High-throughput screening technologies for identification and expression of functional domains of proteins of biomedical importance

PUBLISHED VERSION

A thesis submitted to the University of Sussex for the Degree of Doctor of Philosophy

Sarah Parry-Morris

June 2015

Declaration

I hereby declare that this thesis has not been and will not be submitted, in whole or in part, to another University for the award of any other degree.

Sarah Parry-Morris

Acknowledgements

I would like to thank my supervisor Laurence Pearl for the opportunity to undertake a PhD in his laboratory. I would also like to thank my second supervisor Antony Oliver for his helpful insight, ideas and support, and Chris Prodromou for his help and support during my PhD. Further thanks are given to Domainex, my industrial CASE placement for enabling my PhD, with special thanks to my industrial supervisors Trevor Perrior and Stefanie Reich for teaching of the CDH methodology and helpful discussions. A further special thanks goes to my mentor Keith Caldecott for his helpful discussions and overall support during my PhD. I would also like to extend my gratitude to the past and present members of the Pearl laboratory for their input, help and discussion during the duration of my PhD. Special thanks are for Mark Roe for his never ending help with crystallography, especially with fishing all the minute or really fragile protein crystals that I managed to grow. Additional appreciation goes out to the past and present members of the GDSC, who have made this an enjoyable experience.

I would also like to thank my family for getting me this far, especially my dad Ray for his encouragement over the past few years. Furthermore, I would like to thank all my friends who have come on this journey with me, for their great friendship, support and for all the good times had during my PhD; there are too many to mention, but special thanks goes out to Rachel, Heather, Ben, Tom, Grant, Emma, Harry, Nick, Finlay, Sophie and Thelma (the cat).

My research was funded by a Medical Research Council CASE studentship with Domainex, Cambridge, with further funding support provided by the University of Sussex.

Some parts of this thesis are redacted for reasons of confidentiality; these are published in a confidential version of this thesis.

University of Sussex

Sarah Parry-Morris

Doctor of Philosophy Biochemistry

High-throughput screening technologies for identification and expression of functional domains of proteins of biomedical importance

Summary

The ability to produce multi-milligram quantities of a recombinant 'target' protein in a proteolytically stable, soluble, and functional form is often necessary for subsequent biochemical, biophysical and structure-based analyses. Sub-constructs expressing only part of a large target protein can often be useful. Combinatorial Domain Hunting (CDH) is a methodology that allows the rapid production of sub-constructs via a random DNA fragmentation technique. One particular issue with CDH is that it can be used to identify globular regions or domains of a target protein, but does not take account of the functional properties of such domains; therefore some 'hits' are not useful, because they exclude these functional regions. Here, we have attempted to enhance the CDH methodology by including an additional screening step that could specifically identify those constructs expressing functional protein domains. However, whilst rigorous testing of this functionality screen proved it to be successful under selective conditions, it was not considered suitable for inclusion in the CDH method.

CDH was also used to identify highly expressed, proteolytically stable regions of a previously largely uncharacterized protein, and to investigate their functionality. Human Claspin is a large, highly charged, S-phase specific 'molecular scaffold' protein, with no identifiable sub-domains or enzymatic function(s). However, Claspin is known to make multiple different protein-protein interactions at replication forks during the intertwined processes of DNA replication and DNA replication-coupled repair. CDH successfully identified a number of N-terminal

expression constructs that could be expressed and purified to a high degree of homogeneity. Structural and functional analyses of these protein fragments indicated that the N-terminus of human Claspin is intrinsically disordered, and elongated in nature. However, these regions may become ordered upon binding to their respective protein or macromolecular partner(s). Furthermore, several N-terminal fragments were found to be able to bind to both single- or double-stranded DNA when longer than 16 nucleotides/base-pairs in length. Additionally, the phospho-specific protein-protein interaction made by human Claspin, with the checkpoint kinase Chk1 was further investigated.

Table of contents

Abbreviations

Aa	Amino acid
ABS	Absorption
ADP	Adenosine diphosphate
And1	Acidic nucleoplasmic DNA-binding protein 1
AP	Affinity probe
APC	Anaphase-promoting complex
ARM-repeat	Armadillo-repeat
ASEC	Analytical Size Exclusion Chromatography
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine triphosphate
ATR	ATM and RAD3-related
ATRIP	ATR-interacting protein
AUC	Analytical Ultracentrifugation
bp	Base-pair
BP	Basic patch
bDHFR	Bacterial dihydrofolate reductase
bioSAXS	Biological Small Angle X-ray Scattering
BLAST	Basic local alignment search tool
BLI	Bio-Layer Interferometry
BS3	Bis(sulfosuccinimidyl)suberate
CC	Coiled coil
CD	Circular Dichroism
CDC	Cell division cycle
CDH	Combinatorial Domain Hunting
CDH^2	Combinatorial Domain Hunting^2
CDH1	APC activator protein
CDK	Cyclin-dependent kinase)
CDT1	Chromatin licensing and DNA replication factor 1
C/EBPβ	CCAAT/enhance binding protein beta
Chk1 / 2	Checkpoint kinase 1 / 2
CIP	Calf intestinal phosphatase
CKAD	Chk1 activation domain
CKB	Chk1 binding
CKBD	Chk1 binding domain
CLPN	Claspin
CMG	Cdc45-MCM-GINS
CoESPRIT	Co-expressional Expression of Soluble Proteins by Random Incremental Truncation
CoFi	Colony filtration

