-quadruplex topologies. (A) Schematic representation of guanines into G-quartets (blue rhomboid) with central metal cation (M^+). (B) Stacking of quartets to form intramolecular (left) and intermolecular (right) G-quadruplexes. Arrow directions denote direction of the respective DNA strands. The intermolecular arrangements shown (left-right) on the top line are parallel, anti-parallel and hybrid dimer topologies. The same tetramer variants of topologies are reflected in the bottom line of schematics[155]. Reprinted from Nature Reviews Genetics, 13/11, Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. 2012 Nov;13(11):770, with permission from Nature Publication Group.

1.3.2.2 How G-quadruplexes contribute to cancer progression?

It has been noticed that such G-rich motifs with the potential to be quadruplex-forming (PG4s) are over-expressed in regulatory regions of the human genome. These include, but are not limited to, promoter regions of genes involved in replication, oncogenes, introns, untranslated regions, and at the 3' end of telomeric DNA[156-163]. An important consequence of G4-DNA formations is an increased risk of DNA breakage caused by their obstruction of DNA polymerase. This comes with an additional increased risk of non-allelic homologous recombination during DNA replication[147,148,164,165]. Genome wide molecular profiling of G4-DNA within whole DNA extracted from human B lymphocytes reveal a strong association with oncogenes, tumour suppressors, and somatic copy-number alterations related to cancer development[166]. A proto-gene is a normal gene which, when altered by mutations, becomes an oncogene contributing to tumorigenesis. When a proto-gene contains PG4s in their promotor region they are particularly susceptible to go on to form cancer. Examples of this have been shown in the c-MYC, BRCA1, BRCA2, VEGF, HIF-1, Ret, KRAS, bcl-2, c-Kit, PDGF-A, c-Myb, hTERT, HER2 and TP53 genes[166-170]. With such a strong connection between observable G4 DNA structures and oncogene progression, it comes as no surprise that new G4 targeting drugs are of great interest to cancer researchers.

1.3.2.3 G4 Targeting therapeutics and ligands

The intrinsic mutagenicity of G4s likely plays a crucial role in tumorigenesis, and this makes targeting G4s an attractive strategy in fighting cancer. This has justifiably

lead to designing small molecules that stabilise G4 conformations and trap quadruplexes in cellular DNA. These small molecules have the effect to down-regulate oncogene transcription, promote telomere disruption and induce growth arrest in cancer cells[122,123,171]. In the design of novel anti-cancer therapeutics that target G4s, a general consensus exists that the best mode of action for these molecules should be to stabilise quadruplex structures. Stabilisation inside G4s comes mostly via pi-pi stacking and electrostatic interactions which means ligand binding occurs on the external surface at the G-quartet level. This is most commonly achieved by incorporating flat aromatic molecules into ligand design with the intended purpose to increase quadruplex structure selectivity over duplex structure. Based on their nature of cationic stabilisation, there are four broad categories of quadruplex ligands: (1) amine appendage *in situ* protonating molecules, (2) *N*-methylation of an aza-aromatic moiety, (3) central metal ion, and (4) non-canonical ligands[124].

At the core of most G4 ligand designs is the incorporation of a large flat aromatic region prone to pi stacking that must also possess reasonable water solubility. The usual way to ensure both hydrophobic and hydrophilic characteristics around an aromatic core is to introduce a protonated sidearm (in many cases amine groups). An early example is a disubstituted acridine member called BSU6039 whose binding is directed by hydrophobic-pi-stacking of its aromatic core. This stacking is assisted by electrostatic interactions between two protonated sidechains and the grooves of associating quadruplexes[172]. This structure was later optimised with three sidechains and named BRACO-19. The alteration meant a 31 fold binding

preference quadruplex DNA vs duplex DNA and was shown to be a potent telomerase inhibitor (IC_{50}-TRAP =115nM)[158]. With pi-stacking still dominating the binding behavior of G4 ligands many other variants of this class went in other directions to improve potency and increased the size of their aromatic core. The PIPER ligand is characterised by a broader hydrophobic core and this structure meant a 42-fold preference for quadruplex binding. This did however, come at the cost of telomerase inhibition which was a modest IC_{50} of 20uM[173].

Alternate design pathways, beyond quaternization of amine sidechains via *in situ* protonation have been exploited in the form of *N*-methylated ligands. This category of G4 binding ligands are characterized by very high binding affinities, efficient telomerase inhibition, and often result in down regulation of oncogenes (such as *c-myc* and *k-ras*). Despite promising results, this family of ligands is often plagued with poor or even non-selective DNA binding[174] which has no doubt contributed to the reduced frequency of reporting. It is worth noting that a few small molecule derivative structures have emerged with improved selectivity that have even displayed synergistic action with the staple anti-cancer agent Taxol[175] but levels of non-G4 binding remain comparatively high.

Deviating further from the classical organic G4 ligands is a group of metallo-organic complexes that attach to G4s cationic central channel. These molecules arrange around a central metal cation flanked by aromatics. When the central metal is positioned over the cation channel of accessible G4s it optimizes stacking interactions. This type of G4 ligand has two major advantages over its counterparts

– synthetic ease of formation and very promising G4 binding affinity. Like the other categories metallo-organic ligands have been shown to be good G4 stabilisers and modest telomerase inhibitors but by far and away it is their binding affinity and selective preference that makes them an attractive field to expand. One report was even capable of citing an impressive 10 000-fold quadruplex vs duplex DNA selectivity with their Mn(III) based polyphyrin complex[176].

Natural ligands, or non-canonical ligands, are a developing division of G4 stabilising agents. Telomestatin is a stand out molecule in this category. Its polyheteroaromatic 24-ring structure greatly stabilises G4s and appears to be one of the most selective quadruplex ligands. The most exciting feature of telomestatin is a complete absence of duplex DNA affinity and exceptional telomerase inhibition (IC_{50}-TRAP =5nM). The perfect macrocyclic shape of this natural molecule has even been shown to stabilise G4s in salt-deficient conditions[124]. Where Telomestatin and other similar non-canonical ligands fall by the way side is preparation and limited solubility. The complex pathway to formation does not lend itself to large-scale preparation[177]. Without the creation of new cost effective pathways to prepare these molecules, in a plasma soluble form, their value as G4 targeting anti-cancer agents is significantly reduced.

An advance in G4-stabilising-ligand based chemotherapy has been observed in small molecules that target the loops and/or grooves of G4-DNA rather than terminal ends. By doing so, these ligands interact on secondary structural level to increase binding affinity and prioritize G4-DNA specifically over duplex DNA[143].

Triarylpyridines represent a class of compounds that interact with G4-DNA in just such a way. Intramolecular interactions between a host of triarylpyridines analogues and G4s (measured by Förster Resonance Energy Transfer (FRET) melting assays) show a stabilisation effect on DNA. Subsequent MTT assays using these analogues revealed significant cytotoxicity in HeLa cell lines when compared to other benchmark G4-DNA ligands[123].

An abundance of proteins have been reported to interact G4-DNA in an *in vitro* setting. Helicases such as BLM, WRN and FANCJ are able to recognize and unwind G4-DNA[156]. Nucleolin has also been shown to be a G4-DNA stabilizing protein. The mechanisms of nucleolin action begins following binding to G4-DNA in the c-myc promoter and result in reduced gene transcription[178]. An important recent breakthrough in engineering G4-targeting proteins has come from work at Cancer Research UK, Cambridge Institute when they designed a G4 structure-specific antibody called BG4[179]. In their comprehensive study Biffi et al. used a phage display library of 2.3 x 10^{10} different single chain antibody clones[180] to identify BG4 as the best hit for G4s among the selected binders. ELISA assays showed nanomolar affinity for intra and intermolecular G4s (K_d = 0.5-1.6 nM and 2.0 nM respectively) with no observable binding in RNA hairpins, single-stranded, or double-stranded DNA. BG4 binding was also found to be competitively selective when incubated in up to a 50-fold excess with various competitors including: yeast tRNA, double-stranded poly $(GC)_n$ or poly $(AT)_n$ and hairpin RNA. BG4 has been used to quantitatively visualise G4 DNA in human cells and M phase trapped human chromosomes. It continues to be used in an effort to map precise locations of G4

structures genome wide but the highly selective, low nanomolar affinity ranges suggest the binding pocket of BG4 is a near perfect model to begin designing therapeutic small molecules that treat cancers by G4 targeting. With BG4 having only been engineered in 2007, its distinct physical structure has not yet been solved by X-ray chromatographic analysis of its crystallised form. This is precisely the challenge I seek to address in the original research presented in the fourth chapter of this thesis.

1.4 Summary

Protein based drugs are a very effective and attractive class of therapeutic molecule. Despite extensive research efforts, comprehensive cures for many types of cancers still do not exist. TNBC is one such subtype of cancer that is currently treated with courses of non-specific chemotherapy that have wide spread systemic effects and high rates of recurrences. Proteins are a multifunctional class of molecules that perform highly specialized jobs in living cells. Fundamental properties like substrate specificity, high binding affinities, potency in physiological conditions, ease in manufacturing and target diversity make protein therapeutics a robust candidate for a targeted treatment of TNBCs.

When designing efficient drug therapies against highly specific diseases like TNBC, strategists often ignore cost as a variable for consideration. It is evident from the literature that recombinant protein expression is the most industrially scalable

means of harvesting therapeutic proteins. Though not without drawbacks, the considerably faster and higher yields resulting from recombinant expression lend them to widespread distribution of protein drugs with which cell-free or synthetic methodologies cannot compete. In trying to maximize these two advantages, researchers should look to employ bacterial hosts during recombinant expression. Though this methodology is usually not suitable for larger protein expressions and is associated with a considerable amount of yield either being insoluble or misfolded, there are means to recover imperfect proteins. Many of the current methods are time consuming processes which reduce the advance of this method to simple large volumetric yields. Shear forces between fluids have been shown to alter protein structures and direct folding. Our laboratory possesses a unique machine called the Vortex fluid device (VFD) that uses shear forces to rapidly and gently synthesize nanomaterials. Through controlled application of the shear forces within the VFD we may be able to refold proteins lost during the recombinant bacterial expression process. By doing this we can drastically cut costs associated with production of therapeutic proteins so the research field can greatly expand and keep chemotherapy cost down for patients.

Current chemotherapy treatment of TNBC is nonspecific and could actually be a contributing factor to the many drug resistant recurrences observed in TNBC patients. Literature is still debating over what precisely defines a TNBC due to the complex gene expression profile of the disease. Many TNBC gene mutations have overlapping and complimentary effects which has led to inconsistencies in disease profile which has hindered our ability to identify targets and produce a TNBC

specific chemotherapy. Nanoparticle therapies have been shown to improve drug outcomes of several drugs while also adding a degree of multi-functionality not otherwise exhibited by different formulations of a drug. Taking the active portions of a therapeutic protein and reducing them into a peptide deliverable by nanoparticle appears to be a promising option for novel TNBC therapies. Previously peptide based drugs have been cleared fast and have low target affinity *in vivo*. When combined with multi-functionable nanoparticles these problems have been silenced. We have a therapeutic interference peptide (iPep) that is a potent killer of cancers cells. It does this by interfering with the Engrailed 1 (EN1) gene transcription factor. The EN1 gene is selectively overexpressed in TNBC which makes it a selective target to be treated with nanoparticles containing iPep as a comprehensive solution to TNBC.

Special guanine rich sequence regions of DNA can form non-Watson-Crick structures called G-quadruplexes (G4). G4s have been reported to occur more frequently in regulatory regions of the genome and particular oncogenes which suggests an involvement in oncogenesis. G4 structures are less stable than traditional DNA structures and this weakness is thought to be an underlying reason behind DNA damage leading to the cancerous state. The unique shape of G4s create added complications when designing DNA-binding proteins that target G4s as an anticancer therapy. While G4s are not identical in structure, a G4 specific antibody has been created that is able to bind any structural variant. The antibody, BG4, shows highly preferential binding to G4 over duplex DNA in its current role as a fluorescence reporter for G4 locations in genomic DNA. The highly selective

binding observed between BG4 and G4 could make an excellent mode of binding for G4-targeting cancer therapies. While the sequence and activity of BG4 is known, its physical structure and binding pocket have not been revealed by X-ray crystallography. By successfully solving the structure of BG4 we can go on to design complimentary binding agents. Such agents could be useful in targeted chemotherapies or incorporate analogous existing structures to mimic the selective binding of the antibody.

REFERENCES

1. Lyons JA, Shahsavar A, Paulsen PA, Pedersen BP, Nissen P. Expression strategies for structural studies of eukaryotic membrane proteins. Current opinion in structural biology. 2016;38:137-144.
2. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Recombinant protein expression in microbial systems. 2014;7.
3. Assenberg R, Wan PT, Geisse S, Mayr LM. Advances in recombinant protein expression for use in pharmaceutical research. Current opinion in structural biology. 2013;23(3):393-402.
4. Terada T, Murata T, Shirouzu M, Yokoyama S. Cell-free expression of protein complexes for structural biology. Structural Genomics: General Applications. 2014:151-159.
5. Proverbio D, Henrich E, Orbán E, Dötsch V, Bernhard F. Membrane protein quality control in cell-free expression systems: tools, strategies and case studies. In: Membrane Proteins Production for Structural Analysis: Springer; 2014.
6. Zheng X, Dong S, Zheng J, Li D, Li F, Luo Z. Expression, stabilization and purification of membrane proteins via diverse protein synthesis systems and detergents involving cell-free associated with self-assembly peptide surfactants. Biotechnology advances. 2014;32(3):564-574.
7. Zimmerman ES, Heibeck TH, Gill A, Li X, Murray CJ, Madlansacay MR, et al. Production of site-specific antibody–drug conjugates using optimized non-natural amino acids in a cell-free expression system. Bioconjugate chemistry. 2014;25(2):351-361.
8. Linser R, Gelev V, Hagn F, Arthanari H, Hyberts SG, Wagner G. Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. Journal of the American Chemical Society. 2014;136(32):11308-11310.
9. Sinz A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. Expert review of proteomics. 2014;11(6):733-743.
10. Heck T, Faccio G, Richter M, Thöny-Meyer L. Enzyme-catalyzed protein crosslinking. Applied microbiology and biotechnology. 2013;97(2):461-475.
11. Hermanson GT. Bioconjugate techniques. Academic press; 2013.
12. Meyer H-P, Schmidhalter D. Innovations in Biotechnology. In Tech, Rijeka; 2012.
13. Freire E, Schön A, Velazquez-Campoy A. Isothermal titration calorimetry: general formalism using binding polynomials. Methods in enzymology. 2009;455:127-155.
14. Carrió M, Villaverde A. Protein aggregation as bacterial inclusion bodies is reversible. FEBS letters. 2001;489(1):29-33.
15. Katoh S, Katoh Y. Continuous refolding of lysozyme with fed-batch addition of denatured protein solution. Process Biochemistry. 2000;35(10):1119-1124.
16. Maeda Y, Koga H, Yamada H, Ueda T, Imoto T. Effective renaturation of reduced lysozyme by gentle removal of urea. Protein engineering. 1995;8(2):201-205.
17. Müller C, Rinas U. Renaturation of heterodimeric platelet-derived growth factor from inclusion bodies of recombinant Escherichia coli using size-exclusion chromatography. Journal of Chromatography A. 1999;855(1):203-213.
18. Vallejo LF, Rinas U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microbial cell factories. 2004;3(1):11.
19. Cyr DM, Langer T, Douglas MG. DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends in biochemical sciences. 1994;19(4):176-181.

20. Thomas JG, Ayling A, Baneyx F. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins fromE. coli. Applied biochemistry and biotechnology. 1997;66(3):197-238.
21. Hartl FU. Molecular chaperones in cellular protein folding. Nature. 1996;381(6583):571.
22. West S, Chaudhuri J, Howell J. Improved protein refolding using hollow-fibre membrane dialysis. Biotechnology and bioengineering. 1998;57(5):590-599.
23. Yoshii H, Furuta T, Yonehara T, Ito D, Linko Y-Y, Linko P. Refolding of denatured/reduced lysozyme at high concentration with diafiltration. Bioscience, biotechnology, and biochemistry. 2000;64(6):1159-1165.
24. Robinson CR, Sligar SG. [18] Hydrostatic and osmotic pressure as tools to study macromolecular recognition. Methods in enzymology. 1995;259:395-427.
25. Foguel D, Robinson CR, de Sousa PC, Silva JL, Robinson AS. Hydrostatic pressure rescues native protein from aggregates. Biotechnology and Bioengineering. 1999;63(5):552-558.
26. John RJS, Carpenter JF, Randolph TW. High pressure fosters protein refolding from aggregates at high concentrations. Proceedings of the National Academy of Sciences. 1999;96(23):13029-13033.
27. Francis DM, Page R. Strategies to optimize protein expression in E. coli. Current protocols in protein science. 2010:5.24. 1-5.24. 29.
28. Fischer B, Perry B, Sumner I, Goodenough P. A novel sequential procedure to enhance the renaturation of recombinant protein from Escherichia coil inclusion bodies. Protein engineering. 1992;5(6):593-596.
29. Clark EDB, Schwarz E, Rudolph R. [15] Inhibition of aggregation side reactions during in vitro protein folding. Methods in enzymology. 1999;309:217-236.
30. Geng X, Chang X. High-performance hydrophobic interaction chromatography as a tool for protein refolding. Journal of Chromatography A. 1992;599(1-2):185-194.
31. Schultz CP. Illuminating folding intermediates. Nat. Struct. Biol. 2000;7:7-10.
32. Radford SE. Protein folding: progress made and promises ahead. Trends in biochemical sciences. 2000;25(12):611-618.
33. Lin Z, Madan D, Rye HS. GroEL stimulates protein folding through forced unfolding. Nature structural & molecular biology. 2008;15(3):303-311.
34. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324-332.
35. Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, Sikor M, et al. Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell. 2010;142(1):112-122.
36. Yim S, Shamlou P. The engineering effects of fluids flow on freely suspended biological macro-materials and macromolecules. In: Influence of Stress on Cell Growth and Product Formation: Springer; 2000.
37. Lin J-J, Meyer JD, Carpenter JF, Manning MC. Stability of human serum albumin during bioprocessing: denaturation and aggregation during processing of albumin paste. Pharmaceutical research. 2000;17(4):391-396.
38. Bekard IB, Asimakis P, Bertolini J, Dunstan DE. The effects of shear flow on protein structure and function. Biopolymers. 2011;95(11):733-745.
39. Bekard IB, Dunstan DE. Shear-induced deformation of bovine insulin in Couette flow. The Journal of Physical Chemistry B. 2009;113(25):8453-8457.
40. Maa Y-F, Hsu CC. Effect of high shear on proteins. Biotechnology and bioengineering. 1996;51(4):458-465.
41. Hill EK, Krebs B, Goodall DG, Howlett GJ, Dunstan DE. Shear flow induces amyloid fibril formation. Biomacromolecules. 2006;7(1):10-13.
42. Chen X, Dobson JF, Raston CL. Vortex fluidic exfoliation of graphite and boron nitride. Chemical Communications. 2012;48(31):3703-3705.

43. Wahid MH, Eroglu E, Chen X, Smith SM, Raston CL. Functional multi-layer graphene–algae hybrid material formed using vortex fluidics. Green Chemistry. 2013;15(3):650-655.
44. Martin AD, Boulos RA, Hubble LJ, Hartlieb KJ, Raston CL. Multifunctional water-soluble molecular capsules based on p-phosphonic acid calix [5] arene. Chemical Communications. 2011;47(26):7353-7355.
45. Yasmin L, Chen X, Stubbs KA, Raston CL. Optimising a vortex fluidic device for controlling chemical reactivity and selectivity. Scientific reports. 2013;3:2282.
46. Zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nature Reviews Cancer. 2002;2(5):342-350.
47. Baseman JG, Koutsky LA. The epidemiology of human papillomavirus infections. Journal of clinical virology. 2005;32:16-24.
48. Kim S-S, Tam JK, Wang A-F, Hegde RS. The structural basis of DNA target discrimination by papillomavirus E2 proteins. Journal of Biological Chemistry. 2000;275(40):31245-31254.
49. Abbate EA, Berger JM, Botchan MR. The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. Genes & development. 2004;18(16):1981-1996.
50. Brannon AR, Maresca JA, Boeke JD, Basrai MA, McBride AA. Reconstitution of papillomavirus E2-mediated plasmid maintenance in Saccharomyces cerevisiae by the Brd4 bromodomain protein. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(8):2998-3003.
51. Schwarz-Linek U, Pilka ES, Pickford AR, Kim JH, Höök M, Campbell ID, et al. High affinity streptococcal binding to human fibronectin requires specific recognition of sequential F1 modules. Journal of Biological Chemistry. 2004;279(37):39017-39025.
52. Schwarz-Linek U, Werner JM, Pickford AR, Gurusiddappa S, Kim JH, Pilka ES, et al. Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. Nature. 2003;423(6936):177-181.
53. Kim JH, Singvall J, Schwarz-Linek U, Johnson BJ, Potts JR, Höök M. BBK32, a fibronectin binding MSCRAMM from Borrelia burgdorferi, contains a disordered region that undergoes a conformational change on ligand binding. Journal of Biological Chemistry. 2004;279(40):41706-41714.
54. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular basis for amyloid fibril formation and stability. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(2):315-320.
55. Harjes P, Wanker EE. The hunt for huntingtin function: interaction partners tell many different stories. Trends in biochemical sciences. 2003;28(8):425-433.
56. Li X-J, Li S-H, Sharp AH, Nucifora Jr FC. A huntingtin-associated protein enriched in brain with implications for pathology. Nature. 1995;378(6555):398.
57. Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. Nature Reviews Cancer. 2017;17(2):93-115.
58. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. cell. 2011;144(5):646-674.
59. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How many etiological subtypes of breast cancer: two, three, four, or more? Journal of the National Cancer Institute. 2014;106(8):dju165.
60. Anderson WF, Matsuno R. In: Breast cancer heterogeneity: a mixture of at least two main types?, 2006 Oxford University Press.
61. Prat A, Adamo B, Fan C, Peg V, Vidal M, Galván P, et al. Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. Scientific reports. 2013;3:3544.

62. Altman J, Bayer SA. Development of the brain stem in the rat. V. Thymidine-radiographic study of the time of origin of neurons in the midbrain tegmentum. Journal of Comparative Neurology. 1981;198(4):677-716.
63. Andersen JK. Does neuronal loss in Parkinson's disease involve programmed cell death? Bioessays. 2001;23(7):640-646.
64. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical cancer research. 2007;13(15):4429-4434.
65. Carey L, O'shaughnessy J, Hoadley K, Khambata-Ford S, Horak C, Xu L, et al. In: Potential Predictive Markers of Benefit from Cetuximab in Metastatic Breast Cancer: An Analysis of Two Randomized Phase 2 Trials. 2009 AACR.
66. Perou CM. Molecular stratification of triple-negative breast cancers. The oncologist. 2011;16(Supplement 1):61-70.
67. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? Nature reviews Clinical oncology. 2010;7(12):683-692.
68. Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. Breast Cancer Research. 2005;7(4):143.
69. Lakhani SR, O'Hare MJ. The mammary myoepithelial cell-Cinderella or ugly sister? Breast Cancer Research. 2000;3(1):1.
70. Viale G, Rotmensz N, Maisonneuve P, Bottiglieri L, Montagna E, Luini A, et al. Invasive ductal carcinoma of the breast with the “triple-negative” phenotype: prognostic implications of EGFR immunoreactivity. Breast cancer research and treatment. 2009;116(2):317-328.
71. Van De Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. The American journal of pathology. 2002;161(6):1991-1996.
72. Shin BK, Lee Y, Lee JB, Kim HK, LEE J, CHO S, et al. Breast carcinomas expressing basal markers have poor clinical outcome regardless of estrogen receptor status. Breast cancer. 2001 (2).
73. Turner N, Tutt A, Ashworth A. Hallmarks of'BRCAness' in sporadic cancers. Nature Reviews Cancer. 2004;4(10):814-819.
74. Tammaro C, Raponi M, Wilson DI, Baralle D. In: BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer. 2012 Portland Press Limited.
75. Nakuci E, Mahner S, DiRenzo J, ElShamy WM. BRCA1-IRIS regulates cyclin D1 expression in breast cancer cells. Experimental cell research. 2006;312(16):3120-3131.
76. ElShamy WM, Livingston DM. Identification of BRCA1-IRIS, a BRCA1 locus product. Nature cell biology. 2004;6(10):954-967.
77. Silver DP, Livingston DM. Mechanisms of BRCA1 tumor suppression. Cancer discovery. 2012;2(8):679-684.
78. Li ML, Greenberg RA. Links between genome integrity and BRCA1 tumor suppression. Trends in biochemical sciences. 2012;37(10):418-424.
79. Walsh T, King M-C. Ten genes for inherited breast cancer. Cancer cell. 2007;11(2):103-105.
80. Schlacher K, Wu H, Jasin M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. Cancer cell. 2012;22(1):106-116.
81. Pathania S, Nguyen J, Hill SJ, Scully R, Adelmant GO, Marto JA, et al. BRCA1 is required for postreplication repair after UV-induced DNA damage. Molecular cell. 2011;44(2):235-251.

82. Bunting SF, Callén E, Kozak ML, Kim JM, Wong N, López-Contreras AJ, et al. BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. Molecular cell. 2012;46(2):125-135.
83. Ashworth A. A synthetic lethal therapeutic approach: poly (ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. Journal of Clinical Oncology. 2008;26(22):3785-3790.
84. Carey LA. Targeted chemotherapy? Platinum in BRCA1-dysfunctional breast cancer. Journal of Clinical Oncology. 2010;28(3):361-363.
85. Muller PA, Vousden KH. Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer cell. 2014;25(3):304-317.
86. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-339.
87. Leroy B, Fournier JL, Ishioka C, Monti P, Inga A, Fronza G, et al. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. Nucleic acids research. 2013;41(D1):D962-D969.
88. Wang Y, Suh Y-A, Fuller MY, Jackson JG, Xiong S, Terzian T, et al. Restoring expression of wild-type p53 suppresses tumor growth but does not cause tumor regression in mice with a p53 missense mutation. The Journal of clinical investigation. 2011;121(3):893-904.
89. Weisz L, Oren M, Rotter V. Transcription regulation by mutant p53. Oncogene. 2007;26(15):2202-2211.
90. Gaiddon C, Lokshin M, Ahn J, Zhang T, Prives C. A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain. Molecular and cellular biology. 2001;21(5):1874-1887.
91. Strano S, Fontemaggi G, Costanzo A, Rizzo MG, Monti O, Baccarini A, et al. Physical interaction with human tumor-derived p53 mutants inhibits p63 activities. Journal of Biological Chemistry. 2002;277(21):18817-18826.
92. Jackson JG, Pant V, Li Q, Chang LL, Quintás-Cardama A, Garza D, et al. p53-mediated senescence impairs the apoptotic response to chemotherapy and clinical outcome in breast cancer. Cancer cell. 2012;21(6):793-806.
93. Nguyen PL, Taghian AG, Katz MS, Niemierko A, Abi Raad RF, Boon WL, et al. Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. Journal of Clinical Oncology. 2008;26(14):2373-2378.
94. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clinical Cancer Research. 2005;11(16):5678-5685.
95. Tan DS, Marchió C, Jones RL, Savage K, Smith IE, Dowsett M, et al. Triple negative breast cancer: molecular profiling and prognostic impact in adjuvant anthracycline-treated patients. Breast cancer research and treatment. 2008;111(1):27-44.
96. Falo C, Moreno A, Varela M, Lloveras B, Figueras A, Escobedo A. HER-2/neu status and response to CMF: retrospective study in a series of operable breast cancer treated with primary CMF chemotherapy. Journal of cancer research and clinical oncology. 2007;133(7):423-429.
97. Rodríguez-Pinilla SM, Sarrió D, Honrado E, Hardisson D, Calero F, Benitez J, et al. Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. Clinical cancer research. 2006;12(5):1533-1539.
98. Byrski T, Huzarski T, Dent R, Gronwald J, Zuziak D, Cybulski C, et al. Response to neoadjuvant therapy with cisplatin in BRCA1-positive breast cancer patients. Breast cancer research and treatment. 2009;115(2):359-363.
99. Sirohi B, Arnedos M, Popat S, Ashley S, Nerurkar A, Walsh G, et al. Platinum-based chemotherapy in triple-negative breast cancer. Annals of oncology. 2008:mdn395.

100. Davis ME, Shin DM. Nanoparticle therapeutics: an emerging treatment modality for cancer. Nature reviews Drug discovery. 2008;7(9):771-782.
101. Ragupathi G, Gathuru J, Livingston P. Antibody inducing polyvalent cancer vaccines. In: Tumor Immunology and Cancer Vaccines: Springer; 2005.
102. Dorsam RT, Gutkind JS. G-protein-coupled receptors and cancer. Nature reviews cancer. 2007;7(2):79-94.
103. Aina OH, Sroka TC, Chen ML, Lam KS. Therapeutic cancer targeting peptides. Peptide Science. 2002;66(3):184-199.
104. Mori T. Cancer-specific ligands identified from screening of peptide-display libraries. Current pharmaceutical design. 2004;10(19):2335-2343.
105. Thorpe PE. Vascular targeting agents as cancer therapeutics. Clinical Cancer Research. 2004;10(2):415-427.
106. Reff ME, Hariharan K, Braslawsky G. Future of monoclonal antibodies in the treatment of hematologic malignancies. Cancer Control. 2002;9(2):152-168.
107. Qiu X-Q, Wang H, Cai B, Wang L-L, Yue S-T. Small antibody mimetics comprising two complementarity-determining regions and a framework region for tumor targeting. Nature biotechnology. 2007;25(8):921-929.
108. Peck TE, Hill SA. Pharmacology for anaesthesia and intensive care. Cambridge University Press; 2014.
109. Frankel AD, Pabo CO. Cellular uptake of the tat protein from human immunodeficiency virus. Cell. 1988;55(6):1189-1193.
110. Green M, Loewenstein PM. Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. Cell. 1988;55(6):1179-1188.
111. Derossi D, Joliot AH, Chassaing G, Prochiantz A. The third helix of the Antennapedia homeodomain translocates through biological membranes. Journal of Biological Chemistry. 1994;269(14):10444-10450.
112. Pooga M, Hällbrink M, Zorko M. Cell penetration by transportan. The FASEB Journal. 1998;12(1):67-77.
113. Rothbard JB, Garlington S, Lin Q, Kirschberg T, Kreider E, McGrane PL, et al. Conjugation of arginine oligomers to cyclosporin A facilitates topical delivery and inhibition of inflammation. Nature medicine. 2000;6(11):1253-1257.
114. Schwarze SSR, Dowdy SSF. In vivo protein transduction: intracellular delivery of biologically active proteins, compounds and DNA. Trends in Pharmacological Sciences. 2000;21(2):45-48.
115. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proceedings of the National Academy of Sciences. 2009;106(33):13820-13825.
116. Beltran A, Graves LM, Blancafort P. Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. Oncogene. 2014;33(39):4767-4777.
117. Alves dos Santos M, Smidt M. In: En1 and Wnt signaling in midbrain dopaminergic neuronal development. Neural Dev 2011; 6: 23; PMID: 21569278.
118. Alvarez-Fischer D, Fuchs J, Castagner F, Stettler O, Massiani-Beaudoin O, Moya KL, et al. Engrailed protects mouse midbrain dopaminergic neurons against mitochondrial complex I insults. Nature neuroscience. 2011;14(10):1260-1266.
119. Yan C, Higgins PJ. Drugging the undruggable: transcription therapy for cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. 2013;1835(1):76-85.
120. Curnis F, Gasparri A, Sacchi A, Longhi R, Corti A. Coupling tumor necrosis factor-α with αV integrin ligands improves its antineoplastic activity. Cancer Research. 2004;64(2):565-571.

121. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nature structural & molecular biology. 2012;19(8):837-844.
122. Smith N, Corry B, Iyer KS, Norret M, Raston C. A microfluidic platform to synthesise a G-quadruplex binding ligand. Lab on a Chip. 2009;9(14):2021-2025.
123. Smith N, Labrunie G, Corry B, Tran PLT, Norret M, Djavaheri-Mergny M, et al. Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands. Organic & biomolecular chemistry. 2011;9(17):6154-6162.
124. Monchaud D, Teulade-Fichou M-P. A hitchhiker's guide to G-quadruplex ligands. Organic & biomolecular chemistry. 2008;6(4):627-636.
125. Kim TH, Ren B. Genome-wide analysis of protein-DNA interactions. Annu. Rev. Genomics Hum. Genet. 2006;7:81-102.
126. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature. 2001;409(6819):533-538.
127. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290(5500):2306-2309.
128. Bulyk ML, Gentalen E, Lockhart DJ, Church GM. Quantifying DNA–protein interactions by double-stranded DNA arrays. Nature biotechnology. 1999;17(6):573-577.
129. Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, et al. A common set of gene regulatory networks links metabolism and growth inhibition. Molecular cell. 2004;16(3):399-411.
130. Jen K-Y, Cheung VG. Identification of novel p53 target genes in ionizing radiation response. Cancer research. 2005;65(17):7666-7673.
131. Kim TH, Xiong H, Zhang Z, Ren B. [beta]-Catenin activates the growth factor endothelin-1 in colon cancer cells. Oncogene. 2005;24(4):597.
132. Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proceedings of the National Academy of Sciences. 2003;100(14):8164-8169.
133. Wei C-L, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006;124(1):207-219.
134. Kono H, Sarai A. Structure-based prediction of DNA target sites by regulatory proteins. Proteins: Structure, Function, and Bioinformatics. 1999;35(1):114-131.
135. Mandel-Gutfreund Y, Schueler O, Margalit H. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. Journal of molecular biology. 1995;253(2):370-382.
136. Choo Y, Klug A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. Proceedings of the National Academy of Sciences. 1994;91(23):11168-11172.
137. Suzuki M. A framework for the DNA–protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. Structure. 1994;2(4):317-326.
138. Chen S, Gunasekera A, Zhang X, Kunkel TA, Ebright RH, Berman HM. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. Journal of molecular biology. 2001;314(1):75-82.
139. Hegde RS. The papillomavirus E2 proteins: structure, function, and biology. Annual review of biophysics and biomolecular structure. 2002;31(1):343-360.
140. Horton NC, Dorner LF, Perona JJ. Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. Nature Structural & Molecular Biology. 2002;9(1):42-47.

141. Olson WK, Gorin AA, Lu X-J, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. Proceedings of the National Academy of Sciences. 1998;95(19):11163-11168.
142. Selvaraj S, Kono H, Sarai A. Specificity of protein–DNA recognition revealed by structure-based potentials: symmetric/asymmetric and cognate/non-cognate binding. Journal of molecular biology. 2002;322(5):907-915.
143. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic acids research. 2006;34(19):5402-5415.
144. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. Proceedings of the National Academy of Sciences. 1962;48(12):2013-2018.
145. Bang I. Untersuchungen über die Guanylsäure. Biochemische Zeitschrift. 1910;26:293-311.
146. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature. 1988;334(6180):364-366.
147. Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. Current Biology. 2008;18(12):900-905.
148. Pontier DB, Kruisselbrink E, Guryev V, Tijsterman M. Isolation of deletion alleles by G4 DNA-induced mutagenesis. Nature methods. 2009;6(9):655.
149. Bates P, Mergny JL, Yang D. Quartets in G-major. EMBO reports. 2007;8(11):1003-1010.
150. De Cian A, Mergny J-L. Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. Nucleic acids research. 2007;35(8):2483-2493.
151. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d [AGGG (TTAGGG) 3] in K+ solution. Bioorganic & medicinal chemistry. 2006;14(16):5584-5591.
152. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic acids research. 2006;34(9):2723-2735.
153. Phan AT, Modi YS, Patel DJ. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. Journal of the American Chemical Society. 2004;126(28):8710-8716.
154. Phan AT, Luu KN, Patel DJ. Different loop arrangements of intramolecular human telomeric (3+ 1) G-quadruplexes in K+ solution. Nucleic acids research. 2006;34(19):5715-5719.
155. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nature Reviews Genetics. 2012;13(11):770-780.
156. Maizels N, Gray LT. The G4 genome. PLoS Genet. 2013;9(4):e1003468.
157. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic acids research. 2007;35(2):406-413.
158. De Cian A, Cristofari G, Reichenbach P, De Lemos E, Monchaud D, Teulade-Fichou M-P, et al. Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. Proceedings of the National Academy of Sciences. 2007;104(44):17347-17352.
159. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie. 2008;90(8):1149-1171.
160. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002;417(6891):876-880.
161. Todd AK, Haider SM, Parkinson GN, Neidle S. Sequence occurrence and structural uniqueness of a G-quadruplex in the human c-kit promoter. Nucleic acids research. 2007;35(17):5799-5808.

162. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proceedings of the National Academy of Sciences. 2002;99(18):11593-11598.
163. Patel DJ, Phan ATn, Kuryavyi V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic acids research. 2007;35(22):7429-7455.
164. Sun D, Hurley LH. Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. G-Quadruplex DNA: Methods and Protocols. 2010:65-79.
165. Boán F, Gómez-Márquez J. In Vitro Recombination Mediated by G-Quadruplexes. Chembiochem. 2010;11(3):331-334.
166. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nature biotechnology. 2015;33(8):877-881.
167. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nature reviews Drug discovery. 2011;10(4):261-275.
168. Sagne C, Marcel V, Bota M, Martel-Planche G, Nobrega A, Palmero EI, et al. Age at cancer onset in germline TP53 mutation carriers: association with polymorphisms in predicted G-quadruplex structures. Carcinogenesis. 2014;35(4):807-815.
169. Sun D, Guo K, Shin Y-J. Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. Nucleic acids research. 2011;39(4):1256-1265.
170. Palumbo SL, Ebbinghaus SW, Hurley LH. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. Journal of the American Chemical Society. 2009;131(31):10878-10891.
171. Smith NM, Raston CL, Smith CB, Sobolev AN. PEG mediated synthesis of amino-functionalised 2, 4, 6-triarylpyridines. Green Chemistry. 2007;9(11):1185-1190.
172. Haider SM, Parkinson GN, Neidle S. Structure of a G-quadruplex–ligand complex. Journal of molecular biology. 2003;326(1):117-125.
173. Mazzitelli CL, Brodbelt JS, Kern JT, Rodriguez M, Kerwin SM. Evaluation of binding of perylene diimide and benzannulated perylene diimide ligands to DNA by electrospray ionization mass spectrometry. Journal of the American Society for Mass Spectrometry. 2006;17(4):593-604.
174. Haq I, Trent JO, Chowdhry BZ, Jenkins TC. Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin1. Journal of the American Chemical Society. 1999;121(9):1768-1779.
175. Cookson JC, Dai F, Smith V, Heald RA, Laughton CA, Stevens MF, et al. Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3, 11-difluoro-6, 8, 13-trimethyl-8H-quino [4, 3, 2-kl] acridinium methosulfate (RHPS4) in vitro: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents. Molecular pharmacology. 2005;68(6):1551-1558.
176. Dixon IM, Lopez F, Tejera AM, Estève J-P, Blasco MA, Pratviel G, et al. A G-quadruplex ligand with 10000-fold selectivity over duplex DNA. Journal of the American Chemical Society. 2007;129(6):1502-1503.
177. Doi T, Yoshida M, Shin-ya K, Takahashi T. Total synthesis of (R)-telomestatin. Organic letters. 2006;8(18):4165-4167.
178. González V, Guo K, Hurley L, Sun D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. Journal of Biological Chemistry. 2009;284(35):23622-23635.

179. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nature chemistry. 2013;5(3):182-186.
180. Schofield DJ, Pope AR, Clementel V, Buckell J, Chapple SD, Clarke KF, et al. Application of phage display to high throughput antibody generation and characterization. Genome biology. 2007;8(11):R254.

CHAPTER 2

Shear-stress-mediated refolding of proteins from aggregates and inclusion bodies

Abstract

Recombinant protein overexpression of large proteins by bacterial cells typically results in insoluble and misfolded proteins forced into inclusion body aggregates. Inspired by the natural processing of cellular chaperonins, we report a novel process using applied mechanical energy as shear stress within thin fluid film vortices. This technique is capable of unfolding substrate protein aggregates as well as isolating proteins away from misfolded intermediates. The shear forces successfully refolded boiled hen egg white lysozyme, recombinant hen egg white lysozyme, and recombinant caveolin-1. In addition, studies with the much larger protein, cAMP-dependent protein kinase A (PKA) also allowed refolding when shear stress was applied in a site-specific manner. The following methods reported require <5 minutes to recover protein activity which is >100-times faster than traditional methods using overnight dialysis and Hens egg white lysozyme (HEWL) expressed in inclusion bodies proved capable of refolding to 82% of its active conformation. The findings here have numerous application to industrial and research associated protein expression as they can significantly shorten refolding times, lower cost and reduce total waste streams associated with conventional practices.

Introduction

Recombinant protein expression currently occupies a market share in excess of $160 billion of world biotechnology tradings[1]. When compared with other expression systems such as mammalian cells, protein expression that occurs in yeast or *Escherichia coli* cell lines has been heavily favoured by laboratories because of rapid growth, low consumable costs and higher yields[2,3]. This preference however, does not extend to overexpression of larger proteins in bacterial systems as they typically results in aggregates forming as isolated inclusion bodies[4-6]. Once isolated to inclusion bodies the process of recovering correctly folded protein is tedious and expensive by conventional methods[7]. Most often these methods for refolding proteins incorporate time consuming multi-day dialysis that has the added drawback of producing large volumes of waste solutions (anywhere from 1-10 litres per mg of desired protein)[8].

When expressing high value proteins, such as G-protein-coupled receptors or therapeutic antibodies, laboratories can apply extensively optimized mammalian or insect cell lines, media and bioreactors in lieu of dialysis-based recovery protocols. Such methods are only used when the perceived reward is great enough as these methods of recombinant protein overexpression are very costly. Ultimately, recovery of native conformation protein from inclusion bodies of misfolded protein intermediates from both bacterial and mammalian cell lines is a challenge for industrial scale production because of their inefficiency. Because of this, new methods that can maximise the output and application of bacterial over-expression could revolutionise industrial and research protein production.

2.1 Materials and methods

Buffers Listed in Tables 2-1 and 2-2

Table 2-1. Expression conditions of recombinantly expressed proteins.

Protein	*Expression time (h)*	*Expression temp (ºC)*	*[IPTG] (M)*	*Lysis Buffer*
Hen egg white lysozyme	4	37	1	50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 1 mM HALT protease inhibitor (Pierce), 10 mM 2-mercaptoethanol, pH 8.0
Caveolin-Δ TM	3	37	0.5	50 mM tris-HCl, 10 mM NaCl, 5 mM EDTA, 100 mM PMSF, pH 8.0
HIV gp41	8	22	0.5	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM HALT protease inhibitor, pH 8.0
His-PKA	5	37	1	50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 1 mM HALT protease inhibitor, 10 mM 2-mercaptoethanol, pH 8.0

Table 2-2. Purification conditions of recombinantly expressed proteins. The protein target used for ELISA binding studies HIV gp41, was purified under non-denaturing conditions.

Protein	*Resin*	*Denaturing Buffer*	*Binding/wash Buffer*	*Elusion Buffer*
Hen egg white lysozyme	UNOsphere S (BIO-Rad)	20 mM tris, 10 mM NaCl, 8 M urea, pH 7.8	20 mM tris, 10 mM NaCl, 8M urea, pH 7.8	Wash buffer, 400 mM NaCl
Caveolin-Δ TM	Ni-NTA (Bio-Rad)	50 mM tris, 50 mM NaCl, 8 M urea, pH 8.0	50 mM tris-HCl, 300 mM NaCl, 10 mM imidazole, 0.2% Na azide, 8M urea, pH 8.0	Wash buffer, pH 4.0
HIV gp41	Ni-NTA	-	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM 2-mercaptoethanol, 20 mM imidazole, pH 8.0	Wash buffer, 250 mM imidazole
His-PKA	Ni-NTA	20 mM NaH2PO4, 500 mM NaCl, 6 M guanidine-HCl, pH 7.0	50 mM NaH_2PO_4, 300 mM NaCl, 1 mM imidazole, pH 7.0	Wash buffer, 500 mM imidazole

2.1.1 Expression and purification of HEWL, caveolin-Δ TM and HIV gp41

Hen egg white lysozyme (HEWL), caveolin-Δ TM and HIV pg41 were overexpressed in cultures of BL21 *E. coli* via induction with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM). The 1 L culture volume was centrifuged at 6000 rpm and the bacterial pellet collected. The pellet was reconstituted in lysis buffer with the aid of wand sonication in 30 s continuous bursts with 1 min cooling on ice for eight cycles (20 watts). HEWL and caveolin-ΔTM, were purified under denaturing conditions, and HIV gp41 was purified under non-denaturing conditions. Specific expression and purification conditions for each protein are outlined in Table 2-1 and 2-2, respectively. The egg whites were obtained from chicken eggs, and diluted 2:3 in PBS, heat-treated at 90 °C for 20 min, and dissolved in 8 M urea overnight at 4 °C. Cleavage of the His_6 tag attached to HIV gp41 was achieved with Tobacco Etch Virus protease, which was subsequently removed by immobilized metal affinity chromatography (IMAC). All protein concentrations were determined by Pierce bicinchoninic acid assay kit.

2.1.2 Protein refolding and determination of native state conformation

Commercial, lyophilized HEWL protein purchased from Sigma was reconstituted in PBS as 'active' HEWL sample. Recombinantly expressed HEWL was pre-treated by 1:100 rapid dilution in PBS, and then refolded using a Vortex Fluid Device (VFD). All samples were treated at room temp (22 °C) within a 16 cm long, 10 mm diameter glass test tube. Experiments conducted in confined mode set the VFD to a 45° tilt angle containing a fluid volume of 1 ml, spun at 5 krpm, unless otherwise specified.

When operating in continuous mode, experiments were conducted by flowing the rapidly diluted protein through the inlet port to the base of the sample at a flow rate of 0.1 ml/min. Caveolin-ΔTM VFD refolding was performed in confined mode (1 ml, 5 krpm, 22 °C). During control testing, caveolin-ΔTM was also refolded using conventional dialysis over 4 days (1:500, 50 mM Tris-HCl, 1 mM EDTA, 4 °C, pH 8.5). Circular dichroism spectra of HEWL were collected immediately following VFD refolding in PBS (20 nm/min, 4 accumulations), and caveolin-ΔTM were collected in 10 mM sodium phosphate, pH 7.5 (10 nm/min, 8 accumulations). All lysozyme activity assays used the Invitrogen brand EnzChek kit after rapid dilution from denatured protein solution into PBS (1:100) according to manufacturer's instructions, except for one deviation where the 37 °C incubation time was reduced from 30 to 10 min. Lysozyme activity was interpolated by least-square regressions fit of lysozyme standards to a Michaelis-Menten curve ($v = \frac{Vmax\,[S]}{Km+[S]} + c$) with Prism 6 software (GraphPad, Figure 2-2). After 1:100 rapid dilution into PBS, protein solution contains 80 mM urea, 0.2 mM Tris and 4 mM NaCl. Lysozyme activity with various levels of urea, Tris, and NaCl was determined to verify that the assay was not affected (Figure 2-6). Purified caveolin-ΔTM from the inclusion body were diluted 1:5 in 10 mM sodium phosphate, pH 7.5 and then briefly dialyzed prior to VFD treatment (1:100 volume, 1 h, 4 °C) for circular dichroism spectra.

2.1.3 ELISA binding assays

The dose-dependent ELISA was conducted by coating HIV gp41 (100 µl of 10 µg/ml in 50 mM sodium carbonate pH 9.6 for 4 h at 4 °C) on a Nunc Maxisorp 96-well

microtiter plate. After removing the coating solution, a blocking solution of 0.2% non-fat milk in PBS was applied. Caveolin-ΔTM, anti-His mouse monoclonal antibody (Sigma, H1029), and anti-mouse HRP-conjugated polyclonal antibody (1:2000, Sigma, A5906) were diluted in 100 µl PT buffer (1:1000, PBS, 0.05% Tween-20) and incubated for 1 h at 4 °C including four wash steps using PT buffer (200 µl). The ELISA was developed by the addition of 1% w/v *o*-phenylenediamine dihydrocholoride in citric acid buffer (0.02% w/v H_2O_2, 50 mM citric acid, 50 mM Na_2HPO_4, pH 5.0), and the absorbance of the solution was measured at 450 nm using a microtiter plate reader.

2.1.4 Shear stress-mediated refolding of IMAC resin-bound His-PKA

The catalytic subunit of PKA was overexpressed in BL21 *E. coli* with an *N*-terminal His_6 tag by induction with IPTG (1 mM). This experiment utilized the residual pellet from a 12 L culture, a waste product more typically discarded. After dissolution in lysis buffer, sonication was applied as described above. His-PKA was then denatured in 6 M guanidine-HCl, 20 mM sodium phosphate, 500 mM NaCl and incubated with Bio-Rad Ni^{2+}- charged Profinity IMAC resin for 2 h at room temperature (1 ml of a 1.72 mg/ml His-PKA to 50 µl or 250 µl bed volume IMAC). A control experiment used uncharged IMAC resin instead. The IMAC-His-PKA solution was then diluted to 1 M guanidine-HCl with binding buffer containing 1 mM imidazole, or with the elution buffer containing 500 mM imidazole as a control. This diluted solution was immediately treated in the VFD (1 ml, 5 krpm, 20 min). After transferring to a 1.5 ml eppendorf tube, the resin was washed by aliquoting 1 ml

wash buffer, inverting the tube three times, and centrifuging the tube at 2000 *x g* for 2 min to separate the beads from the supernatant. This process was repeated two additional times before elution with elution buffer containing 500 mM imidazole. For protein quantification only, samples containing 500 mM imidazole were diluted 1:100 in wash buffer to prevent residual imidazole from interfering with the BCA assay. PKA activity was determined by monitoring substrate depletion in an NADH enzyme-linked assay at 340 nm (300 µl assay volume, 10 mM ATP, 0.5 mM NADH, 1 mM phosphoenolpyruvate, 0.0153 U/µl lactate dehydrogenase, 0.0269 U/µl pyruvate kinase, 0.67 mM kemptide, 100 mM MOPS, 9 mM MgCl2, pH 7.0). Kemptide was synthesized by solid-phase peptide synthesis. All other reagents were purchased from Sigma-Aldrich.