Cy5	Cyanine-5
dATP	Deoxyadenine triphosphate
DBD	DNA binding domain
dCTP	Deoxycytosine triphosphate
DDK	Dbf4/Drf1-dependent CDC7 kinase
dGTP	Deoxyguanine triphosphate
DHFR	Dihydrofolate reductase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSB	Double strand break
dsDNA	Double-stranded DNA
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuridine triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
EMSA	Electrophoretic Mobility Shift Assay
Endo IV	Endonuclease IV
ESPRIT	Expression of Soluble Proteins by Random Incremental Truncation
FHA	Forkhead-associated
FITC	Fluorescein isothiocyanate
Flu	Fluorescein
FP	Fluorescence Polarization
FPC	Fork protection complex
FPLC	Fast Protein Liquid Chromatography
GFP	Green fluorescent protein
GINS	Go-ichi-ni-sans
GST	Glutathione S-transferase
GWL	Greatwall kinase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
His_6	Hexa-Histidine
HSQC	Heteronuclear Single-Quantum Correlation Spectroscopy
HTH	Helix-turn-helix
HU	Hydroxyurea
IDP	Intrinsically disordered protein
IDR	Intrinsically disorder region
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMAC	Immobilized Metal Affinity Chromatography
ITC	Isothermal Titration Calorimetry
Kb(p)	Kilobase(pair)

K_d	Equilibration dissociation constant
KD	Kinase domain
kDa	Kilodalton
Klenow	DNA Polymerase I, Large (Klenow) Fragment
L-agar	Luria-Bertani agar
LB	Luria-Bertani medium
LDS-LB	Lithium Dodecyl Sulphate Sample Buffer – β-mercaptoethanol – Loading buffer
MCM2-7	Mini-chromosome maintenance 2-7
mDHFR	Murine dihydrofolate reductase
MEK1	MAPK/ERK kinase 1
MES	2-(*N*-morpholino)ethanesulphonic acid
MRC1	Mediator of replication checkpoint 1
mRNA	Messenger RNA
Native-PAGE	Native polyacrylamide gel electrophoresis
NFκB p65	Transcription factor p65
Ni-NTA	Nickel nitrilotriacetic acid
NIK	NF-kappa-beta-inducing kinase
NMR	Nuclear magnetic resonance
NMR NI	NMR non-inducing starter medium
NRMSD	Normalized root mean squared deviation
nt	Nucleotide
OCTET	ForteBio Octet RED96
OD	Optical density
ORC	Origin Recognition Complex
ORF	Open reading frame
P1 / 2 / 3	Passage 1 / 2 / 3
p50	Hsp90 co-chaperone Cdc37
PB2	Polymerase basic protein 2
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline tween
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
Plk1	Polo-like kinase
PolH	Polyhedrin
Poly-Glu	Poly-glutamic acid
PPIs	Protein-protein interactions
PTM	Post translational modification
PVDF	Polyvinylidene difluoride
Rad9-Rad1-Hus1	9-1-1 complex
RC	Replicative complex

RFC	Replication factor C
RFID	Replication fork interaction domain
Rg	Radius of gyration
RHR	Rel-homology region
RPA	Replication Protein A
S. cerevisiae/sc	*Saccharomyces cerevisiae*
SA	Streptavidin
SCF	Skp, Cullin, F-box containing complex
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Sedimentation equilibrium
SEC	Size exclusion chromatography
*Sf*9	*Spodoptera frugiperda* 9
SGC	Structural Genomics Consortium
S. pombe/sp	*Schizosaccharomyces pombe*
ssDNA	Single-stranded DNA
STU	Staurosporine
SUMO-tag	Small ubiquitin-like modifier tag
SV	Sedimentation velocity
T7F	T7 forward primer
T7R	T7 reverse primer
TAE	Tris-acetate-EDTA
TB	Turbo broth
TBE	Tris-borate-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N,N,N'N'-tetramethylethylenediamine
TFE	2,2,2-trifluoroethanol
TIM	Timeless
TIPIN	Timeless-interacting protein
*T*m	Melting temperature
TMP	Trimethoprim
TopBP1	Topoisomerase-binding protein-1
T-PCR	Tagged random primer PCR
Tris	Tris(hydroxymethyl)aminomethane
UDG	Uracil-DNA glycosylase
USP	Ubiquitin carboxyl-terminal hydrolase
X-gal	5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside

List of figures

List of tables

List of equations

Chapter 1
Introduction

1.1 Overview

This introduction serves as a reference point for the other chapters presented in this thesis. It starts with a description of the practical approaches currently used to overcome the difficulties of recombinant protein expression in heterologous hosts, especially as many of the so-called 'difficult to express' proteins are actually considered as highly attractive targets for therapeutic intervention in the treatment of disease (Littler, 2010). A number of different approaches have therefore been developed, to enable the production of (recombinant) proteins at sufficient quantity, quality and purity to allow downstream structural, biochemical and biophysical studies, as well as drug discovery efforts. Therefore, specific details have been provided for the construction of high-throughput DNA fragmentation and protein expression libraries as a means of identifying optimal expression constructs. A particular emphasis has been placed on the Combinatorial Domain Hunting (CDH) technique, which is used and referred to in subsequent chapters. The second part of the introduction focuses on the human protein Claspin; a target selected for the aforementioned CDH methodology. Claspin is a large 151 kDa protein required for the intertwined processes of DNA replication and replication-coupled repair. Finally, this chapter culminates with a description of a number of biochemical and biophysical techniques (again used in this thesis) for the characterisation and/or determination of the structure of a recombinant protein.