2.2 Results

2.2.1 Modelling thin film fluid dynamics in the vortex fluid device

To accurately estimate the shear forces exerted on proteins during VFD directed folding the fluid behaviour in the VFD was modelled across various rotational speeds. Our modelling applies the solution of Couette flow which describes the laminar flow of viscous fluid in the space between two parallel plates, when one of which is moving relative to the other[9,10]. This flow is driven by the viscous drag force acting on the fluid and the applied pressure gradient parallel to the plates. The velocity of the solution, V_ϑ, is a function of the radius, *r* (Figure 2-1A). The boundary conditions for the liquid film interfaces are defined as follows. The inner air-liquid interface at $r = R_1$ slips due to discontinuity in viscosity, and results in

vanishingly low shear stress ($\frac{dv_\theta}{dr} = 0$). At the outer liquid-glass interface, the no-slip boundary dictates that the velocity of the liquid at $r = R_2$ matches that of the inner wall of the glass tube ($v_\theta = R_2 \bullet \Omega$), where Ω is the angular velocity of the tube. The resulting velocity profile is a nonlinear function of the form:

$$v_\theta = Ar + \frac{B}{r}$$

Where

$$A = \frac{\Omega}{\frac{R_1^2}{R_2^2}+1} \quad \text{and} \quad B = \frac{\Omega R_1^2}{\frac{R_1^2}{R_2^2}+1}$$

From this velocity profile, shear stress can be calculated as:

$$\tau_{r\theta} = \mu r \frac{\partial}{\partial r}\left(\frac{v_\theta}{r}\right)$$

where μ is the viscosity of water at 20 °C. At a speed of 5 krpm, the calculated shear stress ranges from 0.53 to 0.56 Pa (Figure 2-1B). These values of shear stress are similar to the requirements previously reported for protein unfolding[11].

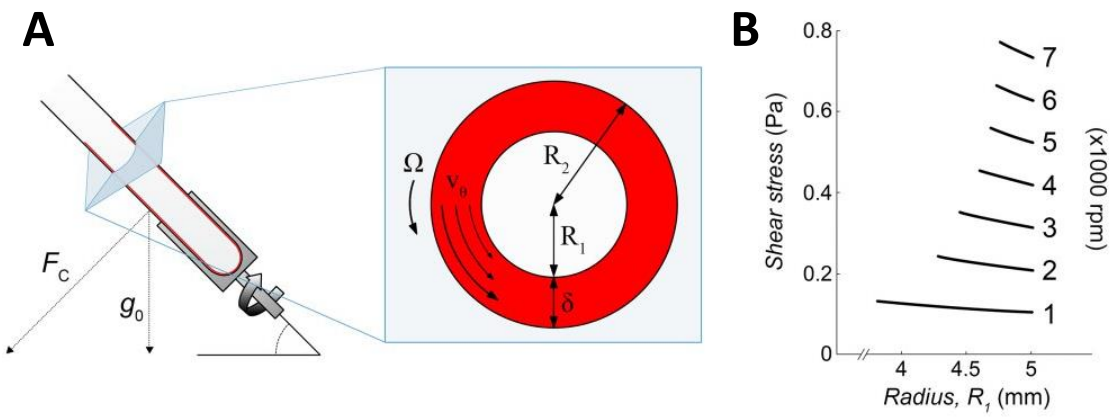


Figure 2-1. *In vitro* protein refolding with the vortex fluid device (VFD) generates shear flow inside thin fluid films. (A) Schematic of the VFD with the insert representing a two-dimensional slice of the rotating tube across the axis of rotation. The radius of the tube, R2, is 5 mm, and the thin fluid film thickness, δ, has been previously measured (not shown to scale)[12]. The velocity of the solution, v_θ, increases from the inner surface of the film to match the velocity of the tube, Ω. (B) Shear stress calculated as a function of rotational speed (rpm) and radius from the centre of the tube.

2.2.2 Refolding denatured hen egg white lysozyme from egg whites

Experiments were conducted using native hen's egg white lysozyme taken from fresh eggs. These experiments determined whether the shear forces of the VFD were able to refold denatured HEWL in a non-homogenous protein solution. Egg whites were separated from the yolk, diluted in PBS and heat-treated at 90 °C for 20 minutes. The hard-boiled egg white was dissolved in 8 M urea, rapidly diluted and then passed through the VFD at the rotational speeds and times shown in Figure 2-3. The Bicinchonic acid assay was then used to determine total protein concentrations – 44 µg/ml. Recovery of HEWL activity was then determined by lysozyme activity assay (Figure 2-2). Our results showed that refolding of HEWL within the native egg white at 5 krpm was able to recover activity after as little as 2.5 minutes of spin time and continued shear forces unfolds the protein. To yield optimal HEWL refolding, VFD processing time was found to be 5 minutes at 5 krpm

(Figure 2-3A). These results were used to establish the parameters for lysozyme refolding by VFD.

Standard	
U	RFU/min
25	0.201
12.5	0.163
6.25	0.071
3.12	0.040
1.5625	0.019
0.78125	0.017
0.390625	0.013
0	0.012

Michaelis-Menton + C	
Best-fit-values	
Vmax	0.4187
Km	0.25.92
C	0.003201
Vmax	0.1.28
Km	14.46
C	0.008513
Degrees of Freedom	5
R square	0.9735
Absolute Sum of Squares	0.001024

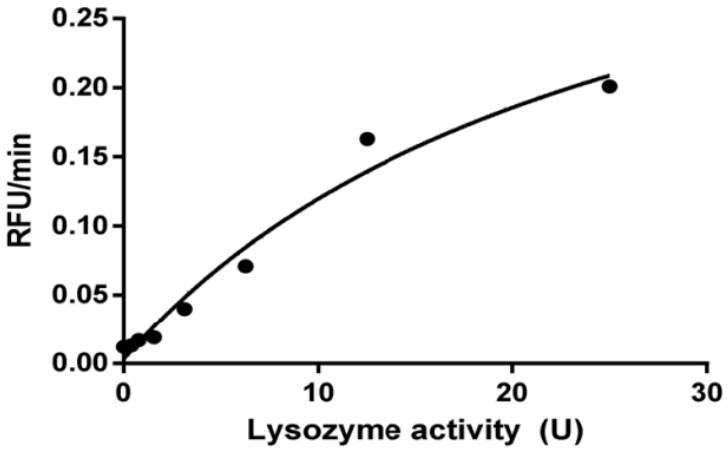


Figure 2-2. Wild-type lysozyme activity interpolated by least-squares regression curve. Relative fluorescence units per minute were fitted to standardized lysozyme activity assayed using the EnzChek Lysozyme Activity kit (Invitrogen) to a Michaelis-Menten model, plus background with Prism 6 software (GraphPad). All assays were conducted in 96-well black opaque microtiter plates, 100 µl reaction volumes, with 10 min incubation at 37 °C.

2.2.3 Refolding recombinant hen egg white lysozyme from inclusion bodies

The VFD successfully refolded recombinantly expressed, reduced HEWL from BL21 *E. coli* bacterial cultures. Cell pellets were reconstituted in lysis buffer containing 2-mercaptoethanol, purified, then urea-denatured and rapidly diluted in PBS at a

1:100 ratio (Table 2-1). 1 ml of recombinant HEWL [44 µg/ml] was immediately transferred into the VFD tube and spun at 22 ^{o}C and 5 krpm for 5 minutes. Circular dichroism spectra of the VFD-refolded protein showed characteristic curves of alpha helical secondary structure from lysozyme isolated from inclusion bodies. CD spectral analysis of identical HEWL samples demonstrated partial recovery of secondary structure after VFD processing when compared to native lysozyme (Figure 2-3B). Protein activity assay of lysozyme following VFD processing were used to determine functional yield. Under the conditions specified above, the so-called 'confined mode' of VFD processing, recombinant HEWL recovered 82% of its activity. As expected HEWL isolated from inclusion bodies that did not undergo VFD processing did not show any lysozyme activity.

The VFD has a second modality capable of operating under continuous flow (Figure 3-3D). This approach delivers additional sample through an inlet at the cylinder base. A 50 ml sample was added at a flow rate of 0.1 ml/min. The method demonstrated significant recovery of HEWL activity and even slightly exceeded recover from the confined mode with >85% of activity restored after this scalable, high volume application. This continuous flow mode is scalable for industrial applications.

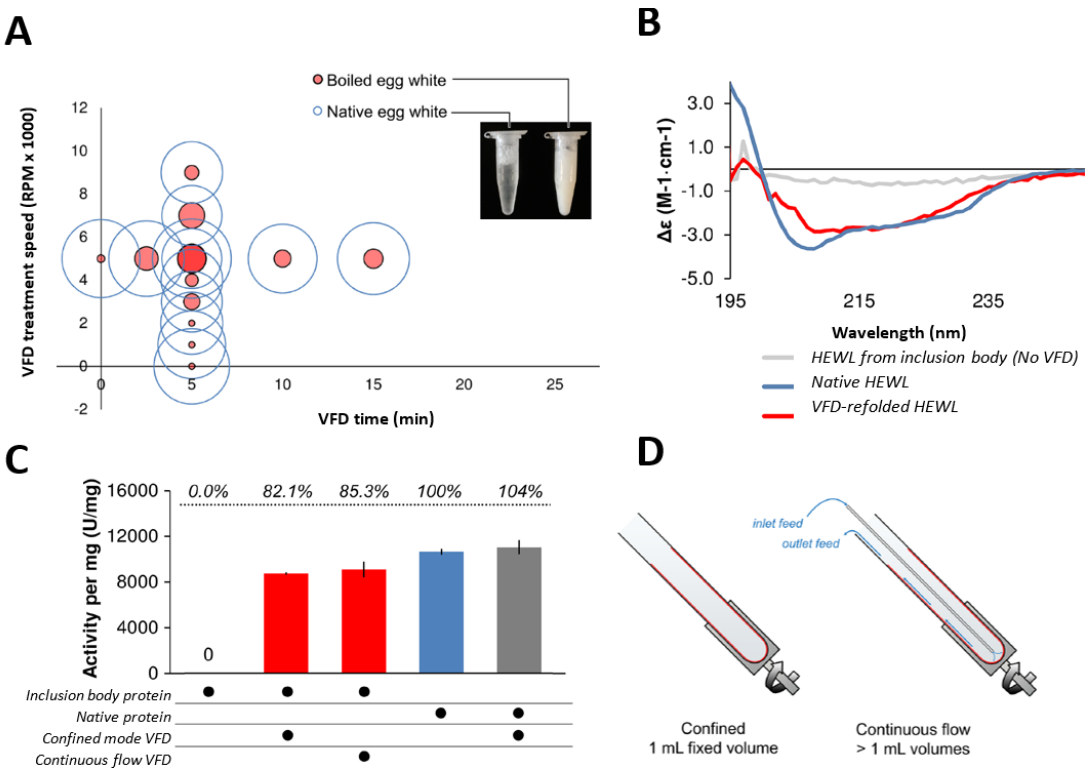


Figure 2-3. Structure and functional analysis of hen egg white lysozyme (HEWL) processed by VFD. (A) Recovered lysozyme activity per mg protein after VFD processing of boiled egg white (90 ºC, 20 min) at fixed VFD speed (5 krpm) for 5 min refolding time. The size of the solid red circle is relative to the clear, blue circle as a fraction of lysozyme activity recovered from hard-boiled egg whites following VFD induced refolding. The inset depicts 40% v/v egg white in PBS with and without 90 °C heat treatment. (B) CD spectra of recombinantly expressed, reduced, and denatured HEWL before (grey) and after (red) VFD refolding. (C) Lysozyme activity per mg protein after VFD refolding of recombinantly expressed HEWL. Here, VFD treatment of purified, active lysozyme does not adversely affect activity. Percentages quoted at top show activity recovered compared to wild-type. Error bars indicate standard deviation (n = 3). (D) Schematic diagram of VFD. In continuous flow mode, the protein solution is fed through a thin line metal tube to the bottom of the sample tube.

2.2.4 Applying shear stress to refold recombinantly expressed caveolin-ΔTM

After successfully refolding denatured lysozyme in both complex (native egg white) and simple (purified recombinant protein) environments, we sought to replicate such refolding in other protein models. Our focus shifted to refolding the protein caveolin-1. Caveolin-1 was chosen because it represents a protein that traditionally requires an especially large amount of processing time by conventional methods

(typically four days of dialysis). The variant used has had the transmembrane domain removed (caveolin-ΔTM). The protein was recombinantly expressed and the inclusion bodies purified under denaturing conditions. Purified caveolin-ΔTM was diluted and underwent rapid dialysis for 1 h to lower urea concentrations. The protein underwent VFD treatment for 0, 10, or 30 min at 5 krpm at a concentration of 186 μg/ml. CD spectra curves of the VFD-treated caveolin-ΔTM showed a distinct minima at 208 nm, which is indicative of α-helical secondary structure (Figure 2-4A)[13]. Following VFD processing there was pronounced decrease in the protein solution turbidity, a mark of VFD induced solubilisation and refolding of partially aggregated proteins (Figure 2-4B). Later, binding studies of the refolded caveolin-ΔTM to HIV glycoprotein 41 (gp41), a known caveolin binding partner[14,15], were conducted via ELISA assay. VFD processing significantly restored protein function, evident by binding to gp41 (Figure 2-4C).

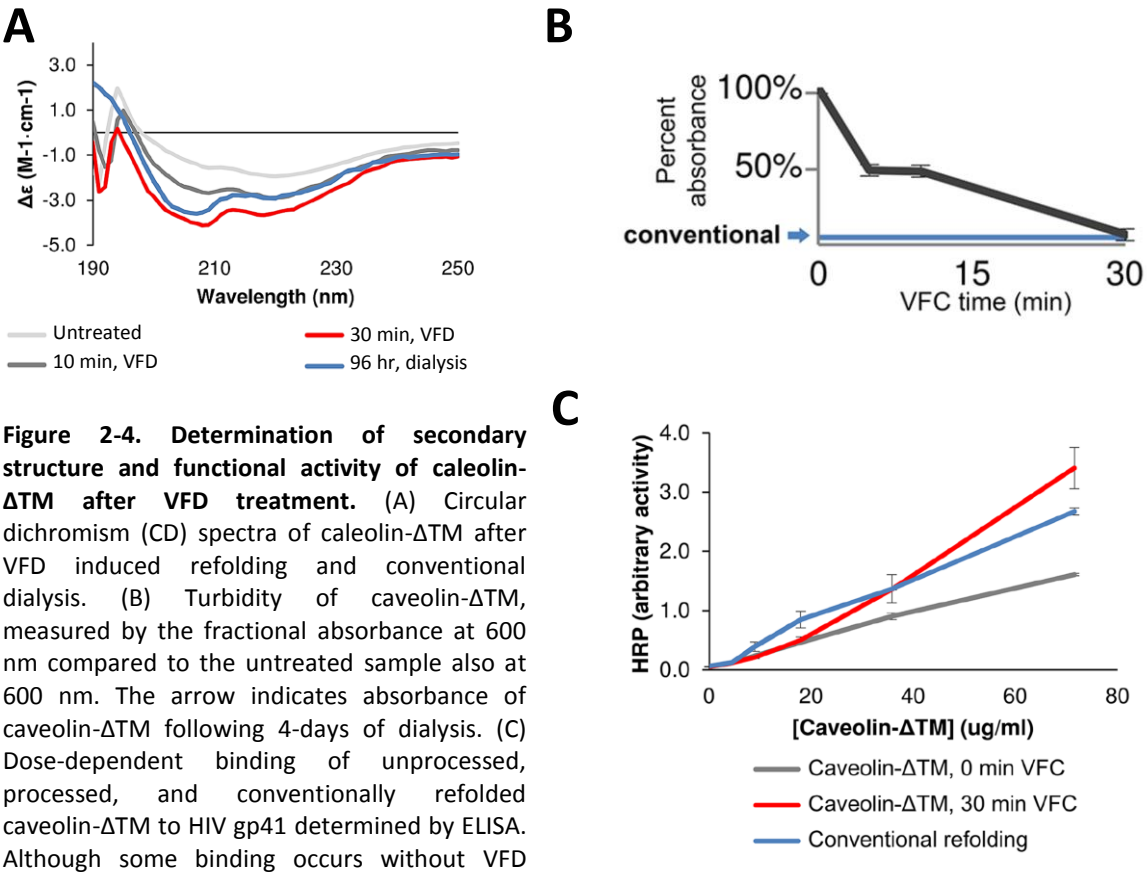


Figure 2-4. Determination of secondary structure and functional activity of caleolin-ΔTM after VFD treatment. (A) Circular dichromism (CD) spectra of caleolin-ΔTM after VFD induced refolding and conventional dialysis. (B) Turbidity of caveolin-ΔTM, measured by the fractional absorbance at 600 nm compared to the untreated sample also at 600 nm. The arrow indicates absorbance of caveolin-ΔTM following 4-days of dialysis. (C) Dose-dependent binding of unprocessed, processed, and conventionally refolded caveolin-ΔTM to HIV gp41 determined by ELISA. Although some binding occurs without VFD treatment, caveolin-ΔTM binds with greater affinity after VFD refolding. Error bars indicate standard deviation (n = 3).

2.2.5 Recovering kinase activity from PKA by attachment to IMAC beads

Initial experiments using the VFD to refold large-sized protein failed to replicate the results obtained with smaller proteins like lysozyme and caveolin. At first, tests using the catalytic domain of protein kinase A (PKA) (42 kDa), a significantly bulkier protein than HEWL (14 kDa) or caveolin-ΔTM (17 kDa), were unable to refold inclusion body protein following similar protocols. An altered protocol was devised to focus shear stress on the *N*-terminus of His-tagged PKA by way of Ni^{2+} charged immobilized metal affinity chromatography (IMAC). The IMAC-His-PKA complex was subsequently subjected to VFD shear stresses (1 ml, 0.2 – 1 mg/ml). After VFD

processing the His-PKA that had been separated from IMAC resin was able to recover 69% of kinase activity (Figure 2-5A). Negative control experiments with identical conditions, but with uncharged IMAC resin, lower charged resin volumes, or 500 mM imidazole to block the Ni^{2+}-His_6 tag interaction, showed only low levels of kinase activity (Figure 2-5A). It is worth noting that remaining His-PKA eluted from the IMAC resin with imidazole yielded considerably less active protein (Figure 2-5B).

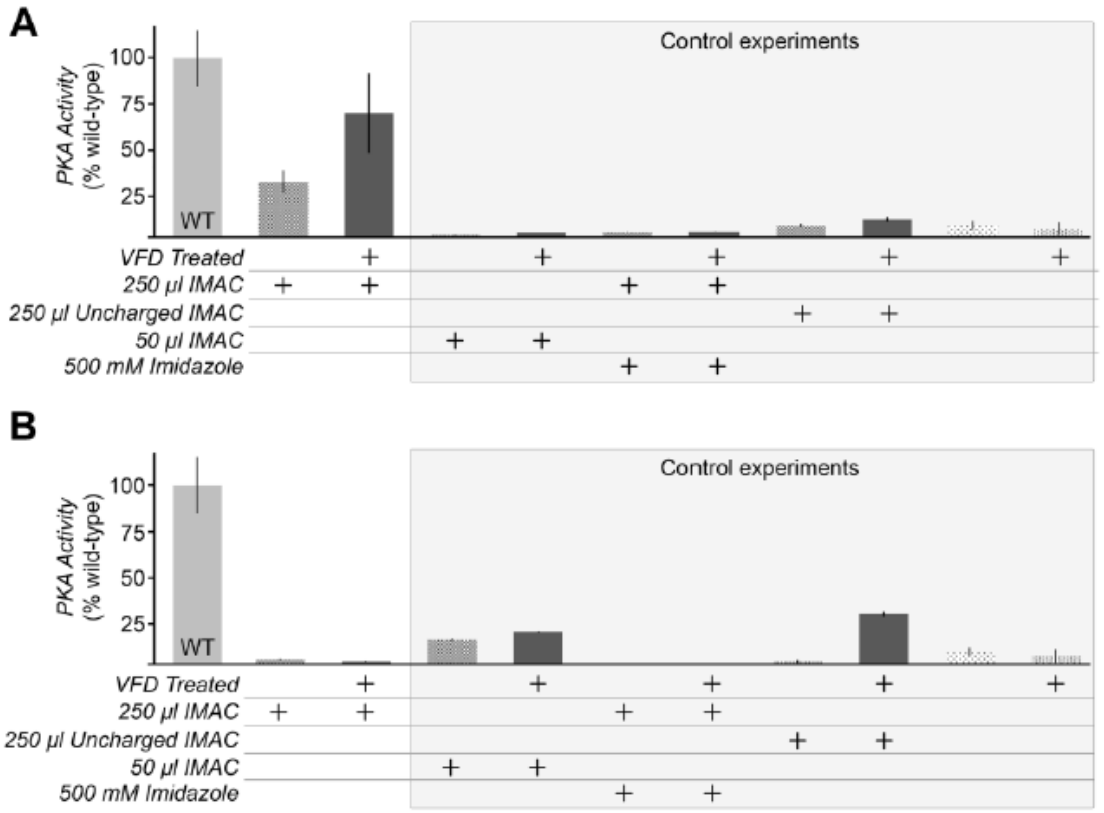


Figure 2-5. Facilitating VFD refolding of PKA by pre-binding to IMAC resin. 1.7 mg PKA was pre-incubated with 250 or 50 µl IMAC resin in 6 M guanidine-HCl prior to dilution to 1 M guanidine-HCl and VFD treatment. (A) PKA was removed by a low imidazole (1 mM) wash buffer. PKA activity per µg of protein measured by nicotinamide adenine dinucleotide (NADH) enzyme-linked assay and shown as percentage of wild-type PKA. Consumption of ATP by active PKA results in consumption of NADH by lactate dehydrogenase. The NADH levels were monitored in this assay by absorbance at 340 nm. Negative control experiments were performed with low quantities of resin, 500 mM imidazole, or resin lacking Ni^{2+} (uncharged) during IMAC incubation. (B) PKA activity per µg of protein of elution fractions following VFD treatment. Steps performed on elution fractions were identical to wash fractions. Error bars indicate standard deviation (n = 3).

The results presented in this chapter identify that protein refolding using VFD processing requires optimization for each protein. Buffers, protein concentration, and processing time were optimized for HEWL, caveolin-ΔTM, and PKA. It appears that complex protein mixtures (like those in the boiled egg whites used) recover folded proteins from inclusion bodies less efficiently when compared to simple solutions. The mechanical energy of the VFD could be misdirected into the other >96% of proteins present in natural egg whites.

We report here a new paradigm in laboratory recombinant protein expression. We have applied shear force to drive rapid equilibration of protein folding. This kind of reaction has long been recognised in nature by cell chaperonins but has never before been accomplished in biochemistry. The VFD also mimics chaperonin function to avoid reaggregation through cage-like isolation during refolding but instead relies upon shear forces of rapidly rotating thin fluid films to maintain protein separation[16,17].

2.3 Discussion

The results reported in this chapter offer a significant contribution to the way we will approach recombinant overexpression of proteins in the future. We describe a paradigm-shifting new method that offers considerable advantages over conventional approaches to protein refolding. The VFD-mediated refolding can be considered to be an important advancement in green chemistry protocols as it represents a method that requires much smaller solution volumes of approximately 1% of the volumes necessary for conventional dialysis. Secondly, our novel protocol alters a key step in protein production to occur >100-times faster than the industry standard of overnight dialysis while simultaneously improving protein production by as much as >1000-fold for proteins such as caveolin. Our method distinguishes itself from others by the use of the VFD, which utilizes high shear forces in thin fluid films to direct protein refolding. The VFD-mediated shear force is a low-energy, inexpensive form of processing that has proven applicable to many different protein models.

We have presented VFD-mediated protein folding in four different systems. Our data has shown the VFD to be highly successful at restoring function to recombinantly expressed smaller proteins such as hens egg white lysozyme isolated from inclusion bodies as well as caveolin-Δ TM. When applied to larger-sized proteins the VFD encountered initial difficulties at restoring the catalytic domain of PKA protein. We hypothesized that to correctly refold the full-length PKA *in vitro* we must mimic the cellular folding process more closely. Rather than trying to refold the entire PKA protein at once we focused shear stress on the *N*-terminus of His-

tagged PKA by immobilisation on Ni^{+2}-charged immobilised metal affinity chromatography (IMAC) beads. Cells do this in nature when the nascent polypeptide is able to fold while the *N*-terminus extrudes from the ribosome. This adjustment proved successful and we were able to recover 69% of the protein's kinase activity. We also recognised that VFD folding is most effective when directed towards purified protein samples. VFD folding in complex protein mixtures appears to share the force across all the protein present. When testing a mixture of boiled egg whites the mechanical energy of the VFD was not solely directed to the lysozyme content of the egg white and instead spread throughout the other proteins present in solution.

The advantages of VFD refolding expose new possibilities to increase protein yields for recombinantly overexpressed proteins in simple cell lines. The VFD has proven capable of untangling complex mixtures, insoluble inclusion bodies and aggregates. The potential for VFD-mediated refolding to be incorporated into laboratory standards could be seen when used with high concentration of a chemical inducer such as IPTG. Such a combination would drive overexpressed proteins into insoluble inclusions that can then be harvested after VFD processing to give an overall increase in yield. Conventional approaches seek to avoid inclusion bodies through painstaking optimization of growth conditions and specialised cell lines. This comes at the expense of larger yields as well as protein purity.

The continuous flow mode of the VFD easily allows for process up-scaling to accommodate much larger volumes of solutions. We offer a path to lower financial costs and shorter timelines required to refold inactive proteins that are conducive

to an industrial scale. The use of mechanical shear forces to drive chemical reactions is a relatively new endeavour and this chapter has highlighted the need to understand and harness shear forces to achieve rapid equilibrium of protein folding. Our methods could certainly be expanded to a wide-range of applications in research and manufacturing.

REFERENCES

1. Meyer H-P, Schmidhalter D. Innovations in Biotechnology. In Tech, Rijeka; 2012.
2. Futaki S. Membrane-permeable arginine-rich peptides and the translocation mechanisms. Advanced drug delivery reviews. 2005;57(4):547-558.
3. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome biology. 2003;4(9):117.
4. Radford SE. Protein folding: progress made and promises ahead. Trends in biochemical sciences. 2000;25(12):611-618.
5. Jemth P, Gianni S, Day R, Li B, Johnson CM, Daggett V, et al. Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(17):6450-6455.
6. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol. 2008;6(8):e195.
7. Vallejo LF, Rinas U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microbial cell factories. 2004;3(1):11.
8. Tsumoto K, Ejima D, Kumagai I, Arakawa T. Practical considerations in refolding proteins from inclusion bodies. Protein expression and purification. 2003;28(1):1-8.
9. Mallock A. Determination of the viscosity of water. Proceedings of the Royal Society of London. 1888;45(273-279):126-132.
10. Rayleigh L. On the dynamics of revolving fluids. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character. 1917;93(648):148-154.
11. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324-332.
12. Bekard IB, Asimakis P, Bertolini J, Dunstan DE. The effects of shear flow on protein structure and function. Biopolymers. 2011;95(11):733-745.
13. Chen X, Dobson JF, Raston CL. Vortex fluidic exfoliation of graphite and boron nitride. Chemical Communications. 2012;48(31):3703-3705.
14. Saxena VP, Wetlaufer DB. A new basis for interpreting the circular dichroic spectra of proteins. Proceedings of the National Academy of Sciences. 1971;68(5):969-972.
15. Hovanessian AG, Briand J-P, Said EA, Svab J, Ferris S, Dali H, et al. The caveolin-1 binding domain of HIV-1 glycoprotein gp41 is an efficient B cell epitope vaccine candidate against virus infection. Immunity. 2004;21(5):617-627.
16. Wang XM, Nadeau PE, Lo Y-T, Mergia A. Caveolin-1 modulates HIV-1 envelope-induced bystander apoptosis through gp41. Journal of virology. 2010;84(13):6515-6526.
17. White DA, Buell AK, Knowles TPJ, Welland ME, Dobson CM. Protein aggregation in crowded environments. Journal of the American Chemical Society. 2010;132(14):5170-5175.

CHAPTER 3

Sensitizing basal-like breast cancer to chemotherapy using nanoparticles conjugated with interference peptide

Abstract

Basal-like breast cancers are a highly aggressive malignant subtype of breast cancer that are associated with very poor prognosis. Typically, these cancers are high grade 'triple negative' tumours that express no oestrogen receptors, progesterone receptors, or Her2 proteins. Commonly, these cancers may initially respond positively to first-line treatment only to become highly resistant to standard chemotherapy later in a metastatic setting. The *Engrailed 1* (EN1) transcription factor plays an important role in the development of basal-like breast cancer drug resistance through highly selective overexpression of the gene homeobox. In this chapter we propose a novel use of poly(glycidyl methacrylate) nanoparticles, decorated with poly(acrylic acid), that facilitate the dual delivery of docetaxel and our own interference peptide (iPep) designed specifically to block, or inhibit, EN1 (EN1-iPep). EN1-iPep is shown to be highly selective in inducing apoptosis in basal-like cancer cells with negligible effects in non-neoplastic human mammary epithelial cell lines. Additionally, treatment with EN1-iPep results in a highly synergistic interaction with docetaxel to inhibit cancer cell growth. By incorporating these two pharmacological agents in a single nanoformulation, we have produced a greater anticancer efficiency than current nanoparticle-based therapies clinically used today.

Introduction

Triple negative breast cancers (TNBC) are hyper-aggressive malignancies observed in approximately 15% of all breast cancer diagnoses. They are recognised clinically and phenotypically by their lack of key cell receptors; oestrogen, progesterone, and the human epidermal growth factor receptor-2 (Her2)[1,2]. As a consequence, TNBCs fail to respond to conventional endocrine and anti-HER2 therapies when treating breast cancer. Currently, there are no approved therapies tailored against TNBCs. With no tumour receptor to target, patients with TNBCs often undergo non-specific chemotherapy after initial diagnosis, followed by an adjuvant chemotherapy using varied compounds such as taxanes, platinum agents, and PARP inhibitors[3]. TNBCs most often belong to the basal-like subtype characterised by the presence of P53 and BRCA1 mutations[4]. Basal-like tumours initially respond positively to chemotherapy, however, in a metastatic setting they exhibit drug resistance and tumour relapse similar to what is observed in TNBCs.

An important feature of basal-like breast tumours is the aberrant overexpression of the Engrailed 1 (EN1) transcription factor. EN1 overexpression promotes proliferation, metastasis, and protects basal-like tumours against apoptosis; thus increasing drug resistance[5]. Previously, our laboratory has demonstrated that the selective knockdown of EN1 induces potent capase-3 apoptosis and a sensitisation in basal-like breast cancer cells to chemotherapy[5]. This discovery means that a target may exist for new selective therapies against basal-like breast cancers in the form of selective EN1 activity inhibitors. Previously, transcription factors (like EN1) have been thought of as ‘undruggable’ targets due

to their small molecular binding pockets. The Blancafort laboratory has designed an interference peptide (iPep) small enough to interrupt the transcription factor behaviour of the EN1[5]. The EN1-iPep inhibits transcription by binding EN1 and in doing so, prevents it from binding its partner molecules and DNA.

Docetaxel (DTX) is a member of the taxoid group of anticancer agents and is routinely used in the treatment of metastatic breast cancers[6]. Taxanes arrest cell cycling through the promotion and stabilisation of microtubule assembly during S-phase, leading to apoptosis[7]. New formulations of nanoparticle-based cargo delivery of DTX have produced promising results, demonstrating greater anti-tumoural effects with reduced off-target effects when compared to free drug delivery in both *in vitro* and *in vivo* models. The same effect has been demonstrated with doxorubicin - another staple drug used to treat metastatic breast cancers[8-11]. Another advantage of nanoparticle drug delivery is a greater systemic circulation half-life. The BRCA1 mutation, a hallmark of basal-like breast cancer, has now been recognised to be an important predictive factor for drug resistance. To date, most drug regimens have been heavily reliant on neoadjuvant therapies to address resistance to docetaxel treatment and most chemotherapies approaches[12]. These have produced limited success.

Here in this chapter, we report the use of poly(acrylic-acid)-decorated poly(glycidyl methacrylate) nanoparticles. These nanoparticles encapsulate docetaxel and electrostatically assemble EN1-iPeps creating a dual delivery platform against basal-like breast cancer. We demonstrate the efficiency of this approach to sensitise basal-like breast cancers to chemotherapy. Using systemic

assessment of individual agents, we demonstrate high synergism in our combinatorial approach, and superior anticancer efficacy when compared to alternate clinical nanoparticle-based treatments.

3.1 Materials and methods

3.1.1 Synthesis and functionalisation of nanoparticles

Poly(glycidyl methacrylate (PGMA) (Mw = 454 270, PDI = 1.79) was synthesised using a free-radical polymerization process. Glycidyl methacrylate was dissolved in MEK and the reaction heated to reflux under an inert atmosphere using 2,2'-azobis(2-methylpropionitrile) as an initiator. PGMA mixture was cooled, then precipitated in methanol and collected by filtration. PAA-PGMA-DTX nanoparticles were prepared according to a modified protocol described previously[13]. 100 mg PGMA and 15 mg docetaxel were dissolved in a 1 : 3 mixture of chloroform and MEK. 6 ml of the mixture was added dropwise into an aqueous solution of Pluronic F-108 (1.25% w/v, 30 ml) under vigorous stirring. The emulsion was homogenised with a probe-type ultrasonicator set at 4 WRMS for 1 min. Organic solvents were removed from the emulsion under reduced pressure. The solution was centrifuged at 3000 *x G* for 45 min to remove large aggregates. The supernatant was further centrifuged at 20 000 *x G* (15 min) to isolate nanoparticles from excess surfactant. Nanoparticles were resuspended in 10 ml of PAA solution (100 mg ml^{-1}, pH 9.0), assisted by short periods of ultrasonication at 4 WRMS. The suspension was stirred at 70 °C overnight to conjugate PAA to the PGMA nanoparticles. The resultant nanoparticles were centrifuged and resuspended in PBS immediately before use. EN1-iPeps, both active and mutant, were electrostatically attached to PGMA-PAA nanoparticles by incubation at room temperature in PBS (pH 8.0) overnight.

3.1.2 Characterisation of nanoparticles

Hydrodynamic size and zeta potential was measured using a Zetasizer Nano ZS (Malvern Instruments). Nanoparticle stability was assessed in an environment analogous to the *in vivo* physical environment. This meant PAA-PGMA-DTX and EN1-iPep-PAA-PGMA-DTX nanoparticle, size, polydispersibility index (PdI), and zeta potential, were determined in the presence of BSA, resuspended in PBS. Fourier transform infrared spectroscopy (FT-IR) determined chemical characterisation of DTX-NPs. FT-IR of PGMA nanoparticles containing DTX shows a peak at 710 cm^{-1} characteristic of DTX.

3.1.3 Determination of docetaxel loading in nanoparticles

PAA-PGMA-DTX nanoparticles were lyophilised and weighed before suspended in methanol (1 ml). The suspension was vortexed regularly over 1 h to assist in the dissolution of encapsulated docetaxel. The nanoparticles were removed via centrifugation (14 000 *x G*, 15 min) with the supernatant containing the dissolved free drug, analysed using RP-HPLC. Measurements were performed on a Waters 2695 separation module with Waters 2489 UV/Vis detector (determination λ = 232 nm) using reverse phase isocratic elution (methanol : water, 70 : 30; 1.5 ml/min^{-1}) through a C18 column (150 × 4.60 mm, 5 µm, 25 ± 5 °C). The measurements were compared against a standard curve done in identical run conditions. Docetaxel loading in the PAA-PGMA-DTX nanoparticles was calculated with the following formula:

$$Docetaxel\ loading\ \%\ (w/w)$$

$$= \frac{mass\ of\ docetaxel\ within\ nanoparticles}{mass\ of\ nanoparticles} \times 100$$

3.1.4 Docetaxel release from nanoparticles

Docetaxel release was quantified according to the method described in Singh et, al[14]. Nanoparticles (1.00 ml, 7.26 mg ml^{-1}) were placed in a dialysis membrane (Mw cutoff 100 KDa) and dialysed against a sink solution (15 ml) containing either PBS or PBS + 1% w/v Tween 80 in a water bath at 37 °C. Aliquots (10 ml) were drawn from the sink solution at each time point, freeze dried, extracted with methanol (200 µl), and quantified by HPLC (isocratic elution with 70 : 30 methanol/water; C18 column, 150 × 4.60 mm, 5 µm, 25 °C; retention time ∼10 min). At each time point, 10 ml fresh solution was added to the sink to maintain flux conditions.

3.1.5 Cell culture

The T11 basal-like breast cancer cell line was derived from the murine basal-like breast tumour of a $p53^{-/-}$ $Brca1^{-/-}$ BALB/c mouse[15,16]. The basal-like breast cancer cell line SUM149 and the normal breast epithelium cell line MCF10A were purchased from Asterand (Detroit; MI, USA) and the American Type Culture Collection (Manassas; VA, USA), respectively. T11 and SUM149 were cultured in RPMI and DMEM-F/12 (Life Technologies; VIC, Australia) medium respectively. Both media were supplemented with 10% foetal bovine serum (Life Technologies) and 1% penicillin/streptomycin (Life Technologies). MCF10A were cultured in a DMEM-

F/12 supplemented with 10% horse serum, 20 ng ml^{-1} EGF, 0.5 µg ml^{-1} hydrocortisone, 10 mg ml^{-1} insulin, 100 ng ml^{-1} cholera toxin and 1% penicillin/ streptomycin (Life Technologies).

3.1.6 Cell viability and assessment of apoptosis

To assess the selective activity of EN1-iPeps, T11, SUM149 and MCF10A cells were treated with EN1-iPep$_{act}$ and EN1-iPep$_{mut}$ for 24 h at concentrations up to 50 µM. To assess the efficacy of EN1-iPep nanoparticles as compared to the commercial chemotherapeutic Abraxane® (paclitaxel-albumin nanoparticles), T11 cells were treated with different concentrations of uncoated DTX nanoparticles (without EN1-iPep; blank), EN1-iPepact-DTX nanoparticles and Abraxane® at up to 0.5 mg/ml^{-1} of nanoparticles, equivalent to a range of EN1-iPep up to 36.3 µM, for 24 and 48 h. Cell viability for the aforementioned experiments was assessed using the CellTiter-Glo assay according to the manufacturer's protocol (Promega; NSW, Australia). Luminescence was measured using the EnVision 2012 Multilabel Reader (PerkinElmer; Waltham, MA, USA). IC50 of EN1- iPeps against the various cell types were calculated using the Graphpad Prism 6 statistical package.

Apoptosis was assessed by means of immunofluorescence visualization of cleaved-caspase 3 positive cells. T11 and SUM149 cells were treated with active or mutant EN1-iPep at 50 µM for 8 h. Following treatments, cells were fixed with 4% paraformaldehyde for 20 minutes, washed twice with PBS, blocked with blocking solution (5% fetal bovine serum, 0.3% Triton X-100 in PBS) for 1 h, incubated with the anti-cleaved caspase 3 primary antibody (1 : 500 dilution; Cell Signaling

Technology, QLD, Australia) in blocking solution overnight and further incubated with an anti-rabbit secondary Alexa Fluor 488-conjugated antibody (Cell Signaling Technology, QLD, Australia). Nuclei were stained with Hoechst 33258. The percentage of positive cleaved-caspase 3 cells was determined by counting green fluorescent cells versus total cells using a fluorescent microscope (Olympus IX71).

3.1.7 Assessment of drug combination effect

The combined effect of doxorubicin and docetaxel with EN1-iPeps was assessed by the median effect method[17] proposed by Chou and Talalay, using the Compusyn software. T11 cells were treated for 24 h with doxorubicin, docetaxel, and EN1-iPepact, in non-constant ratios at doses up to 25 µM, 10 µM, and 10 µM respectively. Cell viability was assessed using CellTiter-Glo as described previously. This method determines the combination index (CI) between two agents. Additive, antagonistic or synergistic interactions between two agents are indicated by CI = 1, >1 or <1 respectively.

3.1.8 Confocal imaging

T11 cells were grown on 12 mm diameter glass coverslips and treated with rhodamine B-conjugated blank, EN1-iPep$_{act}$ or EN1-iPep$_{mut}$ coated nanoparticles at 0.33 mg ml^{-1} in serum free media for 4 h. This concentration of nanoparticles corresponds to the amount of EN1-iPep$_{act}$ attached to the nanoparticles calculated experimentally to reach the IC50 (24 µM) when administered in T11 in a

nanoparticle format. Cells were fixed with 4% paraformaldehyde and washed twice with PBS. Nuclei and actin filaments were stained with 5 μg ml^{-1} Hoechst 33258 and Alexa Fluor 488 phalloidin (Life Technologies) respectively for 20 min. Coverslips were mounted on slides using Slowfade mounting media (Life Technologies). Slides were visualised by confocal microscopy (Hoechst λ_{ex} = 397 nm, λ_{em} = 400–468 nm; Alexa Fluor 488 λ_{ex} = 488 nm, λ_{em} = 497–546 nm; rhodamine B λ_{ex} = 561 nm, λ_{em} = 579–676 nm; Leica TCS SP2).

3.1.9 Animal model and *in vitro* efficacy assessment of EN1-iPep nanoparticles

All animal experiments were performed in accordance with protocols approved by the Animal Ethics Committee of The University of Western Australia. To simulate an advanced model of basal-like breast cancer, T11 cells (2.5 × 104 cells) were resuspended in serum free media and BD Matrigel Matrix High Concentration (BD Bioscience, NSW, Australia) in a 1 : 1 ratio to a total volume of 100 μl and injected subcutaneously into the flanks of 5 week old BALB/cJ females (Animal Resources Centre, WA, Australia) using a 26G needle. Nine days after inoculation of cells, tumours reached an average volume of 279 mm^3. At this point, animals were randomly assigned to one of four different groups. All animals received a total of 5 intratumoural injections every two days containing 2.5 mg of nanoparticles coated with 0.5 mg EN1-iPep diluted in 100 μl of saline solution. The four treatments were: blank nanoparticles with EN1-$iPep_{mut}$ or EN1-$iPep_{act}$, and docetaxel nanoparticles with EN1-$iPep_{mut}$ or EN1-$iPep_{act}$. Width and length of tumours were measured every day using a digital calliper and tumour volumes were calculated using the modified

ellipsoid formula: $V = Width^2 \times \frac{1}{2}\, Length$. Animals bearing tumours >800 mm^3 were humanely sacrificed.

3.1.10 Immunohistochemical analysis of tumours

Tumour tissue was fixed in 4% paraformaldehyde, washed 3 times in PBS and left in 70% ethanol. Tumours were embedded in paraffin and 5 µm sections were prepared. For Haematoxylin/ eosin staining, slides were dewaxed, hydrated using a decreasing solution bank of ethanol, stained with Gill's haematoxylin, dehydrated using 70% ethanol, stained with eosin, further dehydrated using 100% ethanol, cleared using toluene and mounted in coverslips using Acrymount IHC mounting media (McKinney, Texas, USA). Tumour cell apoptosis was determined in tissue sections by TUNEL assay (In Situ Cell Death Detection Kit; Roche, VIC, Australia).

3.1.11 Statistical analysis

All *in vitro* experiments were performed in triplicate and repeated at least three times. Results were averaged and the standard deviation (SD) or standard error of the mean (SEM) was calculated as indicated in the figures. To determine statistical significance, a two-tailed unpaired Student's *t* test was used between 2 different independent groups. * indicates $p < 0.05$, ** indicates $p < 0.005$ and *** indicates $p < 0.0005$.

3.2 Results

3.2.1 *In vitro* toxicity assays in normal and basal-like breast cancer cells treated with EN1-iPeps

The specificity and efficiency of EN1-iPep to target basal-like breast cancer cells was assessed using *in vitro* cell viability assays. The selectivity of EN1-iPep for basal-like breast cancer cells over healthy breast cells was performed using comparative results from CellTiter-Glo assay from T11 and SUM149 basal-like cell lines and the normal breast epithelium cell line MCF10A. Cells were treated with increasing concentrations of EN1-iPep$_{act}$ and EN1-iPep$_{mut}$ (0 – 50 μM) for 24 h and it was observed that the EN1-iPep$_{act}$ induced cancer cell death in murine claudin-low T11 cells, derived from p53$^{-/-}$ transgenic mice, as well as the human basal-like SUM149 cells (Fig 3-1A and B). IC50 values for EN1-iPep in T11 and SUM149 cell lines were 9.26 μM and 13.02 μM respectively. The same induced cell death effect was not observed in normal human breast epithelial cells (MCF10A) (Fig 3-1C).

An immunofluorescence assay was used to detect levels of cleaved capase-3 (a hallmark of apoptosis) in T11 and SUM149 cells treated with 50 μM of EN1-iPep$_{act}$ and EN1-iPep$_{mut}$ for 8 h. The same conditions were also used for cell viability assay to assess cell death percentages. The results showed that 50 μM of EN1-iPep$_{act}$ induced predominantly caspase-3-dependent apoptosis (Fig 3-1D and E). Immunofluorescence images of treated cells exhibited typical apoptotic features such as nuclei condensation.

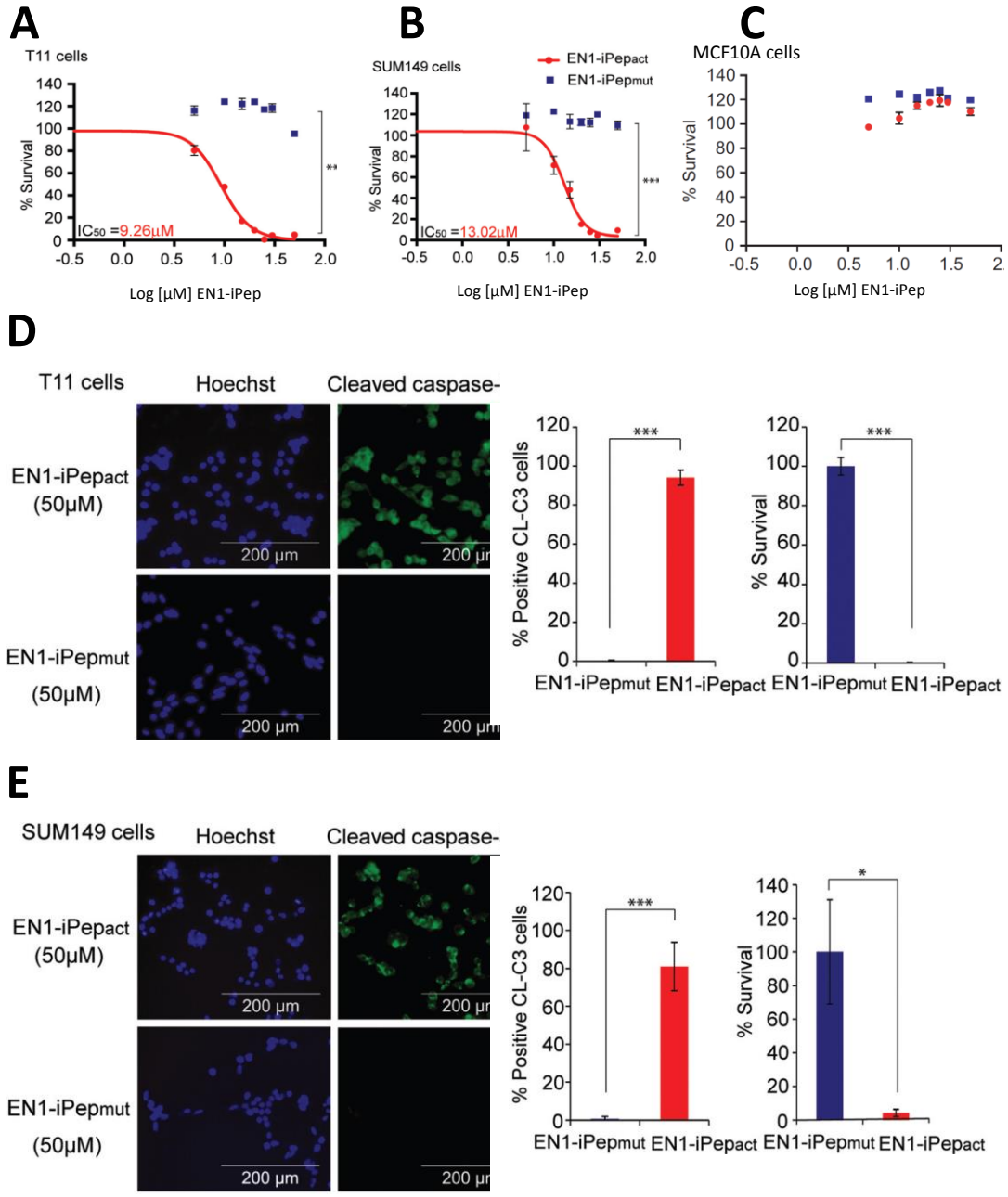


Figure 3-1. *In Vitro* toxicity assays in normal and basal-like breast cancer cells treated with EN1-iPeps. (A) Survival plot (%) of basal-like breast cancer cells (T11), (B) SUM149 and (C) immortalised mammary epithelial (MCF10A) cell lines treated with increasing concentrations of active and mutant form EN1-iPep for 24 hours. (D) Immunofluorescence assay showing cleaved caspase-3 levels in T11 and (E) SUM149 cells treated with 50 µM of active and mutant form EN1-iPeps for 8 hours. Percentages of cleaved caspase-3 positive cells and dead/alive cells shown to the right. Student's t-test $*p < 0.05$, $**p < 0.005$, $***p < 0.0005$.

3.2.2 Assessment of synergistic effects between docetaxel, doxorubicin and EN1-iPeps

Given that our overarching aim is to sensitise TNBCs to chemotherapy, we set out to examine the nature of how the EN1-iPep$_{act}$ would interact with existing anticancer drugs. Two common chemotherapeutics used to treat metastatic breast cancer are doxorubicin and DTX. Their interaction with the EN1-iPep was measured using the median dose effect method in triple negative breast cancer cells (T11 cells)[17]. This method derives a combination index (CI) that is a valuable indication of the nature of interactions between two agents. Synergistic effects were tested for using T11 cells treated with increasing concentrations of DTX alone (0 – 10 μM), doxorubicin alone (0 – 25 μM), EN1-iPep$_{act}$ alone (0 – 10 μM) and the respective combined drug-iPep formulations for 24 h. Following treatment cell viability and mortality (as a fraction affected) were determined by CellTiter-Glo (Fig 3-2). The CI value, obtained with the Compusyn software, scored lower than 1 for all combinations of DTX and EN1-iPep$_{act}$ tested, indicating a synergistic interaction between these two agents. In contrast, the combination of doxorubicin and EN1-iPep$_{act}$ resulted in an antagonistic effect (CI > 1). Based on these results, DTX was selected for inclusion into the nanoparticle formulation.