1.2 Recombinant protein expression

1.2.1 Traditional approaches to recombinant protein expression

Typically, recombinant proteins are expressed in the heterologous host *Escherichia coli* (*E. coli*) and this often yields recombinant protein of the required quality and quantity, whilst remaining highly cost effective. However, expression in *E. coli* does not always result in highly expressed, proteolytically stable and/or soluble recombinant protein. Approaches to overcoming such difficulties in recombinant protein expression, solubility and/or stability are often lengthy and low throughput, requiring many iterative rounds of trial and error, with focus usually on a single target at any one time; this type of approach can be particularly

expensive. There are a large number of protocols and trouble-shooting guides that have been published in order to enable the overcoming in difficulties in recombinant protein expression and purification including; Graslund et al., (2008a), Savitsky et al., (2010), and EMBL (2015). Some of the more traditional approaches are detailed below, but these are in no ways exhaustive, and successful modifications to established protocols are normally highly protein specific.

The DNA sequence of a gene can be modified in order to optimize its codon usage, such that it is specifically tailored to the selected heterologous host. The amino acid sequence of the protein can itself be mutated, such that it resembles those of successfully expressed homologous proteins, or in the case of enzymes, produce an inactive form (recombinant protein functionality can interfere with the function of the host cell resulting in cellular toxicity). Expression constructs can split large and/or multi-domain proteins into their functional components to enhance recombinant protein expression; this approach is discussed in greater detail in section **1.2.2**.

Protein expression tests, typically involve the use of an N-terminal Hexa-Histidine-(His_6-) affinity tag; a small tag that enables a specific purification strategy to be used and can also be used for protein identification (western blot), it also does not typically interfere with the functionality of the protein or enhance the solubility of the protein. If enhancement of solubility is desired, the recombinant protein can be expressed as a fusion with other proteins that are intrinsically highly soluble, such as GST (Glutathione S-transferase) or SUMO (Small Ubiquitin-like Modifier). A number of different *E. coli* expression strains can also be used to enhance protein expression; these include but are not limited to, protease-deficient strains to enhance recombinant protein stability (e.g. BL21); strains co-transformed with plasmid encoding tRNAs for rarely used codons (e.g. RIL, pRARE, pRARE2); strains which also express lysozyme in order to tightly control recombinant protein expression and typically used for the expression of toxic proteins (e.g. pLysS, pLysE). Different types of expression medium can be used for cell culture, along with changes in expression conditions: for example, growth temperature, length of induction, and the type and applied concentration of the inducer of recombinant

protein expression [e.g. IPTG (Isopropyl β-D-1-thiogalactopyranoside), arabinose]. Co-expression of *E. coli* molecular chaperones (e.g. GroEL-GroES, DnaK-DnaJ-GrpE) or a known interaction partner protein can enable the correct folding of a recombinant protein or can result in its stabilisation.

During purification of the recombinant protein, the method of cell lysis, the extraction buffer and subsequent purification steps can all be modified in order to enhance yields and purity of the recombinant protein. If a protein can be expressed, but remains insoluble, it can be purified under denaturing conditions. Downstream attempts can be made at refolding the protein, but can be notoriously difficult and time-consuming, as some form of assay is generally required in order to confirm that the protein is now correctly folded.

Many alternative heterologous hosts are available for recombinant protein expression including insect (e.g. *Spodoptera frugiperda* (*Sf*9), High FiveT) or mammalian cells (e.g. HEK293T), as well as cell-free expression systems (e.g. *E. coli* or wheat germ lysates). However, these are often more time-consuming in the preparation of constructs and for recombinant protein expression, and more expensive due to the more specialist equipment and medium required for culture.

1.2.2 Design of expression constructs

As previously mentioned, large and/or multi-domain proteins typically do not express well from *E. coli*. Furthermore, it has long been known that fragments (sub-constructs) of target genes that encode domains are often more tractable, yielding soluble, proteolytically stable protein that is correctly folded (Graslund et al., 2008b, Littler, 2010). A ‘domain’ can be defined as a region of a protein that forms a stable and folded tertiary structure, even when isolated from the rest of a parent protein. However, the accurate prediction of the amino acid boundaries of a given domain within a protein sequence, typically requires some degree of sequence similarity and/or homology to other proteins where the structure of the domain is already known. As of yet there are no universally accepted formulas to define the amino acid boundaries of a sub-construct, which will guarantee the production of high levels of recombinant protein. The presence or absence of a few

amino acid residues at either end can greatly affect the levels of expressed protein as shown by Klock et al., (2007).

If any recombinant protein can be expressed, limited proteolysis experiments can be used to identify the presence of stable sub-fragments of protein, the boundaries of which can be ascertained by Edman degradation (N-terminus) and / or mass spectrometry. This method of construct design is highly conditional and low-throughput in nature, but can be highly effective.

In recent years, there has been a movement towards high-throughput recombinant protein expression trials, to find both optimal expression conditions and expression constructs, as reviewed in; Prodromou et al., (2007), Savva et al., (2007), Littler, (2010), Savitsky et al., (2010), Yumerefendi et al., (2011), and Hart and Waldo, (2013). High-throughput expression construct design methods either involves a systematic-type approach through bioinformatics, using approximate domain boundaries and nested construct design which is described in more detail in section **1.3**, or can involve 'random' approaches by DNA mutation or fragmentation, in order to create a library of random expression constructs and to subsequently identify domain-containing constructs, described further in section **1.4**. Both types of approaches use high-throughput expression screening in order to identify optimal expression constructs.

1.3 Bioinformatics-based construct design

Bioinformatics-based construct design methods use the amino acid sequence of a protein target as an input for a number of prediction algorithms in order to design expression constructs for a specific protein domain/s. Many groups and consortiums have been established in order to solve the tertiary structure of proteins in the genome using high-throughput bioinformatics-driven initiatives, termed 'Structural Genomics'. One collaborative consortium, the Structural Genomics Consortium (SGC), was set up to in 2000, to solve and publically release the tertiary structures of biologically important proteins (Williamson, 2000), and to date the SGC have released over 1500 structures.

For a given protein, the secondary structure can be predicted from the amino acid sequence; such as regions of α-helices and β-strands, and regions of likely sequence disorder, whilst the tertiary structure can be modelled to identify globular folded regions, by template-based methods; such as homology modelling and fold recognition / protein threading (Dorn et al., 2014). Homology modelling identifies sequence homology between a target sequence and a sequence of a known structure, and using the known structure of the template protein, the tertiary structure of the target protein is predicted (Xu et al., 2008). For accurate modelling this requires high sequence homology; reduced sequence homology results in inaccuracies in the predicted model. Protein threading is used for identifying folds in proteins for where a homologous structure is not known. This identifies known sequence folds in the protein sequence based on amino acid composition and not on sequence homology. A model is built based on the fold template/s identified in the amino acid sequence, where confidence in the predicted structure is based on the similarity of amino acids between the sequences (Marti-Renom at al., 2000). Additionally, the tertiary structure of a protein can be computationally derived directly from the amino acid sequence by *ab initio* / *de novo* structural prediction algorithms. However, the generation of *ab initio* tertiary structures requires vast computing power and there has only been limited success for small proteins. Together, these programs are used to search for domains within the amino acids sequence by comparison with related structures, regions of secondary structure and disorder, and regions of sequence conservation, which are also designed to enable estimation of the N- and C-termini of constructs (Graslund et al., 2008a, Savitsky et al., 2010).

For bioinformatics based techniques once a domain or other 'region of interest' has been identified in the amino acid sequence of a target, a series of expression constructs are created, typically with nested boundaries (2-10 amino acid steps) spanning the domain of interest but also taking into consideration the amino acid sequence surrounding the selected domain. These are cloned into expression vectors, such that they are in-frame with an N-terminal His_6 affinity tag, and small-scale high-throughput expression and purification studies are undertaken. Constructs identified as expressing recombinant protein, can be subsequently fine-

tuned by refinement of the encoded domain boundaries, i.e. with a smaller step increase and/or decrease, thus allowing the iterative improvement of the construct and domain boundaries. More 'traditional' modifications to expression can then be used to improve recombinant protein expression, as explained previously in section **1.2.1**.

Although methods to accurately predict protein domain boundaries are being constantly improved, many eukaryotic proteins do not have close homologues or contain domains with known folds, making it much more difficult to rationally design expression constructs. Additionally, a particular domain may require additional amino acids at either the N- and / or C-terminus for soluble expression of a protein fragment. An alternative to the bioinformatics-driven approach is the generation of a so-called 'random DNA-fragmentation library' approach as discussed in the following section.

1.4 Random DNA-fragmentation libraries

Termed the 'domain-hunting' paradigm, combinatorial libraries of randomly generated DNA fragments can be created, which are typically designed to give full coverage of a particular gene of interest. Unlike bioinformatics-lead approaches, this method does not require any prior knowledge or prediction of domain boundaries. However, if a domain of the protein is of specific interest, a library can be built for just this region of a protein. As with bioinformatics driven approaches, fragmentation / mutation library building is typically followed by high-throughput expression screening of the generated constructs. A number of reviews on library building methods and high-throughput screening techniques are available including; Prodromou et al., (2007), Savva et al., (2007), Littler, (2010), Yumerefendi et al., (2011), and Hart and Waldo, (2013).

Many different laboratories have individually developed domain-hunting methodologies, which whilst often unique in their experimental approaches, follow a similar step-wise methodology for the identification of individual high-expression constructs. Briefly, the first step starts with random mutation or fragmentation of an individual DNA clone, to generate a library of (typically)

hundreds of thousands of clones. Generation of such DNA libraries can be by random sequence mutation, or by DNA fragmentation using physical, enzymatic or PCR-based fragmentation methods (explained in more detail in subsequent sections). The mutated or fragmented DNA is then sub-cloned into an expression vector suitable for high-throughput expression testing. The expression profile of the library is then sampled, looking at expression levels and solubility of recombinant protein fragments with respect to a reporter fusion protein. However, a number of assumptions are made during library screening including: that the fusion reporter is not interfering with or affecting the solubility of the recombinant protein, and that the soluble recombinant proteins are in the correctly folded state. Screening leads to a number of 'initial hits' each of which meets certain criteria; these are identified by DNA sequencing. Selected constructs are subjected to rigorous tests examining protein expression, solubility and proteolytic stability.

Explained in more detail below are a number of different DNA-fragment based methodologies, each of which have been implemented and developed in order to overcome the shortcomings of recombinant protein expression; some have also been developed to enable study of protein-protein interactions (PPIs).

1.4.1 DNA shuffling

An early technique used for the enhancement of protein expression was 'DNA shuffling' (Stemmer, 1994a, Stemmer, 1994b). This method involves random DNaseI fragmentation of a gene of interest, with a series of gene homologues. The DNA is denatured and allowed to reassemble together, mixing homologue gene sequences together randomly. Full-length DNA constructs are identified by a PCR reaction and these are cloned. These are subsequently used to transform *E. coli*, to screen for the phenotype (section **1.4.3**). The best constructs are selected, and the fragmentation and reassembly is repeated a number of times to ensure 'DNA shuffling'. To eliminate non-essential mutations, this is repeated a couple of times using the shuffled DNA and the original DNA sequence of interest.

1.4.2 T-PCR

The Tagged random primer PCR (T-PCR) method, uses a PCR reaction to amplify a gene of interest, but the primers are designed such that they can randomly associate with the DNA template (Grothues et al., 1993). The primers are typically 21 bp in length, with a restriction site at the 5' end, to allow downstream cloning, whilst the 3'-15 bp are random bases that do not completely anneal to the gene to be amplified. These constructs are subsequently used to transform *E. coli*, to screen for the phenotype (section **1.4.3**). This was the first detailed method created to identify soluble domains from large proteins (Kawaski and Inagaki, 2001).