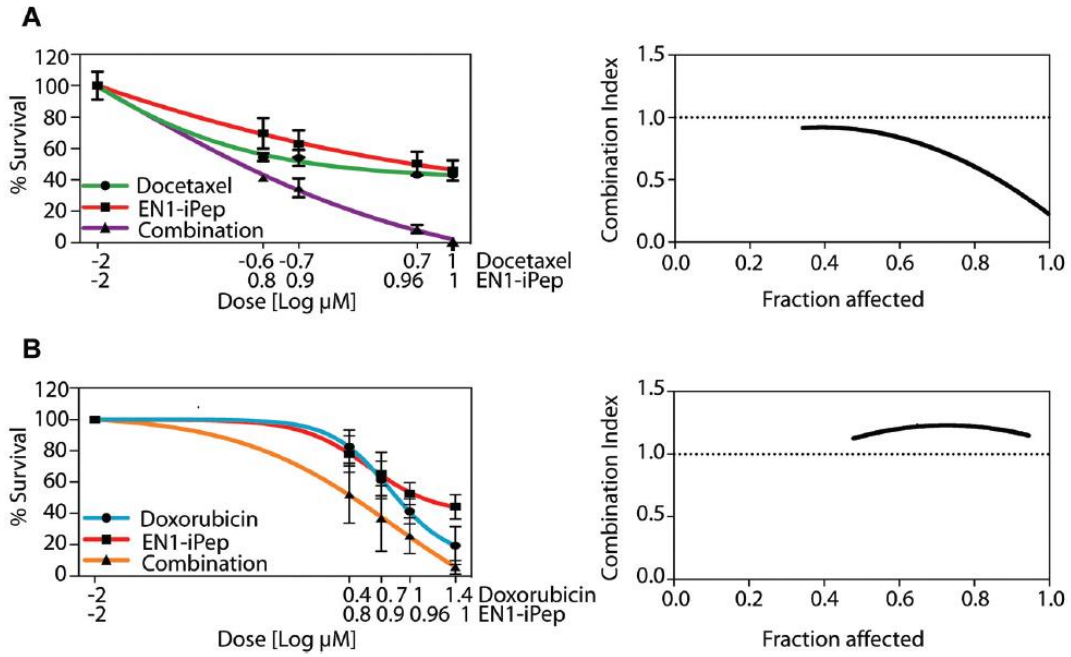


Figure 3-2. Determination of synergism between docetaxel, doxorubicin with EN1-iPep. (A) Left: Survival plot (%) of T11 cells treated with different concentrations of DTX, EN1-iPep$_{act}$ and the combination of both. Right: Combination index obtained for different fractions affected. The solid line indicates the combination index calculated by the median dose effect method. (B) Left: Survival plot (%) of T11 cells treated with different concentrations of doxorubicin, EN1-iPep$_{act}$ and the combination of both. Right: Combination index plot.

3.2.3 Characterization of nanoparticles

The nanoparticle used in the current study consisted of a poly- (glycidyl methacrylate) (PGMA) core. These nanoparticles were additionally functionalized with rhodamine B (RhB) for confocal imaging. The DTX-loaded PGMA nanoparticles were prepared using an oil-in-water emulsification process, which yielded nanoparticles with a Z-average hydrodynamic diameter of 160.9 ± 0.8 nm (PDI: 0.08). Precise hydrodynamic size diameters and zetapotential surface charge of each nanoparticle formulations are summarised in Table 3-1. Nanoparticles had 3% (w/w) DTX drug loading determined using HPLC. PAA was covalently bound to the DTX-loaded PGMA nanoparticles to provide an anionic surface suitable for

electrostatic conjugation with the positively charged EN1-iPep. Binding efficiencies (by mass ratio) of the EN1-iPep$_{act}$ and EN1-iPep$_{mut}$ to various nanoparticle formulations are summarized in Table 3-2. Incubation of EN1-iPep with the nanoparticles was monitored by a positive shift in the zeta potential (Fig. 3-3) indicating successful attachment of the EN1-iPep on the nanoparticle surface. A schematic representation of the EN1-iPep nanoparticle containing DTX has been provided in Figure 3-4.

Table 3-1. Hydrodynamic diameter and zetapotential surface charge of different nanoparticle formulations.

Nanoparticle Formulation	*Size*	*PdI*	*Zeta potential*
PGMA-DTX NPs	135.3 nm	0.087	-14.2 mV
PAA-PGMA-DTX NPs	167.9 nm	0.135	-34.2 mV
PAA-PGMA-DTX NPs + BSA	138.1 nm	0.092	-2.2 mV
EN1-iPep- PAA-PGMA-DTX NPs + BSA	162.4 nm	0.159	(not measured)

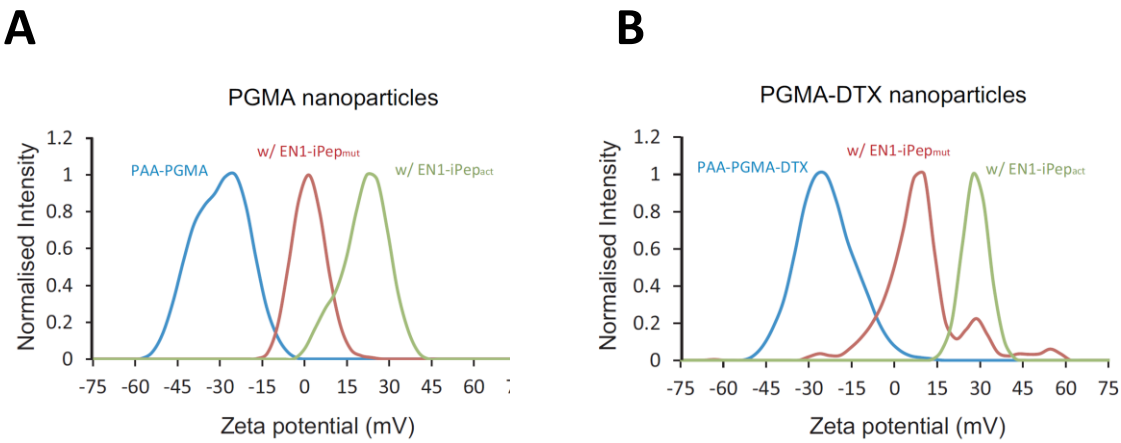
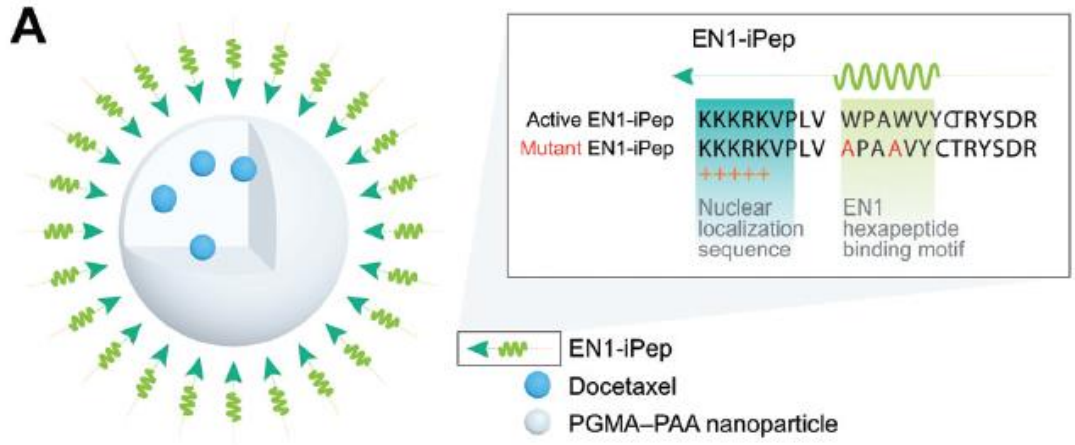


Figure 3-3. Zeta potential of PGMA nanoparticles and associated shifts following EN1-iPep incubation with active and mutant form. (A) Zeta shifts with empty PGMA nanoparticles. (B) Zeta shifts with DTX loaded PGMA nanoparticles.

Table 3-2. Binding efficiencies (by mass ratio) of the EN1-iPep$_{act}$ and EN1-iPep$_{mut}$ to various nanoparticle formulations.

Nanoparticle Formulation	*Binding efficiency (% mass)*	*Standard deviation*
EN1-iPep$_{act}$ PAA-PGMA	16.8	± 2.5
EN1-iPep$_{act}$ PAA-PGMA-DTX	20.1	± 3.7
EN1-iPep$_{act}$ PAA-PGMA-RhB	14.2	± 1.7
EN1- iPep$_{mut}$ PAA-PGMA	16.2	± 1.6
EN1- iPep$_{mut}$ PAA-PGMA-DTX	23.6	± 0.4
EN1- iPep$_{mut}$ PAA-PGMA-RhB	22.7	± 1.8



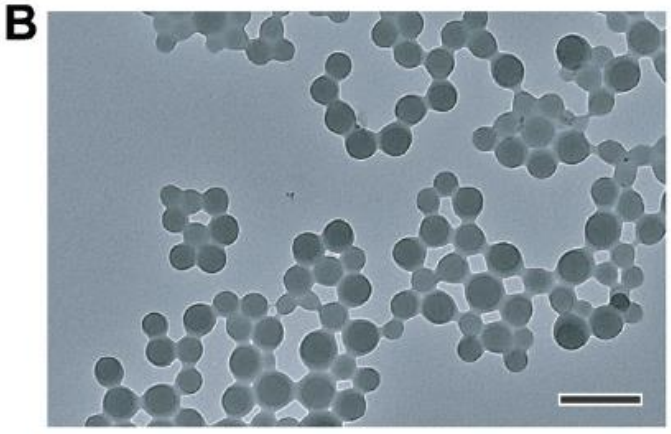


Figure 3-4. Schematic representation of EN1-iPep nanoparticles. (A) Schematic representation of combined therapeutic DTX, EN1-iPep$_{act}$ in PAA-PGMA nanoparticle. (B) Transmission electron microscopy of DTX-loaded nanoparticles. Scale bar provided = 100 nm.

3.2.4 Internalization and in vitro effects of EN1-iPep-docetaxel chimeric nanoparticles

To investigate the effect of EN1-iPep attachment on cellular association of our nanoparticles, T11 cells treated with nanoparticles bearing either EN1-iPepact, EN1-iPepmut, or no EN1- iPep ("blank"), were examined using confocal microscopy. To achieve this, nanoparticles were modified by the addition of fluorescent rhodamine B dye that facilitates visualisation. Figure 3-5A shows EN1-iPep decorated nanoparticles efficiently associate with T11 cells. Similar imaging of T11 cells treated with blank nanoparticles did not show any such association. A 43% induction of apoptosis assessed by cleavage of caspase-3 was identified in T11 cells treated with EN1-iPep$_{act}$-decorated nanoparticles after 8 h but not in cells treated with EN1-iPep$_{mut}$ or blank nanoparticles (Fig 3-5B).

Subsequent investigation demonstrated that EN1-iPep and DTX inhibited cell viability when these agents were assembled in PGMA nanoparticles. In order to test this, T11 cells were treated with increasing concentrations of PAA-PGMA nanoparticles for 24 and 48 h (0 to 0.28 mg/mL^{-1}). Nanoparticles were coated with an equivalent concentration of EN1-iPep (act or mut), (0 to 75 μM). After the treatments, the percentage of cell survival was determined by CellTiter-Glo and the IC50 of formulations determined (Fig. 3-5C). The IC50 of nanoparticles bearing both EN1-iPepact and DTX was 0.23 mg mL^{-1} (13.9 μM EN1-iPep equivalent) after 24 h and 0.10 mg/mL^{-1} (6.03 μM) after 48 h. The combination of DTX and EN1-iPepact was more potent than either component alone when delivered using nanoparticles. Nanoparticles bearing EN1-iPepmut did not reach 50% cell mortality at 24 or 48 h post-treatment regardless of DTX content.

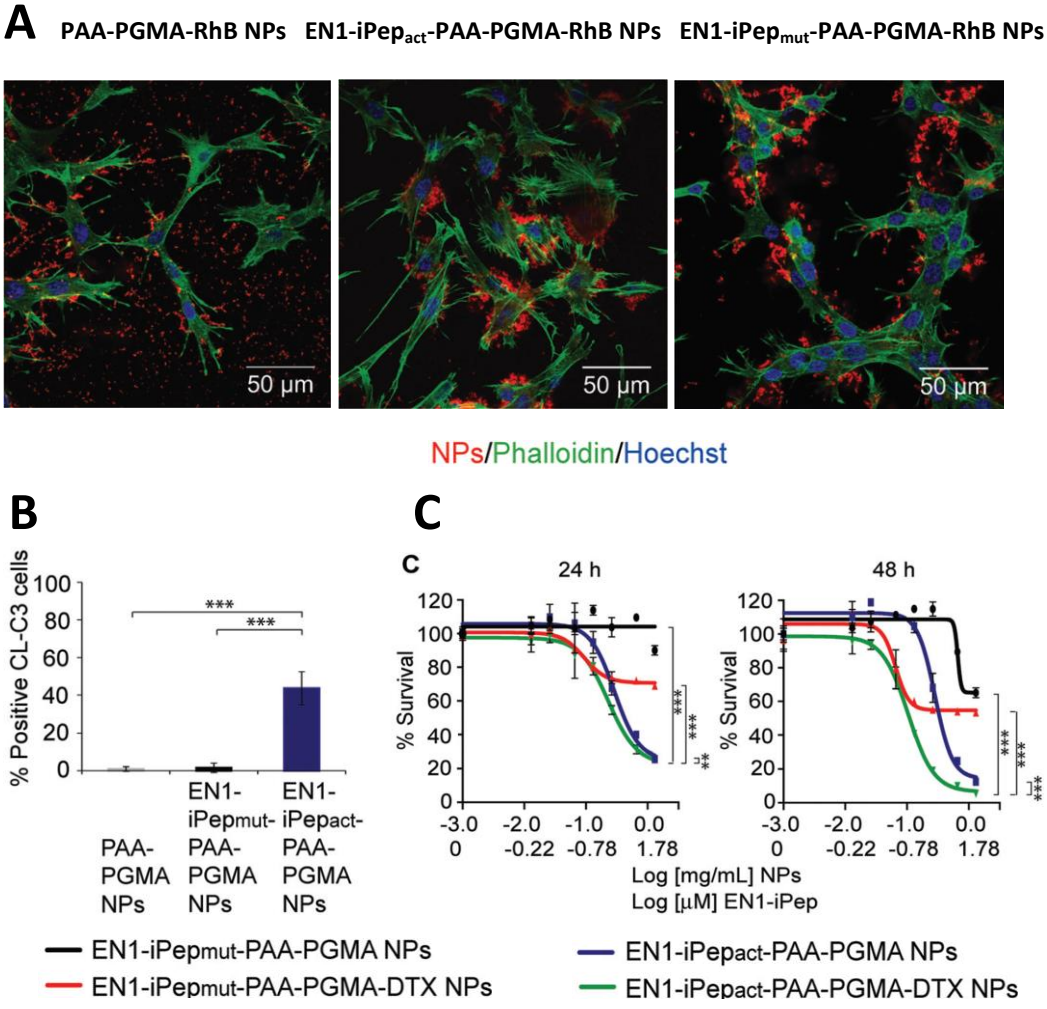


Figure 3-5. Association and *in vitro* effects of EN1-iPep$_{act}$-DTX nanoparticles. (A) Confocal microscopy images of T11 cells treated for 4 hours with blank nanoparticles or with nanoparticles coated with either EN1-iPep$_{act}$ or EN1-iPep$_{mut}$. Rhodamine B tagged nanoparticles appear red. Actin cytoskeleton appears green and nuclei appear blue. (B) Percentage of caspase-3 positive cells after treatments for 24 hours by immunofluorescence. (C) Survival plots of T11 cells after treatments for 24 and 48 hours. Statistical analysis was performed with Student's t-test. *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$.

3.2.5 Anti-tumoural effect of EN1-iPep/DTX nanoparticles relative to clinically approved Abraxane®

Abraxane® is currently the only taxane based nanoformulation approved by the FDA for treatment of breast cancer. This nanoformulation has been proven to improve patients' response and delayed tumour progression in phase III trial[18]. We investigated whether our nanoformulations were more potent than the current industry standard Abraxane® in inducing cell death in basal-like breast cancer cell lines. Viability assays by CellTiter-Glo in T11 cells treated for 24 h and 48 h across varied concentrations of Abraxane® (0–0.28 mg/ml^{-1}), and the same concentrations of PAA-PGMA-DTX NPs and PAA-PGMA-DTX NPs coated with EN1-iPep$_{act}$. The concentrations of EN1-iPep in the last NPs ranged from 0 to 20 µM. We observed a significantly more potent anticancer effect in our DTX-EN1-iPep nanoparticles when compared to Abraxane®. Equivalent concentrations of both nanoparticles showed our nanoparticles to be more efficient inducers of cell death in T11 cells at both the 24 and 48 h time points (Fig 3-6).

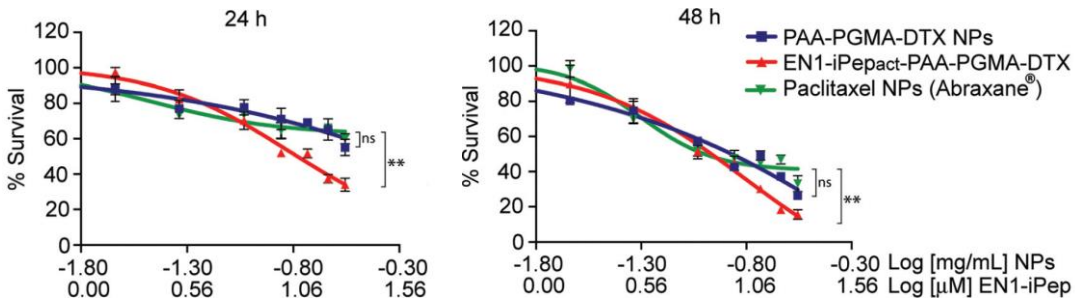


Figure 3-6. Antitumoural effect of EN1-iPep$_{act}$ nanoparticles compared to clinically approved Abraxane®. Survival plots (%) of T11 cells treated with DTX nanoparticles, DTX nanoparticles coated with EN1-iPep$_{act}$ and Abraxane® for 24 and 48 hours at different nanoparticle concentrations. Statistical analysis was performed with Student's *t*-test. *p < 0.05, **p < 0.005, ns means not significant.

3.2.6 In vivo anti-tumour activity of EN1-iPep-PAA-PGMA nanoparticles

After establishing that our EN1-iPep nanoparticles were potent inhibitors of cancer cell growth in an *in vitro* environment we then tested four different NP formulations (Blank NPs coated with EN1-iPep$_{mut}$, Blank NPs coated with EN1-iPep$_{act}$, DTX NPs coated with EN1-iPep$_{mut}$ and DTX NPs coated with EN1-iPep$_{act}$) for their *in vivo* ability to inhibit tumour growth. The model used was of mice implanted with subcutaneous T11 tumour allografts using a model derived from serial orthotopic transplantation of mammary tumours from a p53 null mouse into a p53 wild-type syngeneic recipient[15-18]. This model uses animals with an intact immune system while mimicking human claudin-low tumours of which the majority are triple negative phenotypically. Tumour allografts were injected in the flank of BALB/c mice and were treated via intratumoural injection of nanoparticle solution in PBS (2.5 mg nanoparticles per injection) once tumours reached ~300 mm^3 (determined by digital callipers). Injections were performed every 2 days, with a total number of 5 injections occurring (Fig 3-7A). The animal group treated with DTX nanoparticles bearing EN1-iPepact showed a reduction in tumour growth at the end of the treatment phase and exhibited the highest survival relative to the other formulations (Fig 3-7B and C). In comparison, animals treated with formulations containing no DTX and bearing EN1-iPep$_{mut}$ were the first group to reach the endpoint. A particularly noteworthy observation for tumours treated with EN1-iPep$_{act}$ nanoparticles was their marked inhibition between days 7 and 9, which reached end point of the experiment at day 14 ($p < 0.03$).

Histopathological analysis of tumours determined by terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick-end labelling (TUNEL) assay demonstrated a significant increase (>50%) in apoptotic cells in the combinatorial treatment (iPep$_{act}$ and DTX) relative to single treatments (25%) (p = 0.002 and 0.001 respectively) (Fig 3-8).

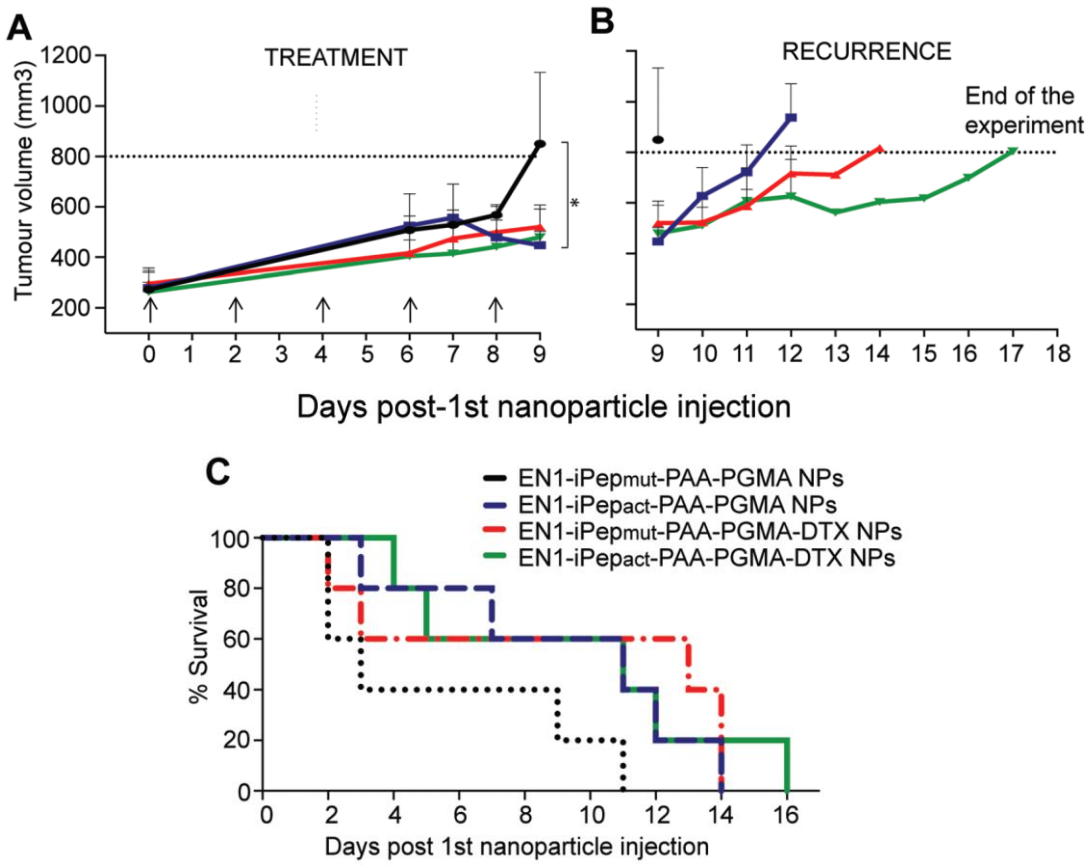


Figure 3-7. *In vivo* antitumour activity of EN1-iPepact nanoparticles. Antitumour activities of various nanoparticle formulations in an allograft model of basal-like breast cancer in BALB/c mice. Mice bearing T11 tumours (~300 mm3) were treated with blank nanoparticles coated with EN1-iPepmut, blank nanoparticles coated with EN1-iPepact, DTX nanoparticles coated with EN1-iPepmut and DTX nanoparticles coated with EN1-iPepact (n = 5). (A) Treatment phase with the timing of the injections indicated with black arrows. (B) Recurrence of tumour growth after cessation of treatment. (C) Percent survival of mice treated with the four different nanoformulations.