1.4.3 GFP reporter assay

GFP has been used as a fusion protein in order to screen colony recombinant protein expression. GFP is expressed as a C-terminal fusion partner of a target protein, and the level of GFP fluorescence is directly correlated with the level of recombinant protein expression (Waldo et al., 1999). The fluorescence of GFP is wholly dependent on the correct folding of the recombinant protein by the heterologous host. This type of reporter assay was initially used to identify soluble proteins from DNA shuffling, where the 'best' expression clones were selected based on GFP fluorescence (section **1.4.1**) (Waldo et al., 1999, Waldo et al., 2003). This screen was subsequently used for colony expression screening of a random-fragment library generated from T-PCR, measuring fluorescence intensity for each expression fragment clone (section **1.4.2**). Subsequently the highest GFP fluorescence clones are identified, and the solubility of the protein from each expression construct was tested without fusion to GFP (Waldo et al., 1999, Kawaski and Inagaki, 2001, Waldo, 2003).

This screening method was used to obtain a soluble expression construct of NPD (nucleoside diphosphate) kinase from *Pyrobaculum aerophilum* for protein crystallography (Pedelacq et al., 2002). DNA shuffling was also used to develop a GFP-variant that was more susceptible to the misfolding of the upstream fusion-protein to facilitate the ease of library screening (Cabantous et al., 2008). Additionally, this screening approach was used to investigate the multi-domain murine Vav (guanine nucleotide exchange factor) protein, successfully identifying

four soluble domains that corresponded to the known domain architecture of the protein (Kawaski and Inagaki, 2001). Furthermore, this was also used to study the multi-domain protein TERT (Telomerase Reverse Transcriptase) from *Tetrahymena thermophila*, and this identified constructs that covered all of the identified domains of the protein (Jacobs et al., 2005).

1.4.4 Error-prone PCR

A similar GFP-reporter assay was also developed for the mutation method 'error prone PCR' which incorporates random mutations into the DNA, whilst using a fusion to ZsGreen (reef coral fluorescent) protein as the reporter protein (Heddle and Mazaleyrat, 2007). Error prone PCR is carried out using Taq DNA polymerase with an unbalanced concentration of nucleotides, to introduce random DNA point mutations throughout the amplicon. These fragments are cloned in-frame with an N-terminal His_6 affinity tag and a C-terminal ZsGreen reporter, where highly expressing clones are selected by the fluorescence levels of the ZsGreen reporter. This method was successfully used to identify mutant expression constructs for a number of receptor tyrosine kinases (RTKs) with improved soluble protein expression (Heddle and Mazaleyrat, 2007).

1.4.5 Split-GFP Fusion

A known problem with the GFP-reporter fusion approach was the solubilisation of the recombinant proteins during a screen, leading to a high number of false positives. The split-GFP-reporter was developed as an alternative (Cabantous et al., 2005a, Cabantous et al., 2005b, Cabantous and Waldo, 2006). GFP is an 11 β-strand protein that forms a β-barrel structure, and this was successfully split into two parts: β-strands 1-10 (GFP1-10) and β-strand 11 (GFP11) [optimized by DNA shuffling (section **1.4.1**)]. Both parts of the protein remained soluble and maintained the ability to interact, resulting in a detectable fluorescent signal. GFP11 is fused to the C-terminus of a recombinant protein, and is co-expressed with GFP1-10. This method was shown to not result in passenger protein solubilisation of the fused recombinant protein and could be used as a downstream step in a number of different DNA fragmentation methods (Listwan et al., 2009, Cabantous et al., 2005a, Cabantous et al., 2005b, Cabantous and Waldo,

2006). Additionally, a tripartite split-GFP fusion system was developed for the examination of PPIs (Cabantous et al., 2013). This system splits the GFP protein further, into β-strands 1-9 (GFP1-9), β-strand 10 (GFP10) and GFP11. GFP10 and GFP11 are used to individually tag two different proteins, and association of these proteins is required for the interaction of the three separate GFP components for fluorescent signal maturation. This system has been tested in *E. coli* and has been extended into mammalian cells to study both *in vivo* and *in vitro* interactions (Cabantous et al., 2013).

A further modification to the Split GFP-fusion was the introduction of a 'bead screen' (Lockard et al., 2011). Fragment constructs were cloned between an N-terminal His_6 affinity tag and GFP11 at the C-terminus. Recombinant split-GFP protein expression of transformed *E. coli* colonies is induced on L-agar plates containing IPTG. The induced *E. coli* colonies were replica-plated using a Durapore membrane (colony side up) onto agarose plates that contained Immobilised Metal Affinity Chromatography (IMAC) resin to bind the N-terminal His_6 affinity tag. Colonies were lysed and the soluble recombinant protein can pass through the Durapore membrane to bind to the affinity resin. Any tagged protein that was fused to folded GFP formed a fluorescent spot in the agarose, enabling detection of colonies that expressed a soluble, in-frame (at the N- and C-terminus) recombinant protein (Lockard et al., 2011). Additionally, this Split GFP-fusion bead screen method was further developed to allow the analysis of multi-protein complexes: the affinity tag is fused to one protein and the GFP 11-peptide to another and these are co-expressed with GFP 1-10. This method was tested with 8 different dimeric and trimeric complexes (not subjected to library screening), and was successful in identifying stable complex formation (Lockard et al., 2011).

1.4.6 DHFR reporter assay

A further example of a random fragmentation library approach for the identification of soluble expression constructs is one using a dihydrofolate reductase (DHFR) reporter (Liu et al., 2006a, Dyson et al., 2008). The DHFR protein (conserved from *E. coli* to higher eukaryotes) converts dihydrofolate to tetrahydrofolate for the production of essential co-factors for the *de novo* synthesis

of purines and certain amino acids (Chen et al., 1984). The bacterial DHFR (bDHFR) protein is essential for cellular viability and this can be inhibited by trimethoprim (TMP). One method using this system creates mutations by DNA shuffling (section **1.4.1**), and constructs are ligated into an expression vector upstream of a murine DHFR (mDHFR) reporter gene. Transformed *E. coli* are grown on L-agar plates containing IPTG and increasing concentrations of TMP; the assay uses mDHFR as the reporter protein as is not inhibited by TMP. Soluble recombinant protein expression enables the survival of the transformed *E. coli* colony. This method was used to identify a number of solubilizing mutations in an acetyltransferase from *Vibrio fischerii* (Liu et al., 2006a).

Recently, the DHFR reporter assay has been used for screening a fragmentation library using a deoxyuridine triphosphate (dUTP)-containing PCR reaction created as described below for Combinatorial Domain Hunting (CDH) (section **1.5**). DNA fragmentation is carried out using endonuclease V [with manganese (II)] to cleave double strand DNA 3' to the inserted dUTP. The expression screen was carried out as described above, and was used to identify soluble domains from FLI1 (Friend Leukaemia Integration-1) transcription factor, and the cytoplasmic domain of the trans-membrane protein Pecam1 (platelet endothelial cell adhesion molecule-1) (Dyson et al., 2008).

1.4.7 Two-body *E. coli* DHFR assay with split GFP-fusion

As with fusion to GFP, fusion of recombinant proteins to DHFR can result in passenger protein solubilisation effects, i.e. false positives. Therefore this method requires downstream re-cloning and expression testing to determine if the fragment protein is soluble (Dyson et al., 2008). In order to circumvent the problems with 'false positives', a "two-body *E. coli* DHFR" assay (Pedelacq et al., 2011) was developed using a split *E. coli* DHFR gene (Smith and Matthews, 2001). This multistep-screen enables the detection of 'in-frame' expression, and due to their small size eliminates / reduces passenger protein solubilisation. A DNA fragment library generated by DNaseI or Exonuclease treatment is ligated into an expression plasmid between the two halves of the *E. coli* DHFR gene. Transformed *E. coli* are then are grown on L-agar plates containing IPTG and TMP, and only

colonies expressing functional mDHFR survive. These clones are picked and the coding DNA is re-cloned into the split-GFP system (section **1.4.5**). Subsequently, transformed *E. coli* expressing a soluble GFP-fusion are selected. This method was used to study the PpsC (polykeytide synthase) multifunctional, multi-domain enzyme required for virulence of *Mycobacterium Tuberculosis*; this identified constructs expressing each of the six domains of the protein, and the X-ray crystal structure was solved for three of them (Pedelacq et al., 2011).

1.4.8 Nested deletion library

The nested deletion method was developed to 'fine-sample' the 5' and 3' ends of particular gene constructs, and hence the N- and C- terminal ends of the encoded protein (King et al., 2006). The PCR amplified gene is treated with Mung Bean nuclease, which has relatively poor activity on blunt-ended double-stranded DNA, resulting in 'nested' ends of a particular DNA construct. The resulting fragments are cloned into an expression vector containing an N-terminal His_6 affinity tag and a C-terminal chloramphenicol acetyltransferase (CAT) fusion. The expression library is grown on L-agar plates containing chloramphenicol. Only *E. coli* transformed with in-frame expression constructs, expressing soluble protein are able to grow on the selective medium. The CAT-screening protocol was initially used for the investigation of point mutations in the HIV (Human Immunodeficiency Virus) integrase domain (Maxwell et al., 1999). Additionally, this was used to identify expression constructs for two *Saccharomyces cerevisiae* (*S. cerevisiae*; budding yeast) proteins essential for gene silencing, specifically to investigate homodimerisation of Sir3 and the interaction of Sir3 with Sir4 (King et al., 2006).

1.4.9 Shotgun proteolysis for stable domain identification

An alternative screening technique using phage display libraries combined with protease resistance was developed to identify protease resistant protein fragments (Christ and Winter, 2006). The coding DNA of a protein target is first fragmented using DNA shearing, and the fragments are cloned via adapter sequences into a phagemid vector. These fragments are expressed as fusion proteins with a N-terminal affinity tag (barnase) and a C-terminal pIII protein (a phage coat protein), and are thus displayed on the surface of the resulting phage. The phage display

library is incubated with trypsin, and the barnase affinity tag used to capture any protease resistant fusion-proteins. As proof of concept, this technique was applied to the *E. coli* genome, where it identified a number of compact globular proteins (that had already previously been identified). It was also shown this method could be used to capture the individual domains of a number of multi-domain proteins for which structures had already been determined (Christ and Winter, 2006).