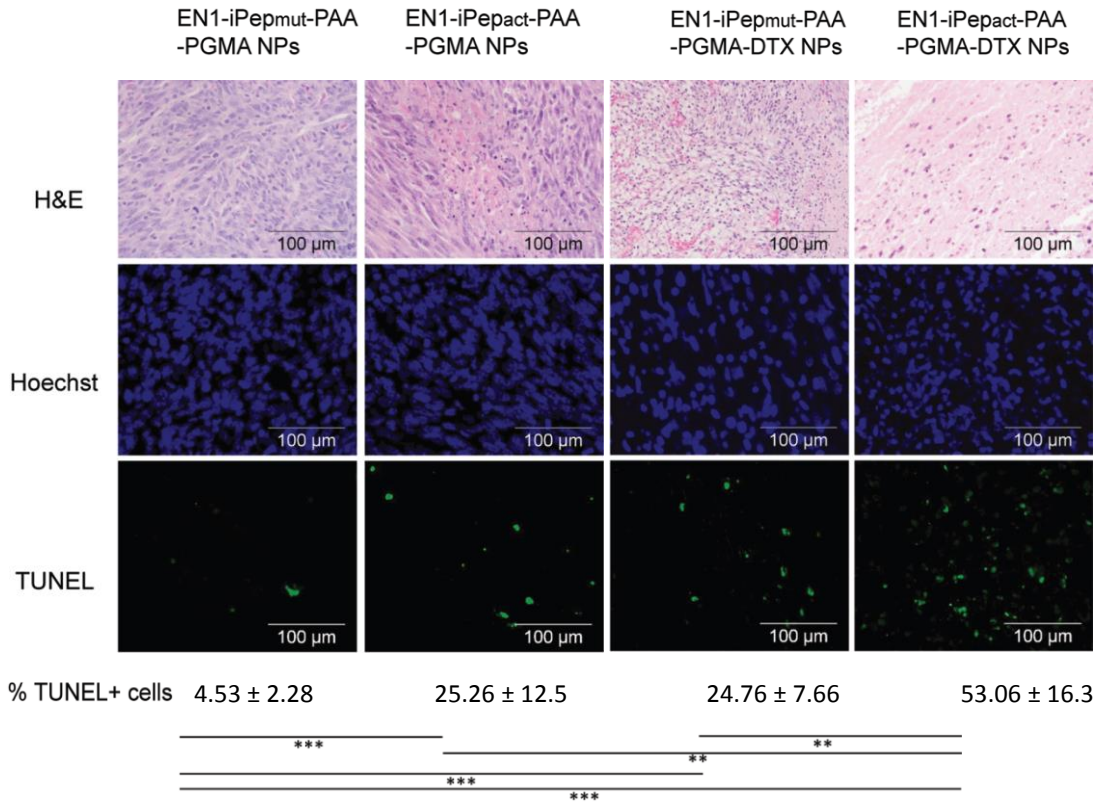


Figure 3-8. Representative H&E, Hoechst and TUNEL assay photos. Tumours treated with the four nanoformulations and percentages of TUNEL positive cells obtained for each condition. Statistical analysis was performed with Student's t-test. *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$

Haematoxylin/eosin staining of the tumour sections demonstrated increased tissue damage, and areas of intense necrosis, in tumours treated with nanoparticles delivering both EN1- iPep$_{act}$ and DTX over the other treatment groups, which correlated with the enhanced apoptosis in the double treatments. Our findings indicate a benefit with the co-administration of DTX and EN1-iPep$_{act}$ in the nanoparticle formulation, which delayed the progression and the recurrence of the tumours post-administration.

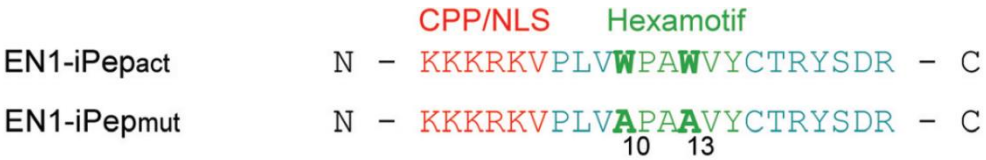


Figure 3-9. Peptide sequence of the EN1-iPep$_{act}$ and EN1-iPep$_{mut}$. Peptides coloured red belong to the sequence responsible for coding the cell penetrating and nuclear localization region. Peptides coloured green are the hexamotif shared by the homeodomain superfamily of transcription factors.

3.3 Discussion

The overarching aim of the experiments presented in this chapter was to devise a novel drug formulation that could efficiently sensitise triple negative and basal-like breast cancers to chemotherapy. To this end, we have successfully demonstrated that our EN1 transcription factor-inhibiting peptide acts in a potent and cancer cell specific manner that does not affect normal breast epithelium. The EN1-iPeps are 22-residue peptides derived from the EN1 transcription factor with the following sequence: N-KKKRKVPLVWPAWVYCTRYSDR-C. This sequence corresponds to the active form of the peptide, which carries anticancer properties in its EN1-iPep$_{act}$ form. In the mutant (EN1-iPep$_{mut}$) form, the two tryptophan residues at positions 10 and 13 were replaced by alanines (Fig. 3-8). The tryptophans are contained in a highly conserved hexamotif, WPAWVY, which is shared by the homeodomain superfamily of transcription factors. These two tryptophan residues are essential for the interaction between EN1 and their binding co-factors (PBX1 and HOX members) and DNA. Thus, the mutant form EN1-iPep$_{mut}$ did not affect either the cancer model cells (T11 or SUM149) or the normal breast cells (MCF10A) owing to the altered hexamotif. This is promising evidence that indeed transcription factors can be targeted for use in anticancer treatments through this highly conserved homeodomain hexamotif common to many basal-like breast cancers. The N-terminus of the peptide carries a cell penetration/nuclear localization sequence (CPP/ NLS), KKKRV which is present in the Simian Virus 40 (SV40) large T antigen and is necessary for the internalization of the peptide through plasma and nuclear membranes.

We can confidently attribute the hallmark features of induced apoptosis seen in cells treated with EN1-iPep$_{act}$ due to EN1 transcription factor interference and not simply due to foreign substances introduction into the cells. Previous reports using EN1 and EN2 knockout mice confirm this assertion[19-21].

3.3.1 Docetaxel and EN1-iPep synergy

The nanoparticles used in this chapter are a modified formulation of a previously reported nanoparticle structure created in our lab in 2011 by Evans et, al[22]. In 2011 we reported that similar PGMA nanoparticles were capable of endocytosis into neuronal cells via a charge-mediated affinity with cell membranes. Here the negatively-charged cell membrane endocytosed PMGA nanoparticles were decorated with the positively charged polymer polyethylenimine (PEI). In the experiments presented in this chapter we demonstrated that nanoparticles made of the same PMGA core, modified with PAA (a negatively charged polymer) and then further modified with electrostatic addition of the EN1-iPep$_{act}$ (a positively charged peptide) (monitored by a positive shift in the zeta potential in Fig 3-3) were successfully up taken by T11 and SUM149 cell lines resulting in potent apoptosis. The same result was observed in DTX loaded PAA-PGMA-NPs decorated with the EN1-iPep. We attribute the successful cellular uptake to the EN1-iPep sequence being rich in positively charged residues that facilitate electrostatic interactions with anionic PAA and cell membranes. It should be mentioned that cellular uptake is most likely aided by the SV40 cellular internalisation motif (which is particularly rich in positive residues) present in the EN1-iPep. We also observed a lack of internalisation in our nanoparticles decorated with the mutant form of the EN1-

iPep, and our blank nanoparticles, which were not decorated with any variant of iPep. In the latter case, this is most likely due to a repelling action between cell membranes and the negative surface charge of the PAA on the nanoparticles. Whereas for the EN1-iPep$_{mut}$ nanoparticles, we suggest that failure to internalise is attributed to the substitution of negatively charged alanine residues reducing electrostatic attraction. Because the alanine substitution occurs in the homeodomain hexamotif, and not the cell penetrating region of the iPep sequence, this suggests that the internalisation effect is dominated more by the surface charge of our nanoparticles than the CPP region of the outer peptides decorating.

When we compared the efficiency of the combined delivery of free EN1-iPep$_{act}$ and DTX to a single formulated nanoparticle delivery of DTX loaded EN1-iPep$_{act}$ nanoparticles we found a significantly more potent apoptotic response. Nanoparticles bearing EN1-iPepmut did not reach 50% cell mortality at 24 or 48 h post-treatment regardless of DTX content. These results indicate that our formulations were successfully delivered into basal-like breast cancer cells, and that the EN1-iPep on the nanoparticle surface enhanced the effects of DTX in inhibiting breast cancer cell growth. Similarly, other agents assembled in nanoparticles have been used in order to enhance the therapeutic potential of chemotherapeutic drugs such as antibodies, siRNAs or plasmid DNAs[23-27]. EN1 is only expressed in dopaminergic neurons and in certain malignancies such as neuroblastomas[28] and basal-like breast carcinomas[16,29,30] which are highly enriched in stem cell-like characteristics[16,30,31] but not in normal breast tissue. EN1 expression in basal-like breast cancers could explain their intrinsic multidrug resistance. Similar to our work, some researchers attempted to abrogate the stem cell phenotype using

plasmidic cDNA or special nanocarriers which lead to the sensitization of glioblastoma and hepatic tumours to chemotherapeutic drugs such as temozolomide[27] and epirubicin[32].

3.3.2 Anti-tumour effect of EN1-iPep/DTX NP vs Abraxane

Abraxane® consists of albumin-bound paclitaxel nanoparticles (130 nm in size) and was approved in 2005 for the treatment of metastatic breast cancer. Other paclitaxel nanoformulations are currently in phase II/III clinical trials, and one DTX PEGylated nanoparticle formulation, NKTR-105® is still in its phase I clinical trial. We observed that our DTX nanoparticles bearing EN1-iPepact were significantly more potent in comparison with Abraxane® at both time points when used at an equivalent nanoparticle concentration (Fig. 3-6).

It is important to note that the paclitaxel loading in Abraxane® is superior to DTX loading in our DTX nanoparticles, 10% *versus* 3%. In the clinic, free DTX was demonstrated to perform slightly better than free paclitaxel in metastatic breast cancer patients when used at 100 mg m^{-2} and 175 mg m^{-2} respectively. Taking this into account, we can conclude that our synthesised DTX nanoparticles are more potent than Abraxane®[33]. The finding that our DTX nanoparticles performed better than Abraxane® is of potential interest because of the improved potency in this study and the high specificity of EN1-iPep for basal-like breast cancers over non-basal breast cancer cell lines[5] and normal cells (Fig. 3-1C). This data further validates the concept that EN1-iPep could be used to decrease the dose and potentially the off-target toxicity of current chemotherapeutic regimes.

REFERENCES

1. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? Nature reviews Clinical oncology. 2010;7(12):683-692.
2. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-752.
3. Mayer IA, Abramson VG, Lehmann BD, Pietenpol JA. New strategies for triple-negative breast cancer—deciphering the heterogeneity. Clinical cancer research. 2014;20(4):782-790.
4. Carey LA. Targeted chemotherapy? Platinum in BRCA1-dysfunctional breast cancer. Journal of Clinical Oncology. 2010;28(3):361-363.
5. Beltran A, Graves LM, Blancafort P. Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. Oncogene. 2014;33(39):4767-4777.
6. Martin M, Pienkowski T, Mackey J, Pawlicki M, Guastalla J-P, Weaver C, et al. Adjuvant docetaxel for node-positive breast cancer. New England Journal of Medicine. 2005;352(22):2302-2313.
7. Schiff P, Horwitz SB. Taxol stabilizes microtubules in mouse fibroblast cells. Proceedings of the National Academy of Sciences. 1980;77(3):1561-1565.
8. Gradishar WJ. Taxanes for the treatment of metastatic breast cancer. Breast cancer: basic and clinical research. 2012;6:159.
9. Park J, Fong PM, Lu J, Russell KS, Booth CJ, Saltzman WM, et al. PEGylated PLGA nanoparticles for the improved delivery of doxorubicin. Nanomedicine: Nanotechnology, Biology and Medicine. 2009;5(4):410-418.
10. Palma G, Conte C, Barbieri A, Bimonte S, Luciano A, Rea D, et al. Antitumor activity of PEGylated biodegradable nanoparticles for sustained release of docetaxel in triple-negative breast cancer. International journal of pharmaceutics. 2014;473(1):55-63.
11. Liang M, Fan K, Zhou M, Duan D, Zheng J, Yang D, et al. H-ferritin–nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection. Proceedings of the National Academy of Sciences. 2014;111(41):14900-14905.
12. Kriege M, Jager A, Hooning MJ, Huijskens E, Blom J, van Deurzen CHM, et al. The efficacy of taxane chemotherapy for metastatic breast cancer in BRCA1 and BRCA2 mutation carriers. Cancer. 2012;118(4):899-907.
13. Evans CW, Viola HM, Ho D, Hool LC, Dunlop SA, Fitzgerald M, et al. Nanoparticle-mediated internalisation and release of a calcium channel blocker. RSC Advances. 2012;2(23):8587-8590.
14. Singh R, Norret M, House MJ, Galabura Y, Bradshaw M, Ho D, et al. Dose-Dependent Therapeutic Distinction between Active and Passive Targeting Revealed Using Transferrin-Coated PGMA Nanoparticles. Small. 2016;12(3):351-359.
15. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast cancer research. 2010;12(5):R68.
16. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC, et al. Combined PI3K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models. Clinical cancer research. 2012;18(19):5290-5303.
17. Chou T-C, Talalay P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Advances in enzyme regulation. 1984;22:27-55.
18. Gradishar WJ, Tjulandin S, Davidson N, Shaw H, Desai N, Bhar P, et al. Phase III trial of nanoparticle albumin-bound paclitaxel compared with polyethylated castor oil–based

paclitaxel in women with breast cancer. Journal of clinical oncology. 2005;23(31):7794-7803.
19. Albéri L, Sgadò P, Simon HH. Engrailed genes are cell-autonomously required to prevent apoptosis in mesencephalic dopaminergic neurons. Development. 2004;131(13):3229-3236.
20. Alvarez-Fischer D, Fuchs J, Castagner F, Stettler O, Massiani-Beaudoin O, Moya KL, et al. Engrailed protects mouse midbrain dopaminergic neurons against mitochondrial complex I insults. Nature neuroscience. 2011;14(10):1260-1266.
21. Sgadò P, Albéri L, Gherbassi D, Galasso SL, Ramakers GMJ, Alavian KN, et al. Slow progressive degeneration of nigral dopaminergic neurons in postnatal Engrailed mutant mice. Proceedings of the National Academy of Sciences. 2006;103(41):15242-15247.
22. Evans CW, Fitzgerald M, Clemons TD, House MJ, Padman BS, Shaw JA, et al. Multimodal analysis of PEI-mediated endocytosis of nanoparticles in neural cells. Acs Nano. 2011;5(11):8640-8648.
23. Li J, Cai P, Shalviri A, Henderson JT, He C, Foltz WD, et al. A multifunctional polymeric nanotheranostic system delivers doxorubicin and imaging agents across the blood–brain barrier targeting brain metastases of breast cancer. ACS nano. 2014;8(10):9925-9940.
24. Su S, Tian Y, Li Y, Ding Y, Ji T, Wu M, et al. “Triple-punch” strategy for triple negative breast cancer therapy with minimized drug dosage and improved antitumor efficacy. ACS nano. 2015;9(2):1367-1378.
25. Inoue S, Patil R, Portilla-Arias J, Ding H, Konda B, Espinoza A, et al. Nanobiopolymer for direct targeting and inhibition of EGFR expression in triple negative breast cancer. PLoS One. 2012;7(2):e31070.
26. Xu W, Luo T, Li P, Zhou C, Cui D, Pang B, et al. RGD-conjugated gold nanorods induce radiosensitization in melanoma cancer cells by downregulating α. International journal of nanomedicine. 2012;7:915-924.
27. Kim S-S, Rait A, Kim E, Pirollo KF, Nishida M, Farkas N, et al. A nanoparticle carrying the p53 gene targets tumors including cancer stem cells, sensitizes glioblastoma to chemotherapy and improves survival. ACS nano. 2014;8(6):5494-5514.
28. Ceballos-Picot I, Mockel L, Potier M-C, Dauphinot L, Shirley TL, Torero-Ibad R, et al. Hypoxanthine-guanine phosphoribosyl transferase regulates early developmental programming of dopamine neurons: implications for Lesch-Nyhan disease pathogenesis. Human molecular genetics. 2009;18(13):2317-2327.
29. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How many etiological subtypes of breast cancer: two, three, four, or more? Journal of the National Cancer Institute. 2014;106(8):dju165.
30. Su Y, Subedee A, Bloushtain-Qimron N, Savova V, Krzystanek M, Li L, et al. Somatic cell fusions reveal extensive heterogeneity in basal-like breast cancer. Cell reports. 2015;11(10):1549-1563.
31. Bunting SF, Callén E, Kozak ML, Kim JM, Wong N, López-Contreras AJ, et al. BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. Molecular cell. 2012;46(2):125-135.
32. Wang X, Low XC, Hou W, Abdullah LN, Toh TB, Mohd Abdul Rashid M, et al. Epirubicin-adsorbed nanodiamonds kill chemoresistant hepatic cancer stem cells. ACS nano. 2014;8(12):12151-12166.
33. Jones SE, Erban J, Overmoyer B, Budd GT, Hutchins L, Lower E, et al. Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer. Journal of Clinical Oncology. 2005;23(24):5542-5551.

CHAPTER 4

Recombinant expression and crystallisation of G-quadruplex binding BG4 antibody

Abstract

Four-stranded G-quadruplexes (G4) DNA secondary structures have attracted particular interest since they have been shown to concentrate in gene regulatory regions within cells. Recently a structure-specific antibody, BG4, has been engineered to visually quantify DNA G4s in human cells. The BG4 antibody binds G4 structures without any binding to regular Watson-Crick secondary structures. G4s association with gene regulatory and oncogenic regions make them a logical target for cancer therapy and the selective binding displayed by BG4 appears to be an ideal mode of potential therapy binding. We have attempted to solve the three dimensional structure of BG4 using molecular replacement modelling from crystallized recombinant protein samples. Though unsuccessful in this goal we do identify potential contamination sources for false positive crystals in the form of bacterial *Hfq* proteins. We go on to describe where in our methodology such contaminants could have gone undetected in the protocols used and offer solutions to correcting these problems. We also present useful data for improving the amount of protein yielded from each bacterial expression of BG4. Together, these reports provide the first steps towards a successful methodology for the recombinant expression, purification, and subsequent crystallization needed for molecular replacement modelling to solve the three dimensional structure of the BG4 antibody.

Introduction

Building on observational reports from the 1960s and even as far back as 1910, the now seminal work of Sen and Gilbert in 1988 showed that regions of DNA sequences containing short guanine (G) rich motifs can self-assemble in distinct four-stranded secondary structures we now know as G-quadruplexes[1]. G-quadruplexes were proposed to be involved in telomere association, recombination and replication after it was noted that their structural stability was cation-dependent and their thermodynamic viability reliant on physiological conditions. Despite this early association it wasn't recognised until later that G-quadruplexes occur in abundancy in regulatory regions of the human genome[2-10]. Concentrating in replication promoter regions, oncogenes and telomeric DNA, G-quadruplexes began being considered as a possible target for anticancer chemotherapies. The atypical shape created by G-quadruplexes obstructs DNA polymerase leading to an increased risk of non-allelic homologous recombinant replication, DNA breakage as well as being implicated in germline deletions. The intrinsic mutagenicity of G-quadruplex DNA has rationally led to a desire to design small molecules that could stabilize the quadruplex conformation by trapping them in cellular DNA, down-regulating oncogene transcription, promoting telomeric disruption and inducing cell growth arrest in cancer cells[11-15]. There are some 370,000 identified G-quadruplex motifs in the human genome with most research focused on describing their location or structural orientations. In the process of mapping G-quadruplex locations in the genome researchers have used a fluorescently-tagged antibody (*BG4*) that binds to G-quadruplexes *in vitro* in a highly specific manner[16]. This result is especially exciting because it demonstrates that G-quadruplexes are a viable

structurally specific target when designing new anticancer therapies. What remains to be understood is the precise nature of the binding between the BG4 antibody and the G-quadruplex structure. In this chapter we have endeavoured to define the currently unknown structure of the BG4 antibody using molecular replacement modelling of crystals grown from recombinantly expressed BG4. In doing so we have identified means to improve expression yields and identified pitfalls that can occur in the purification and crystallization processes involved.

4.1 Materials and methods

Buffers used during recombinant expression and purification of BG4 antibody

Table 4-1. Purification conditions of recombinantly expressed BG4 antibody. *elusion buffer had no addition of protease inhibitor tablets.

Protein	*Resin*	*Lysis Buffer*	*Elusion Buffer**	*SEC Buffer*
BG4 antibody	Ni-NTA (Bio-Rad)	50 mM tris, 1M KCl, 20 mM imidazole, pH 8	50 mM tris, 250 mM KCl, 500 mM imidazole, pH 8	50 mM tris, 250 mM KCl, pH 8

Table 4-2. Crystal screening conditions (Hampton Research). 96-well sitting drop vapour-diffusion format via an Art Robbins Phoenix robot using concentrated protein samples plus 150 nl reservoir solution equilibrated against 80 µL reservoir solution in 96-well format ARI LVR Intelli-Plate (Hampton Research) at 293 ± 0.5 K.

Protein Sample No.	*Concentration*	*Crystal Screen Condition(s)*	*Positive Crystal Condition*
1	[4.5mg/ml]	Crystal Index	Nil Nil
2	[6.19mg/ml]	Crystal Index	91. 0.2M NHHPO, 0.1M Tris pH 8.5, 5.% v/v $(CH_3)_2CCH_2CHCH_3$ Nil
3	[9.65mg/ml]	PEG	Nil
4	[11.74mg/ml]	Crystal Crystal Index	17. 0.2M $Li_2SO_4 \cdot H_2O$, 0.1M Tris-HCl pH 6.5, 20% w/v PEG 4,000 18. 0.2M $Mg(CH_3COO)4H_20$, 0.1M $(CH_3)_2AsO_2Na$ $\cdot 4H_20$ pH 6.5, 20% w/v PEG 8,000 22. 0.2M $(CH_3)_2AsO_2Na$ $\cdot 3H_20$, 0.1M Tris-HCl pH 8.5, 30% w/v PEG 8,000 Nil
5	[23.4mg/ml]	Crystal Index PEG	Nil Nil Nil

4.1.1 Recombinant expression and purification of BG4 antibody

4.1.1.1 Transformation of BL21 (DE3) cells

The BG4 antibody gene was sub-cloned into the pSANG10 plasmid and expressed in BL21 (DE3) *Escherichia coli (E. coli)* cells. The BG4 plasmid (3.6 µl, ~30 ng/µl) was added to of BL21 (DE3) cells (100 µl). Following a 20-minute incubation on ice, the cells were heat shocked at 42^{o}C for 45s. LB media (800 µl) was added and the cells were incubated at 37^{o}C for 45 minutes. The transformed cells were then plated on kanamycin-treated LB-agar plates and stored overnight at 37^{o}C.

4.1.1.2 Bacterial cell culture of BG4 antibody

A single colony was used to inoculate a pre-culture solution of kanamycin-treated LB media (100 ml) and the culture was stored at 37^{o}C overnight in an orbital shaker. The pre-culture solution (1% v/v) was used to inoculate 2YT antibiotic-treated media (1% w/v D-glucose, 50 mM Tris at pH 8.0). The culture was incubated at 37^{o}C in an orbital shaker until the ΔOD600nm = 0.6. It was then induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM) and incubated overnight at 25^{o}C in an orbital shaker. Cells were harvested the following day, snap frozen and stored at -80^{o}C.

4.1.1.3 Cell lysis and BG4 purification via NiTA and size exclusion column

BG4 BL21 (DE3) culture pellet (harvested from collective yield of 6 x 500 ml main cultures) was defrosted on ice for ~ 30 minutes. *BG4* lysis buffer (40 ml, 50 mM tris, 1M KCl, 20 mM imidazole, pH 8) was added to pellets and re-suspended with pipetting. It should be noted here that early binding buffers used during Ni^{2+} affinity columns contained 500 mM KCl, this was later increased to 1M KCl to improve binding. The homogenous sample was passed through a cell cruncher (Emulsiflex C5 high-pressure homogeniser (Avestin)) with the lysate clarified by centrifugation at 20 K x *G* for 40 minutes at 4°C and 0.22 µm filtration. The supernatant was loaded onto a 1 mL Ni-NTA column (GE Healthcare) at a rate of 1 ml/min at 4°C. *BG4* was eluted using a ten-column-volume imidazole gradient (25-500 mM). Relevant elution fractions were collected and concentrated using a 10kDa centrifugal concentration device (Amicon) centrifuged 4.5K x *G*. The concentrated fractions were loaded in 5 ml injections onto HiLoad 16/60 Superdex 200 column (GE Healthcare) developed with gel-filtration size exclusion (SEC) buffer at 1 ml min^{-1}. *BG4* was again concentrated following gel filtration using 10 kDa molecular cut off concentrators (4.5K x *G*) to a range of concentrations ([4.5mg/ml], [6.19mg/ml], [9.65mg/ml], [11.74mg/ml] and [23.4mg/ml]) determined from absorption at 280 nm using an estimated absorption coefficient of 55600 M^{-1} cm^{-1}. Purified *BG4* antibody was snap frozen with liquid nitrogen for storage AT -80 °C or ideally used immediately in 96/24-well crystal trays.

4.1.2 Cell culture and confocal microscopy of recombinantly expressed *BG4*

HEK-293T and MCF-7 cells were seeded at 1.5 and 2.5 × 10^5 cells per well respectively in 6-well plates on top of poly-L-lysine coated glass coverslips (ProSciTech) overnight. The cells were further incubated at 5% CO_2, 37°C for 48 h. The samples were incubated with cytoplasm removal buffer (20 mM HEPES-KOH, 20 mM NaCl, 5 mM $MgCl_2$, 300 mM sucrose, 0.5% (v/v) NP-40, pH 7.9) for 20 min before being fixed in 2% paraformaldehyde/PBS followed by permeabilization with 0.1% Triton-X100/PBS. Samples were then treated with 10 µM of GQC-05 (a highly effective G4 stabilizing ligand) for 1 h[17]. Thereafter, all samples were blocked with 2% skimmed milk/PBS and subjected to standard immunofluorescence methods with BG4, anti-FLAG (#2368, Cell Signaling Technology). Sections were imaged by confocal microscopy (Nikon A1Si, Centre for Microscopy, Characterisation and Analysis, The University of Western Australia). DAPI was imaged at λ_{ex} 405 nm and λ_{em} 450/25 nm. Images were captured at 120x magnification and the entirety of the cell nuclei were imaged at 0.5 mm Z-plane intervals over 15 to 25 Z-sections. G4 foci counts for samples were analyzed using the open-source image processing package, FIJI. The Z-stack images were converted to a single-plane maximum intensity projection (MIP) in order to accurately count the G4 foci. Thresholding was applied using appropriate levels to the MIP of the red channel (with G4 foci) to further reduce noise before doing an automated particle count using FIJI. G4 foci were counted for each individual nucleus in the images.