1.4.10 ESPRIT

Expression of Soluble Proteins by Random Incremental Truncation (ESPRIT) uses either uni- or bidirectional DNA digestion by endonuclease III in order to create a DNA fragmentation library (Tarendeau et al., 2007). This DNA fragment library is ligated into an expression vector encoding an N-terminal His_6 affinity tag and a C-terminal biotin acceptor peptide, a 15 amino acid acceptor peptide, which is biotinylated *in vivo* by the BirA enzyme (produced by the heterologous host) (Beckett et al., 1999). Transformed *E. coli* are then are grown on L-agar plates containing IPTG, and the expression of both tags is determined using colony blotting. This enables the detection of 'in-frame' expression, and due to their small size eliminates / reduces passenger protein solubilisation (Tarendeau et al., 2007, Yumerefendi et al., 2010, An et al., 2011). The ESPRIT method has been used to study the PB2 (Polymerase Basic 2) domain protein of the human influenza A polymerase (H3N2), for the identification and structural / functional analysis of the domains within this protein (Tarendeau et al., 2007, Guilligay et al., 2008, Tarendeau et al., 2008). This protein had no identifiable homologues and rational construct design had previously failed to design expression constructs that produced soluble protein (Tarendeau et al., 2007). Additionally, the ESPRIT method has been used for the following protein targets: a DNA terminase nuclease from herpes virus 5 (human cytomegalovirus) (Nadal et al., 2010), *Bacillus subtilis* SpoIIE (sporulation protein) (Rawlings et al., 2010), Rif1, a protein involved in the regulation of origin firing and DNA repair (Sukackaite et al., 2014), CagA domains from *Helicobacter pylori* (Angelini et al., 2009), and the 320 kDa human neurofibromin protein (Bonneau et al., 2009).

The ESPRIT methodology was further developed to enable the study of PPIs. Termed CoESPRIT, this method involves co-expressing a DNA fragment library with a N-terminal FLAG-tagged 'bait' protein to identify complex formation. CoESPRIT has been used to identify two interactions made by PB2, with PB1 and importin α-1 (An et al., 2011).

1.4.11 Colony Filtration blotting

The Colony Filtration (CoFi) blot, screens a fragmentation library created through exonuclease III resection of the target gene with a high-throughput expression and solubility screen (Cornvik et al., 2005, Dahlroth et al., 2006, Cornvik et al., 2006). The DNA fragmentation library is ligated into an expression vector encoding a C-terminal His_6 affinity tag and used to transform *E. coli*, plated on L-agar medium. The expression is sampled by colony lift onto L-agar plates containing IPTG, and *in situ* colony lysis (section **1.5.3**), is carried using a 'CoFi blot sandwich' where only soluble protein passes through the filter paper to the nitrocellulose membrane. The CoFi protocol was initially used to successfully identify soluble expression constructs for a number of higher eukaryotic proteins termed "hard to express" in *E. coli* (Cornvik et al., 2006). This technique was used to identify optimal expression constructs for the C-terminal cytoplasmic domain of the membrane protein rotavirus nsp4 (non-structural protein 4) from feline coronavirus, and enabled determination of the X-ray crystal structure (Manolaridis et al., 2009).

An adapted CoFi protocol had also been shown to be suitable for screening random mutagenesis libraries, created by polymerases with a mutator phenotype, for expression of membrane proteins in *E. coli* (Martinez Molina et al., 2008). A further adaptation of CoFi, termed Screening Colonies of ORFeome Pools (SCOOP) was used to screen in *E. coli* all of the open reading frames (ORFs) of the KSHV (Kaposi's sarcoma associated herpes virus), this identified a number of soluble expression constructs (Dahlroth et al., 2009). Yet another adaptation to CoFi termed 'Hot-CoFi', investigates the random high-throughput mutagenesis of a target protein as an approach to engineer more thermally and proteolytically stable recombinant proteins, specifically for drug development pipelines. Briefly, 'Hot-CoFi' involves testing the aggregation status of proteins in response to

increasing temperature, as indicated by a reduction of protein diffusion through a filter membrane. This method was found to improve the stability of a number of unnamed "biotechnological important proteins" (Asial et al., 2013).

1.5 Combinatorial Domain Hunting (CDH)

In order to help combat the many difficulties of producing multi-milligram quantities of high-quality protein from the heterologous host *E. coli*, Dr Renos Savva and his colleagues from Birkbeck College and UCL, London, developed the Combinatorial Domain Hunting (CDH) methodology (Reich et al., 2006). The first published study of CDH, was a 'proof-of-concept' study using the p85α regulatory subunit of class 1A phosphoinositide 3-kinase as a target. CDH identified a number of soluble, highly expressed constructs, corresponding to the known globular regions of the protein (Reich et al., 2006). Subsequently, CDH was used to generate a soluble, non-aggregating protein construct of MEK1 (MAPK/ERK kinase 1), known as MEK1 4F11 (Meier et al., 2012). The CDH methodology has been now successfully used for over 50 different target proteins, many of which have been for industrial clients (Littler, 2010). As this particular technique is used in this thesis, a comprehensive review of the method is provided below; and a schematic of the methodology can be found in **Figure 1.1**.

1.5.1 Amplification and fragmentation of the target gene

This section is redacted for reasons of confidentiality.

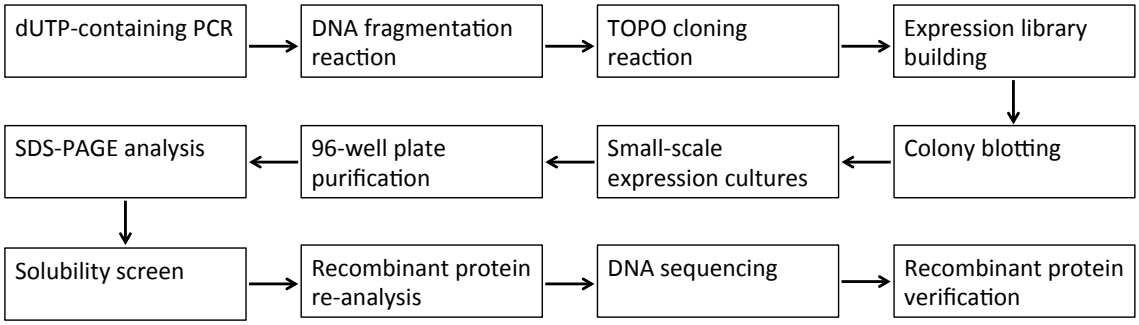


Figure 1.1 The CDH methodology pipeline.
A flow diagram showing the steps in the CDH methodology. Figure adapted from Reich et al., (2006) and Littler, (2010).