4.1.3 Crystallisation and X-ray data collection

Crystallisation space was explored using Index, PEG and Crystal screen (Hampton Research) set up in 96-well sitting drop vapour-diffusion format via an Art Robbins Phoenix robot as follows: 1 or 2 nL of concentrated protein solutions (See table 4-2) plus 150 nL reservoir solution was equilibrated against 80 µL reservoir solution in 96-well format ARI LVR Intelli-Plate (Hampton Research) at 293 ± 0.5 K. Clusters of rod like crystals formed in condition No. 17, 18 and 22 of the crystal screen using a stock *BG4* solution concentrated to 11.74 mg/ml. A single large plate formed in condition No. 91 of crystal screen using 6.19mg/ml *BG4*. 24-well hanging-drop optimization was repeated in the 4 condition that produced crystals but failed to yield crystals a second time. Crystals were first observed 4 weeks after screens were prepared. All crystal conditions were negative when checked at the 3 week mark and did not grow any larger after being first observed. Crystal clusters were harvested and transferred to cryoprotection solution using a 0.02 mm nylon loop [30% (*w/v*) glycerol in MilliQ water mixed in 1:1 ratio with respective well crystallizing solution and pH] before flash-cooling with liquid nitrogen. Diffraction experiments were carried out on the MX2 beamline at the Australian Synchrotron, Melbourne, Victoria, Australia. Data were collected using X-rays of energy of 13 keV (λ = 0.953 Å) and a crystal-to-detector distance of 250 mm (2.5 Å at the edge of an ADSC Quantum 315r detector).

4.1.4 Data processing and crystal structure solution

Data-collection and processing statistics are presented in Table 4-3. Data were processed and scaled using *XDS*[18] and *AIMLESS*[19] respectively. A likely space group of P6 was identified by AIMLESS. The *L* statistic analysis[20] of cumulative intensity distribution indicated apparent crystal twinning. Solvent-content analysis[21] indicated the most likely asymmetric unit content to be approximately 2.5 at the edge of plate. The self-rotation function was calculated with *MOLREP*[22] using data between 30.64 and 2.5 Å resolution with an integration radius of 14.12 Å Molecular replacement calculations were carried out with MOLREP and crystallographic refinement with REFMAC5[23].

4.1.5 Proteolysis stability assessment of *BG4* antibody over time

In an effort to assess whether protein crystals were whole *BG4* or smaller fragments that had crystallised following partial digestion over time SDS-PAGE gels were run with BG4 samples containing trypsin in 1:10 and 1:100 ratio solutions over 1 hr, 3 hrs and overnight time points. 9 µl of *BG4* [12mg/ml] was added to 1 µl of trypsin (1M) ~16 hrs, 3hrs, 1hr and immediately prior to being loaded onto gels. The same process was repeated and then serial diluted once to yield the 1:100 proteolysis condition.

4.2 Results

4.2.1 Expression and purification of *BG4* antibody

Successful HisTrap column binding of *BG4* was confirmed with SDS-PAGE (Figure 4-1(A)). Gels were run containing samples of the 8 elution fraction that were collected along with samples of the cell lysate and NiTA column flow through. As expected, the cell lysate is loaded with all the protein content of the bacterial cell. When we compare this to the flow through of the HisTrap column we see missing bands around the 80, 40 and importantly 30 kDa regions. These missing bands are seen in the 8 fractions collected that were loaded onto the Superdex 200 column. Following size exclusion column purification, gels reveal successful removal of the 80 kDa protein and most of the 40 kDa protein (Figure 4-1 (B)). Very faint amounts of the 40 kDa protein proved very difficult to remove from solution. 0.22 μm syringe filtration and via centrifugation did remove most of the remaining 40 kDa protein but faint streaking is still visible in some samples. Importantly gels show pronounced retention of *BG4* antibody at 30 kDa marker.

To confirm that 30 kDa protein marker was the desired BG4 antibody, functional immunoflurescence assays were performed in HEK-293T (Figure 4-1 (C)) and MCF-7 cell lines (Figure 4-1 (D)) and imaged under confocal microscopy. Quantitative visualization of G4 foci are clearly present and consistent with previous studies[16].

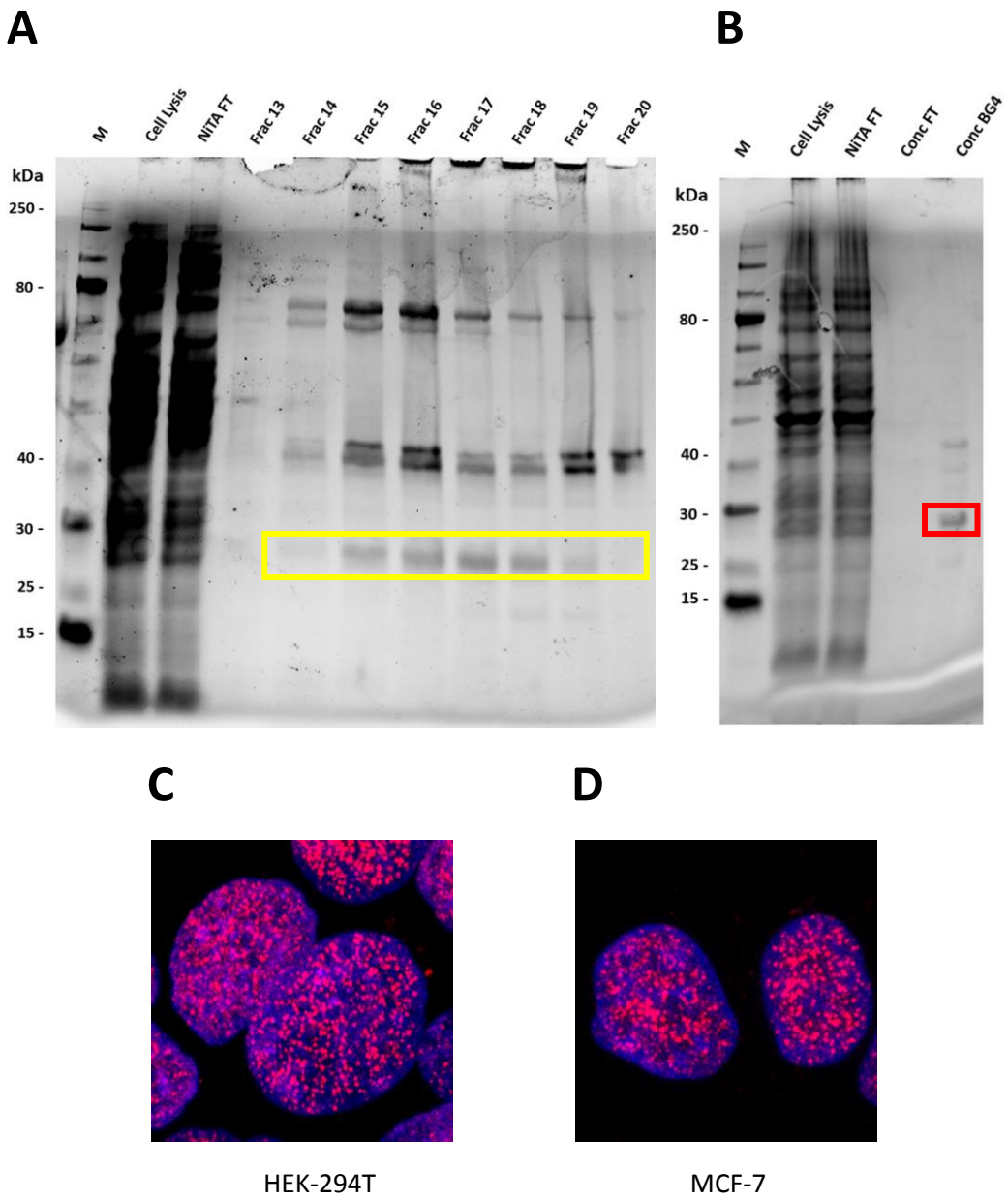


Figure 4-1. Recombinant *BG4* expression. Gel electrophoresis of BG4 elution fractions following HisTrap column purification (A) and subsequent S200 size exclusion column purification (B). BG4 elusion fractions were on stain-free pre-cast polyacrylamide gel and imaged using ChemiDoc MP system. Yellow square depicts presence of BG4 in HisTrap elution fractions, whereas the red square shows purified and isolated BG4 after SEC. Visualisation of G-quadruplexes by BG4 binding in HEK-293T (C) and MCF-7 (D). Pink shows G4 loci were BG4 fluro tag has bound. Purple is discrete cell nuclei after cell cytoplasm removal.

4.2.2 Crystallization and X-ray data processing

From the three crystal conditions that successfully grew crystals, 4 individual crystals were harvested and analysed at the Australian Synchrotron. Two crystals from conditions number 17 and 18 were selected for X-ray structural analysis. The large and atypically shaped plate crystal from well 91 using [6.19mg/ml] BG4 was not sent for synchrotron analysis as it did not absorb UV light at 280 nm and was therefore concluded not be protein. Of the four samples sent to the Australian Synchrotron only one crystal belonging to well number 18 (BG4 [11.74mg/ml)] produced a diffraction pattern (Figure 4-2 (A)).

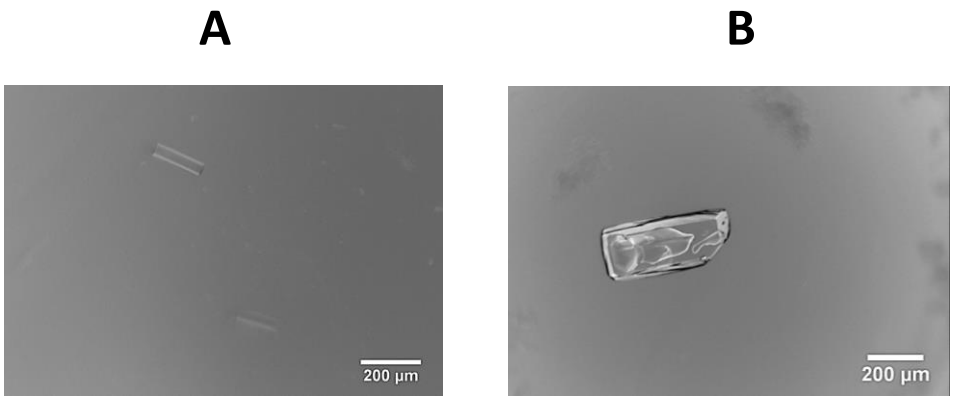


Figure 4-2. Crystals of E.coli Hfq 4RCB hexamers. (A) rod-like crystals grown in 96-well sitting drop vapour-diffusion crystal screen format (Hampton Research).The crystal shown here is the same crystal the produced X-ray diffraction pattern at Australian Synchrotron. (B) Large plate crystal grown in similar crystal screen conditions but failed to absorb UV light at 280 nm and was excluded from further analysis.

Using the data-collection strategy, protein crystals diffracted to beyond 2.99 Å, producing a tightly spaced reflection pattern on the detector (Figure 4-3). Data from the Australian Synchrotron was processed using AIMLESS function to reveal a P6 space group with unit cell volume of 91826.72. A Matthews' coefficient function using the approximate molecular weight of 30 kDa returned no compatible results. Troubleshooting, a second function was run this time using 1 kDa molecular weight with the results summarised in Table 4-4. Importantly these results suggest the content of the unit cell to be consistent with a protein of mass 7 kDa, far smaller than the known 30.7 kDa of *BG4*.

Table 4-3. X-ray data-collection and processed statistics for crystal screen condition No.18

Protein crystal data-collection and processing statistics	
Space group	*P6*
Unit-cell parameters (Å, º)	*a* = 61.28, *b* = 61.28, *c* = 28.23, *α* = 90.00, *β* = 90.00, *γ* = 120.00
Molecules per asymmetric unit	7
Resolution (Å)	30.64-1.99
Measured reflections	47078
Unique reflections	4286
Completeness (%)	99.9
Multiplicity	11.0
‹*I*/σ (*I*)›	23.8
$CC_{1/2}$	0.99
R_{merge} (%)	0.08
Average mosaicity (º)	0.09
Wilson *B* factor ($Å^2$)	11.3
L statistic	No twinning

Table 4-4. CCP4 Matthews function report using estimated molecular weight 1000 Da.

Estimated molecular weight 1000.

Nmol/asym	Matthews Coeff	%solvent	P(1.99)	P(tot)
1	15.30	91.97	0.00	0.00
2	7.65	83.94	0.00	0.00
3	5.10	75.90	0.00	0.00
4	3.83	67.87	0.01	0.03
5	3.06	59.84	0.09	0.13
6	2.55	51.81	0.32	0.35
7	2.19	43.78	0.44	0.38
8	1.91	35.74	0.13	0.10
9	1.70	27.71	0.00	0.00
10	1.53	19.68	0.00	0.00
11	1.39	11.65	0.00	0.00
12	1.28	3.62	0.00	0.00

As the Matthews function reveals the protein crystals could not be *BG4* antibody, in order to assess which contaminating protein it could be, a search of the PDB (protein data base) was conducted using the unit cell parameters obtained from the diffraction experiments. Using the abc and αβγ parameters of the resultant unit cell a match was found - *4RCB Hfq* protein. PDB data belonging to *4RCB* was run through a molecular replacement with MOLREP and a rigid body refinement with REFMAC resulting in a 3D model with near perfect overlapping electron density map indicating that the crystals were in fact crystals of Hfq.

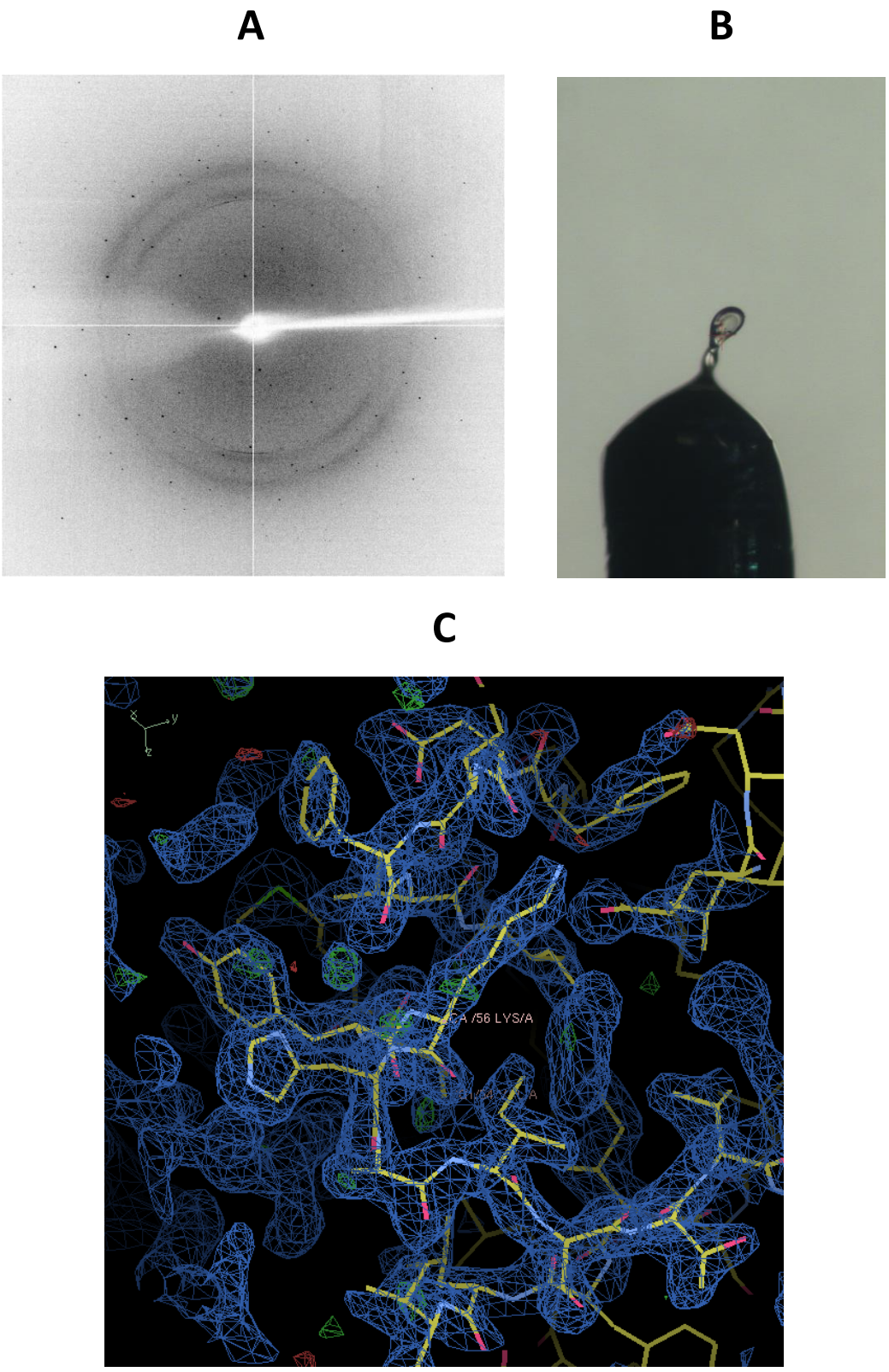


Figure 4-3. X-ray diffraction pattern from data set collected and resulting Hfq 3D model. (A) Representative X-ray diffraction pattern (ϕ = 90º) obtained on MX2 beamline of Australian Synchrotron at 13 keV, diffracted to beyond 2.99 Å. (B) Photo take at Australian Synchrotron of Hfq crystal on puck loop. (C) 3D model of Hfq hexamer made with rotation function calculate with MOLREP and a rigid body refinement with REFMAC.

4.3 Discussion

4.3.1 Recombinant expression and purification of *BG4* antibody

Using a model protocol reported earlier by Biffi et, al.[16] it has been shown that with the addition of a fluorescent tag to the *BG4* antibody we can successfully identify G4 DNA through structure specific binding. Since the antibody is not commercially available it had to be recombinantly expressed in *E.coli* using a method previously developed in our laboratory. This method successfully observed protein via peak absorbance (1349 mAU) at a concentration of 65% elution buffer. Elution fractions collected from the corresponding peak were run on polyacrylamide gel and displayed prominent 30 kDa bands indicative of BG4. The same 30 kDa band was also observed in the column flow-through pointing to poor binding of the antibody to the column. This poor binding was reflected in low yields in early expression attempts. Early expressions produced in the range of 375-515 µg of protein per litre of culture. By altering our laboratory protocol with an increased concentration of KCl in our lysis buffer (used to load sample onto the column) from 500 mM to 1 M KCl we were able to boost the amount of BG4 harvested from culture in the range of 804-978 µg per litre of culture. One expression even yielded as much as 2.3mg/l. Predictably, after this change the 30 kDa bands were no longer observable in our column flow-through.

4.3.2 X-ray data and Contamination

The processed data collected from the Australian Synchrotron reveal the crystals grown to have a 6 fold line of symmetry with a unit cell volume of 91826.727. Solvent content analysis of this crystal using a Matthews coefficient with an input molecular weight equal to that of BG4 (30743.6 Da) was unable to return any data for a protein of the that size. Together these results immediately excluded *BG4* from being the protein that had crystallized in our test. In an effort to identify an approximate size for a molecule that fit the X-rayed data a second Matthews coefficient was run using 1000 Da molecular weight. The results indicated that each asymmetric unit was occupied by a molecule approximately 7000 Da with a solvent content percentage of 43.78%.

A search of the online PDB (Protein Data Bank) using the unit cell dimensions returned for the P6 space groupings (a = 61.28, b = 61.28, c = 28.23,α = 90.00, β = 90.00, γ = 120.00) found E.coli *Hfq* 4RCB to fit our orientation model. A MOLREP function test run through REFMAC produced a model with significant overlapping electron density even correctly predicting location of internal water molecules inside the crystal. These results conclusively indicate the crystal grown was the E.Coli *Hfq* 4RCB a native E.Coli protein, a known contaminant in protein crystallographic methods that use IMAC during purification[24-26].

E.Coli Hfq is a host factor protein required for phage Qβ RNA replication. The Hfq family of proteins range in size between 70 - 110 amino acid residues[27-30]. Six crystallographic structures of 4 different Hfq protein expressed by E.coli have published. Identified in 2015 the *Hfq* variant 4RCB has a molecular weight of

7653.96 Da and length 71 amino acids arranged in a toroidal hexamotif. A conserved feature among *Hfq* protein members is their hexameric shape with a central canal that is lined with six surface histidine residue[25] (Figure 4-4) that have a high affinity for divalent cations (particularly Zn^{2+} and Ni^{2+}) present in metal-chelating resins. Simultaneously each monomer possesses a C-terminal run of 24 histidine residues in each monomer[31] cable of binding column resins in the same manner as the central cluster. When considering E.coli proteins that bind metal-chelating resins, a native His-tag along with the possession of surface exposed clusters of histidine are two of the most frequently cited reasons for contamination. *Hfq* contains both of these features and thus offers multiple sites for interaction with metal-chelating resins.

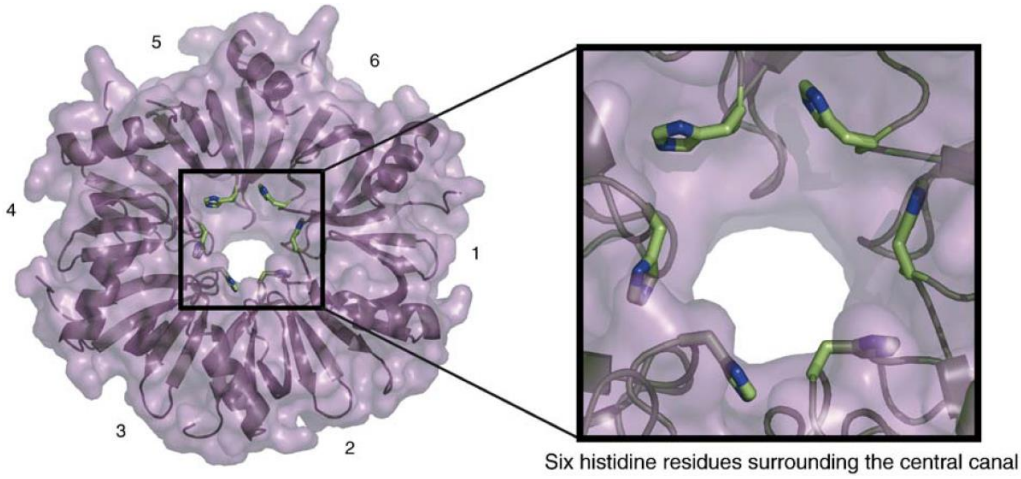


Figure 4-4. 3D space filling structural model of Hfq hexamer with ribbon model inlay. (Left) Complete hexamer toroid of functional Hfq protein. Each of the 6 monomer subunits have been labelled. (Right) Magnified view of the hexamer central canal. 6 exposed histidine residues are visible. Hfq may bind metal-chelating resin through this cluster of surrounding histidine residue[25]. Reprinted from Biochimica et Biophysica Acta (BBA)-General Subjects, 1760/9, Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. 2006 Sep 1;1760(9):1304-13, with permission from Elsevier.

In addition to native sequence based avidity for column binding, another functional property could also be responsible for *Hfq* co-purifying in the high concentration. The *Hfq* protein has been extensively studied as a modulator of gene expression via its binding to RNA[28,32-35]. Most of the studies focus on its binding to sRNAs and mRNAs, identifying as many as 22 different sRNA sequences capable of binding *Hfq*[35]. However, multiple studies have also observed Hfq protein binding DNA albeit at a reduced rate than RNA[36-38]. This DNA binding behaviour could contribute to Hfq copurifying in our tests. Gel analysis of our samples confirm *BG4* binding to our column and we showed it functionally capable of binding DNA by immunoflurescence assay. If the *BG4* bound to our column were also bound to DNA it could be offering an addition binding site for *Hfq* proteins as they pass through the column. A study by Updegrove and colleagues in 2010 showed *Hfq*-DNA interaction consist of multiple *Hfq* hexamers per DNAs complex (~500 bp in length). The study additionally demonstrated a higher affinity for curved DNA than so called mixed sequence DNA. This finding was supported by earlier reports from 1999 using 60 bp DNA duplexes[37]. It may be that this increased affinity to curved nucleic acid structure is also applicable to the curved quadruplex exposed surfaces. Considering that only a small amount of BG4-bound DNA could be capable of retaining several *Hfq* hexamers this may well explain why our sample became contaminated.

REFERENCES

1. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature. 1988;334(6180):364-366.
2. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nature structural & molecular biology. 2012;19(8):837-844.
3. Maizels N, Gray LT. The G4 genome. PLoS Genet. 2013;9(4):e1003468.
4. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic acids research. 2007;35(2):406-413.
5. De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou J-F, et al. Targeting telomeres and telomerase. Biochimie. 2008;90(1):131-155.
6. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie. 2008;90(8):1149-1171.
7. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002;417(6891):876-880.
8. Todd AK, Haider SM, Parkinson GN, Neidle S. Sequence occurrence and structural uniqueness of a G-quadruplex in the human c-kit promoter. Nucleic acids research. 2007;35(17):5799-5808.
9. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proceedings of the National Academy of Sciences. 2002;99(18):11593-11598.
10. Patel DJ, Phan ATn, Kuryavyi V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic acids research. 2007;35(22):7429-7455.
11. Smith NM, Raston CL, Smith CB, Sobolev AN. PEG mediated synthesis of amino-functionalised 2, 4, 6-triarylpyridines. Green Chemistry. 2007;9(11):1185-1190.
12. Smith NM, Corry B, Iyer KS, Norret M, Raston CL. A microfluidic platform to synthesise a G-quadruplex binding ligand. Lab on a Chip. 2009;9(14):2021-2025.
13. Smith N, Labrunie G, Corry B, Tran PLT, Norret M, Djavaheri-Mergny M, et al. Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands. Organic & biomolecular chemistry. 2011;9(17):6154-6162.
14. Monchaud D, Teulade-Fichou M-P. A hitchhiker's guide to G-quadruplex ligands. Organic & biomolecular chemistry. 2008;6(4):627-636.
15. Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, et al. Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. Nature chemical biology. 2012;8(3):301-310.
16. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nature chemistry. 2013;5(3):182-186.
17. Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. Journal of Biological Chemistry. 2011;286(47):41018-41027.
18. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallographica Section D: Biological Crystallography. 2010;66(2):133-144.
19. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallographica Section D: Biological Crystallography. 2013;69(7):1204-1214.
20. Padilla JE, Yeates TO. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. Acta Crystallographica Section D: Biological Crystallography. 2003;59(7):1124-1130.

21. Matthews BW. Solvent content of protein crystals. Journal of molecular biology. 1968;33(2):491-497.
22. Vagin A, Teplyakov A. Molecular replacement with MOLREP. Acta Crystallographica Section D: Biological Crystallography. 2010;66(1):22-25.
23. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography. 2011;67(4):355-367.
24. Milojevic T, Sonnleitner E, Romeo A, Djinović-Carugo K, Bläsi U. False positive RNA binding activities after Ni-affinity purification from Escherichia coli. RNA biology. 2013;10(6):1066-1069.
25. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochimica et Biophysica Acta (BBA)-General Subjects. 2006;1760(9):1304-1313.
26. Samuelson JC. Purifying Recombinant His-Tagged Proteins. 2016.
27. Franze dFMT, Hayward WS, August JT. Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. The Journal of biological chemistry. 1972;247(3):824-831.
28. Valentin-Hansen P, Eriksen M, Udesen C. MicroReview: The bacterial Sm-like protein Hfq: a key player in RNA transactions. Molecular microbiology. 2004;51(6):1525-1533.
29. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, et al. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Molecular cell. 2002;9(1):23-30.
30. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Molecular cell. 2002;9(1):11-22.
31. Beich-Frandsen M, Večerek B, Konarev PV, Sjöblom B, Kloiber K, Hämmerle H, et al. Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. Nucleic acids research. 2011;39(11):4900-4915.
32. Majdalani N, Vanderpool CK, Gottesman S. Bacterial small RNA regulators. Critical reviews in biochemistry and molecular biology. 2005;40(2):93-113.
33. Jousselin A, Metzinger L, Felden B. On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends in microbiology. 2009;17(9):399-405.
34. Papenfort K, Vogel J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. Research in microbiology. 2009;160(4):278-287.
35. Brennan RG, Link TM. Hfq structure, function and ligand binding. Current opinion in microbiology. 2007;10(2):125-133.
36. Kajitani M, Kato A, Wada A, Inokuchi Y, Ishihama A. Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q beta. Journal of bacteriology. 1994;176(2):531-534.
37. Azam TA, Ishihama A. Twelve species of the nucleoid-associated protein from Escherichia coli sequence recognition specificity and DNA binding affinity. Journal of Biological Chemistry. 1999;274(46):33105-33113.
38. Updegrove TB, Correia JJ, Galletto R, Bujalowski W, Wartell RM. E. coli DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2010;1799(8):588-596.

CHAPTER 5

Conclusion and future work

In this project, multiple applications of proteins as therapeutic agents were designed and tested. These studies attempt to solve some of the problems associated with the production, administration and targeting of drugs used in chemotherapy treatment of aggressive cancers. While the primary focus is directed to TNBCs much of the work has broader application for other cancer subtypes and protein therapeutics as a whole. TNBCs are a highly aggressive metastatic form of cancer that commonly become resistant to radiotherapy and experience frequent recurrences even after surgical removal. Current TNBC chemotherapies lack specificity which causes several systemic side effects for patients. When these nonspecific chemotherapies do successfully reduce tumours the positive results are all too often followed by multi-drug-resistant recurrences within five years. By looking at DNA structure and gene expression profiles rather than cell surface receptors it was thought that TNBC specific targets could be identified and exploited for improved chemotherapy. Building from this belief, nanoparticle delivery could be used to improve drug delivery and potency. In addition to novel application of protein therapeutics this project also investigates new means to produce high purity protein therapeutics, faster and by more cost effective methodologies. In this chapter, the further consequences of each of the studies will be discussed, speculation as to future research directions made, and a general conclusion provided.

5.1 Shear stress mediated refolding of proteins

We have presented a novel methodology for correctly folding proteins that harnesses mechanical energies which occur between fluid layers within rapidly rotating surfaces. This method distinguishes itself from existing methods that rely on time consuming dialysis or, either chemical or thermal means to alter protein structures. Recombinant protein overexpression is the most widely used method of producing proteins for laboratory use. However, as laboratories continue to identify new proteins for research interests they must optimize unique methodologies to express and harvest these proteins. Each unique protocol aims to produce biologically active forms of the new protein in large enough quantities to be useful for testing. Most often this means expressing a protein in bacterial cell cultures that typically result in insoluble and misfolded proteins. Our study has addressed this problem with a unique table top device called the Vortex Fluid Device (VFD). The VFD has been shown to take the protein inclusion bodies of three proteins; hens egg white lysozyme, caveolin-ΔTM and cAMP-dependent protein kinase A (PKA), and apply shear-stress in micrometre-wide thin fluid films to force unfolding of substrate proteins while segregating samples away from misfolding intermediates. The final protein product is a bio-active sample that is produced in only minutes at a fraction of the cost of traditional harvesting methodologies. In addition to reporting a single batch mode of operating, the VFD can also function in a continuous flow setting that lends itself to industrial applications as well as a laboratory table top device.

In the set of experiments conducted using the VFD we modulated the rotational speed of the tube container, the angle setting of the tube, the volume of fluid in the tube and the time a sample spent at a rotational speed. My suggestion for future development of this work is experimentation with the sample tubes itself. Two logical alteration that can be made would involve varying the size and surface material of the sample tubes. The VFD has the capability to receive sample tubes of different diameters (that is other than the 10 mm sample tubes used in this thesis). With the calculation of shear stresses offered in this thesis, future studies could test variants of VFD folding that could allow harsher or softer forms of folding should a protein require such processing. Secondly, the tubes used during experimentations were all glass sample tubes. A modified tube material could be developed to amplify or otherwise direct the intensity of shear forces beginning between the fluid and the wall of the container. For instance, a modified surface with a high contact angle, like a plastic or other polymer material, could enhance the turbulent flow and thus the applied shear stresses. Similarly a textured coating could be added to the inner surface of a sample tube to alter flow. These changes to sample tubes could result in changes to shear forces and the fluid film microenvironment. This could allow potentially greater control over protein alterations using the VFD in later works.

The different proteins used in our experiments were chosen because they were proteins of interest for other research conducted in our laboratory at the time. This familiarity meant we had existing protocols for recombinant expression and functional assays to aid our investigation of shear forces using the VFD. Coincidently, the secondary structure of all the proteins was predominately random

coil or alpha helical motifs. None of the investigated proteins were principally beta sheet in their secondary structure. Beta structure proteins represent a large fraction of secondary structures found in proteins and future application of the VFD technology should consider experiments using beta sheet proteins. The folding properties of beta sheets are not well understood, especially in comparison to helix or coil structures. Understanding of beta structure folding is difficult as attempts to create isolated beta-hairpin peptides often result in aggregate formations[1-4]. Refolding by VFD could potentially solve this problem via the same mechanism described in inclusion body aggregates. The small, all beta sheet protein, CspA would be an ideal protein to begin a future study. CspA is the major cold shock protein in *E. coli.* Its structure is well studied along with its folding properties which would aid data interpretation of VFD processed samples[5-7]. Interestingly, CspA represents a general single-stranded nucleic acid-binding motif[6]. Studying the behaviour of CspA in the VFD micro environment could pave the way for future endeavours applying shear forces towards nucleic acids for altered binding interactions.

5.2 Sensitizing basal-like breast cancer to chemotherapy using nanoparticles conjugated with interference peptide

In this work, nanoparticles were prepared using a PGMA polymer core. The PGMA nanoparticles were functionalised with RhB dye for *in vitro* imaging and covalently bonded to a second polymer PAA to provide an anionic surface. The change in nanoparticle surface charge was monitored with zeta potential and attachment was

recorded by size increases determined by dynamic light scattering. The anionic surface on the nanoparticles was used to facilitate electrostatic attachment of the anticancer interference peptide iPep. The intention of this study was to test a nanoparticle delivery system for the interference peptide iPep to confirm it as a TNBC and BLBC specific drug. This aim was achieved. Successfully delivering a nanoparticle formulation for iPep meant the study could expand to investigating combined drug delivery. The established anti-cancer drug docetaxel (DTX) was encapsulated using an oil in water emulsification process during preparation of the polymer nanoparticles. The combined DTX-iPep-nanoparticles had a 3% (w/w) DTX drug-loading that was confirmed by HPLC. DTX was chosen to be encapsulated following the results of combination index test that yielded a combination index score less than 1, indicating synergistic action with iPep.

In order to test the hypothesis that simultaneous nanoparticle delivery of iPep and DTX would sensitise TNBC/BLBC and improve drug potency, as proposed in the third chapter of this thesis, *in vitro* and *in vivo* cell viability tests were conducted. *In vitro* cell viability, assessed by CellTiter-Glo, of the combination nanoparticle drug therapy produced an IC50 value to be 0.23 mg/ml^{-1} after 24 h and 0.1 mg/ml^{-1} after 48 h. This combination was more potent than either component alone when delivered by nanoparticle. When compared to equal concentrations of the commercial gold stand drug Abraxane® the combined DTX-iPep-nanoparticles was significantly more potent at both the 24 h and 48 h time points. This is in spite of paclitaxel loading in Abraxane® being superior to DTX loading in our nanoparticles (10% versus 3%). *In vivo* analysis of DTX-iPep-nanoparticles performed in mice implanted with subcutaneous T11 tumour allographs.

Nanoparticles solutions were delivered via 5 intratumour injection every 2 days. At the end of the 14 day treatment cycle tumours treated with the combined nanoparticle formulation showed the greatest growth reduction and highest survival. Results were also compared to single drug nanoparticle administration and mutant iPep as well as Abraxane® treatment and each time the DTX-iPep-nanoparticles produced the most potent anti-cancer response. These findings are highly significant as they demonstrate a novel therapeutic strategy for triple negative and BLBCs that is both more potent and cancer cell specific than current commercial standard therapies.

In identifying an appropriate drug partner to be delivered with the EN1-iPep we examined two well-known anti-cancer agents DTX and DOX. DTX was chosen for future development over DOX because as it was found to synergistically enhance the action of iPep. In the same group of combination index testing we found DOX to be antagonistic and inhibit iPep potency and was therefore abandoned as a delivery partner for the EN1-iPep. One means to potentially improve the results obtained in this study could be to investigate other drug candidates for co-delivery. This could identify a more synergistic combined therapy while also importantly aid in the understanding why DTX was found to be synergistic and why DOX was not. To rationalise the observed synergist differences we must look more closely at each of the agent's mode of action. DTX induces apoptosis by binding microtubules present during active mitosis phase in the cell cycle. This binding stabilizes microtubules and prevents anaphase completion which tags the cell for apoptosis. Alternately, DOX acts by DNA intercalation, preventing topoisomerase II progression, during cellular transcription. Similarly the EN1-iPep interferes with EN1 gene transcription factor

to prevent gene transcription which leads to apoptosis. We believe the synergy observed between DTX and EN1-iPep occurred because they acted at different stages of cell cycle. In this way DTX and EN1-iPep could act as a check system for one another, where cells that first entered into transcription were targeted by EN1-iPep, and those that escaped targeting and progressed into anaphase were targeted by DTX to induce apoptosis. If this were true the lack of synergistic effect between EN1-iPep and DOX is due to both drugs acting at the same phase in the cell-cycle. In future work I would like to investigate the percentage of cell death that occurs during different stages of cell-cycle progression. Using methods outlined by Jackman and O'Conner, cell cultures can be synchronized in the G_1, S, and M[8]. Once synchronized I would repeat *in vitro* treatment of T11 cells with free EN1-iPep, with free DTX and DOX. Cell treated with EN1-iPep and DTX should show two distinct time points for reduction in cell survival during the S phase (EN1-iPep) and M phase (DTX). In contrast, cells treated with EN1-iPep and DOX should experience the majority of cell death during S phase. By comparing the percentage of cell death in each synchronized cell population we can infer the relative amount of cell death that is a result of each component of the combined drug delivery. This could direct future work to using alternate anti-cancer drugs and allow for new formulations of our existing combined drug to reflex necessary doses relative to component potency.

It remains a point of interest that the findings from combination index tests of EN1-iPep and DOX showed no enhanced effect on cell mortality than either component delivered alone. We suspect this might be due to a comparatively smaller proportion of DOX reaching the nucleus. Nuclear delivery of the EN1-iPep

occurs readily in cells due to both a cell penetrating peptide (CPP) sequence and nuclear localization sequence (NLS) present inside the peptide. In contrast, DOX must enter the cell and then the cell nucleus down a concentration gradient. To overcome this limitation we suggest future investigation into a conjugate form of DOX-EN1-iPep delivery. Previous reports by Liang and Yang have seen improved nuclear distribution and cell killing activity of DOX when conjugated to the well-known cell penetrating peptide TAT[9]. Using similar methods could conjugate EN1-iPep to DOX via amide bonding using NHS esters. A proposed reaction scheme is outlined in Figure 5-1. This conjugate structure potentially improves on the TAT conjugate by two means. Firstly the EN1-iPep is not simply acting as a CPP but also a potent killer of drug resistant cancer cells. Secondly the EN1-iPep's NLD can further improve nuclear localization of DOX to improve potency.

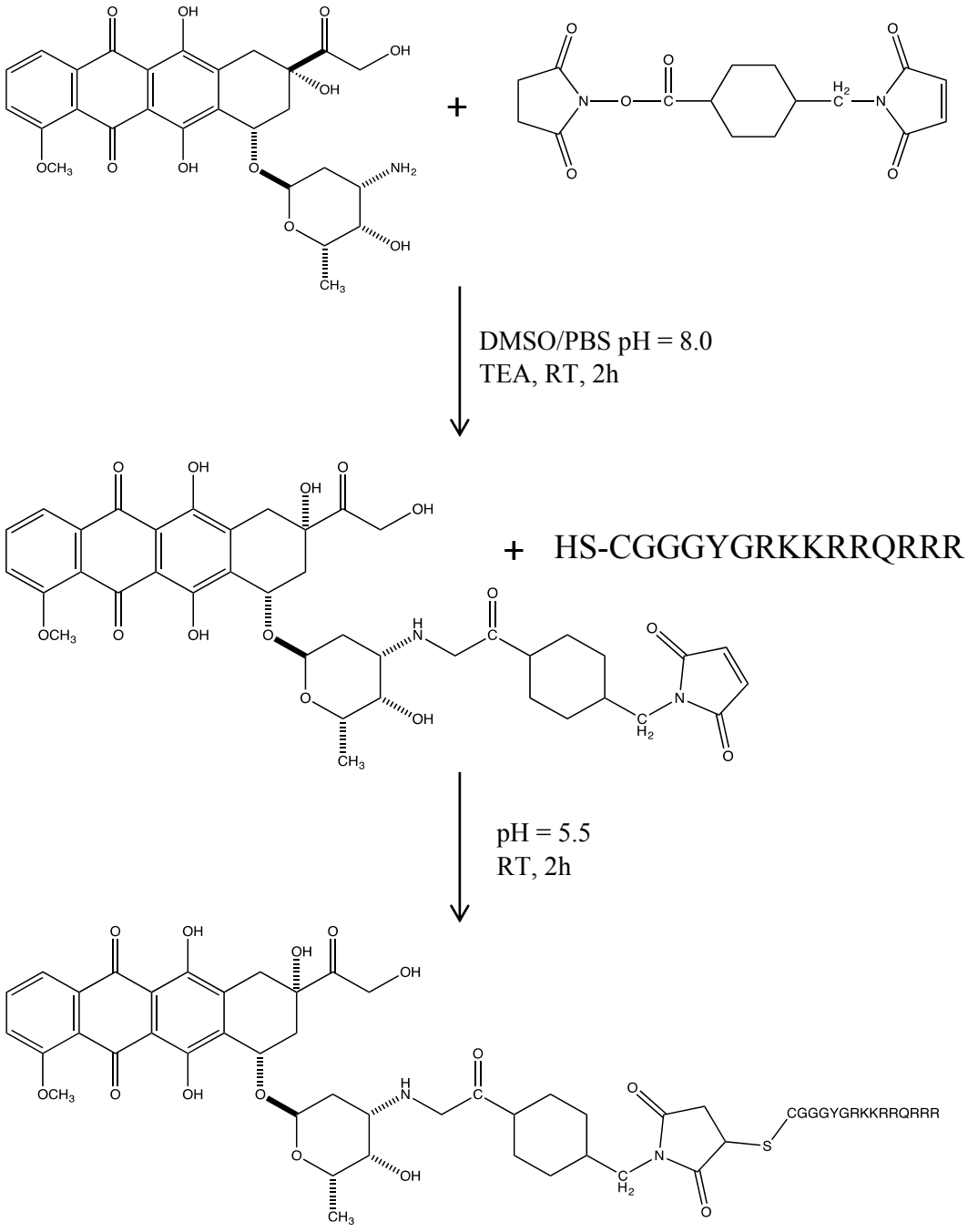


Figure 5-1. Reaction schematic for doxorubicin-EN1-iPep conjugation synthesis. 1, doxorubicin; **2**, succinimidyl-4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SMCC); **3**, doxorubicin-SMCC; **4**, EN1-iPep; **5**, doxorubicin-EN1-iPep conjugate.

5.3 Recombinant expression and crystallisation of G-quadruplex binding BG4 antibody

The aim of this work was to crystallise the recently identified single-chain antibody *BG4*. The *BG4* antibody has been discovered to specifically bind all possible structural conformations of G-quadruplex (G4) DNA. G-quadruplexes are highly mutagenic unique DNA structures that form in G rich regions of human DNA that concentrate in promoter regions, telomeric DNA and oncogenes. By crystallising *BG4* it was my intention to understand the nature of BG4 binding to G4s. This information could then be used to design small molecules with analogues binding pockets that could interfere with G4 stability in new anticancer therapies. Unfortunately the ultimate goal of a solved crystal structure of *BG4* was not achieved. *BG4* was successfully recombinantly expressed and yields improved upon by enriching KCl concentrations in the lysis buffer used during purification. Initially promising crystals were grown from a purified sample of *BG4* that were later found to be contaminated by traces of the E. Coli RNA binding protein Hfq 4RCB.

When reviewing the methodologies implemented in this study several improvements could be made to aid future attempts to solve the crystal structure of *BG4*. When the X-ray data collected from the Australian Synchrotron returned an X-ray diffraction pattern incompatible with BG4 this was quite a shock. Until this point 2D-PAGE gel analysis of the BG4 sample indicated we were using a highly purified sample. Never did we notice any protein bands around the 7653.94 Da (the molecular weight of Hfq) range. When assessing how this contaminant remained undetected I recognised that methodologies used Mini-PROTEAN®TGX Stain-Free™ Precast Gels. These gels were preferred because they drastically reduce run times

and imaging of gels to allow protein samples to be used fast and without the need to freeze samples between analysis and use. The gels include unique trihalo compounds that allow rapid fluorescent detecting when used with Bio-Rad stain-free imaging system Image Lab™. Crucially these trihalo compounds react with tryptophan (W) residues in protein to image samples. This was not thought be a problem as *BG4* contains six W in its sequence and would therefore allow for correct identification. Hfq contains 102 amino acid residues in it sequence, none of which are W. The absence of W residues prevented Hfq from being detected during purification and explains how it came to crystallise in the crystal screens. Subsequently, future work is advised reverting to traditional Coomassie Brilliant Blue staining protocols. This change will allow all proteins present in purified samples to be detected and appropriate purification steps can then be added into protocols to remove any contaminants in subsequent purified samples of *BG4*.

To successfully crystallize BG4, final protein samples must be concentrated and highly purified. The methods used employ a BG4 plasmid containing a non-cleavable His-tag that we used for purification with Ni^{+2} affinity columns. If the plasmid were to be redesigned to incorporate a cleavable His-tag this may improve sample purity. In the current protocol, protein collected from cell lysate is passed through a Ni^{+2} ion affinity column and the His-tag attached to BG4 binds to the column, separating the sample. However, any other protein containing strings of histidine residues will also bind to the column. If a new BG4 with a cleavable His-tag was expressed then the sample washed from the column could be mixed with a protease enzyme to cleave the His-tag. Passing the sample through the Ni^{+2} affinity

column a second time will bind any contaminating proteins with intrinsic histidine rich regions to the column and allow a more purified sample of BG4 to be collected.

Another simple way to improve BG4 purity is by adding additional chromatographic separation steps. The BG4 used for crystal screens in this thesis went through two different chromatographic separations. These were Immobilized Metal Ion Affinity (IMAC), and size exclusion chromatography (SEC). For a protein the size of BG4 (~30 kDa) there are wide range of chromatographic media that can be used during protein purification. Each method of chromatography separates proteins according to different properties inherent to the protein of interest. With each successive chromatography media the probability of any two proteins having the same properties as a desired protein of interest decrease, thereby increasing purity of desired protein in final sample. Following IMAC (using new BG4 with cleavable His-tag) the collected protein sample could additionally undergo both Ion Exchange Chromatography (IEX) and Hydrophobic Interaction Chromatography (HIC). IEX exploits electrostatic interactions between oppositely charged proteins according to their isoelectric point. By using a buffer with a greater pH than the pI of BG4 the protein will gain a net negative charge and bind to a positively charged, anion exchange resin. The current working buffer is adjusted to pH 8 and well suited for use in IEX. This is because IEX buffers are most effective when within 0.5 – 1.5 pH units of the pI of the protein of interest. Using the online tool ProtParam the theoretical isoelectric point of BG4 has calculated to 6.83, well within optimum range for IEX. Additionally, HIC can be used to separate proteins by their degree of hydrophobicity via Van der Waal interactions with column resins. HIC is a useful intermediate purification method, particularly when following IEX. Hydrophobic

interactions can be modified especially well by altering salt concentrations in running buffers. High ionic strength buffers increase hydrophobic interactions between protein residues and column resin. This is convenient when considering the running buffer suited for both IMAC and IEX was also increase from 500mM to 1M KCl. Finally, SEC would be used as a final polishing step as it was in the work presented earlier. Introducing additional chromatographic steps does significantly affect the overall time-cost of production and potentially expose samples to protein degradation. However high sample purity is key to successful protein crystallization and must be addressed in future work. An increased amount of initial protein should also be considered as successive chromatography will consume a small amount of protein in each step.

REFERENCES

1. Nesloney CL, Kelly JW. Progress towards understanding β-sheet structure. Bioorganic & medicinal chemistry. 1996;4(6):739-766.
2. Ramírez-Alvarado M, Blanco FJ, Serrano L. De novo design and structural analysis of a model β-hairpin peptide system. Nature Structural & Molecular Biology. 1996;3(7):604-612.
3. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable β-hairpin in aqueous solution. Nature Structural & Molecular Biology. 1994;1(9):584-590.
4. Searle MS, Williams DH, Packman LC. A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native β-hairpin. Nature Structural & Molecular Biology. 1995;2(11):999-1006.
5. Reid KL, Rodriguez HM, Hillier BJ, Gregoret LM. Stability and folding properties of a model β-sheet protein, Escherichia coli CspA. Protein science. 1998;7(2):470-479.
6. Newkirk K, Feng W, Jiang W, Tejero R, Emerson SD, Inouye M, et al. Solution NMR structure of the major cold shock protein (CspA) from Escherichia coli: identification of a binding epitope for DNA. Proceedings of the National Academy of Sciences. 1994;91(11):5114-5118.
7. Schindelin H, Jiang W, Inouye M, Heinemann U. Crystal structure of CspA, the major cold shock protein of Escherichia coli. Proceedings of the National Academy of Sciences. 1994;91(11):5119-5123.
8. Jackman J, O'Connor PM. Methods for synchronizing cells at specific stages of the cell cycle. Current protocols in cell biology. 2001:8.3. 1-8.3. 20.
9. Liang JF, Yang VC. Synthesis of doxorubicin–peptide conjugate with multidrug resistant tumor cell killing activity. Bioorganic & medicinal chemistry letters. 2005;15(22):5071-5075.