1.5.2 Cloning of DNA fragments

Details in this section have been redacted for reasons of confidentiality.

The DNA fragments are ligated into a specially designed TOPO-cloning expression vector series (pDXV4; proprietary technology of Domainex). The vector series provides three translation starts in three different reading frames, each with a C-terminal His_6 affinity tag and stop codons in three different reading frames. This allows capture of every generated fragment, with a probability of 1/18 of a fragment ligating correctly in frame (**Figure 1.3**). The His_6 affinity tag is used for expression screening in order to identify recombinant protein and to significantly reduce any passenger-protein solubilisation effects often seen with larger affinity tags, which are typically highly soluble. This can result in the artificial solubilisation and/or stabilisation of recombinant protein fragments that are unstable or unfolded. Removing the solubility tag from these fragments identifies the true properties of these fragments, thereby identifying 'false-positives' clones.

1.5.3 Expression screening: Colony blot

Typically around 20,000 individual transformed colonies are screened through the colony blot. An *E. coli* host strain is transformed with the DNA fragment library, with selection on L-agar plates containing antibiotic. A nitrocellulose membrane is placed on top of the plate, capturing a small amount of each transformed colony, and this membrane is placed down (colony side up) on top of an L-agar plate containing antibiotic and IPTG to induce protein expression. Subsequently, the

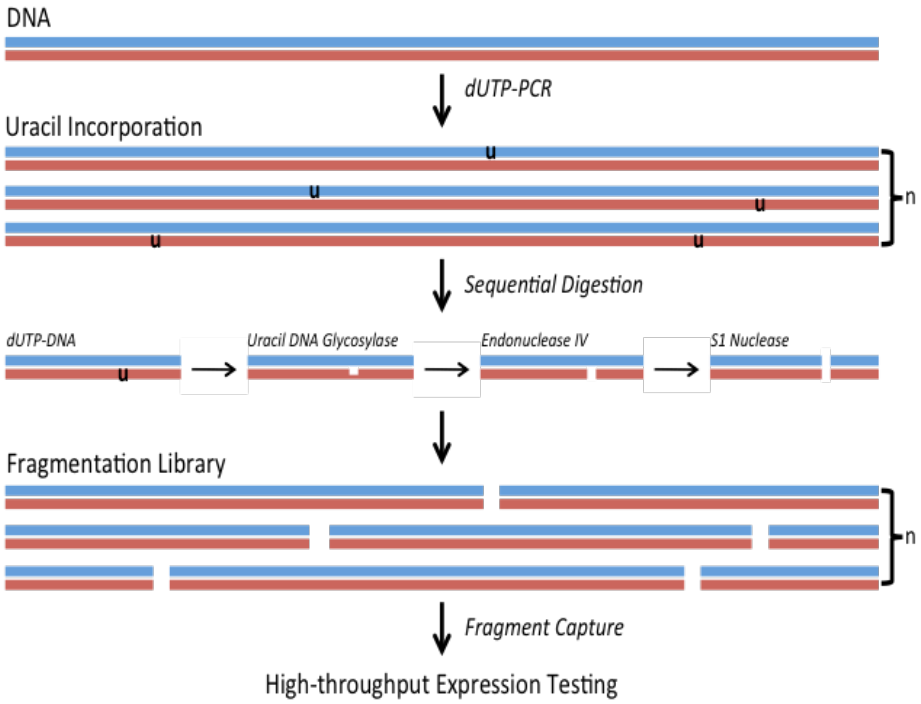


Figure 1.2 A schematic diagram illustrating the CDH methodology.
By including dUTP in PCR reactions, uracil is stochastically incorporated into a DNA amplicon (dUTP-PCR). The amplicon is then treated using a series of sequential enzymatic digests: UDG first creates an abasic site by removing incorporated uracil base; Endo IV then creates a nick in the backbone of the DNA on the 5' side of the abasic site; S1 nuclease cleaves the DNA backbone at the Endo IV nicks, and with extended incubation this also cleaves on the opposite strand of the nick, to create a blunt-end double-strand break at the uracil incorporation site. This process creates a 'random fragment' library, which can then be ligated into an adapted series of TOPO-cloning vectors suitable for protein expression. Figure adapted from Reich et al., (2006).

Figure 1.3 An overview of the pDXV4-TOPO vector series.

This Figure is redacted for reasons of confidentiality.

induced colonies are lysed *in situ* on the membrane. The recombinant protein expression is assessed by a colony blot using a monoclonal anti-His primary antibody, with colourmetric detection using an Alkaline-phosphatase conjugated secondary antibody. Colonies expressing high levels of recombinant protein are ranked and selected for the next stage of CDH.

1.5.4 Expression screening: Small scale cultures

Typically 750 colonies are picked and then used in small-scale, high-throughput, expression studies. Transformed *E. coli* are grown in auto-induction medium, enabling both high-density culture and spontaneous induction of protein expression without the addition of IPTG (Studier, 2005). The harvested cell pellets are lysed, and recombinant protein purified in a 96-well plate batch-format using a one-step IMAC purification. Eluted protein samples are resolved by SDS-PAGE to identify highly expressed recombinant proteins. The identified proteins are subjected to a solubility screen (by centrifugation) to pellet any insoluble material, and the soluble sample fractions are reanalysed by both SDS-PAGE and western blotting. These steps are used to identify highly expressed, soluble and proteolytically stable recombinant proteins, which are verified as containing the C-terminal His_6 affinity tag. Selected constructs are identified by DNA sequencing to determine the location of the DNA fragment with respect to the original full-length sequence. These preliminary constructs are termed 'initial-hits' and the fragment clones of interest can be taken forward for further characterization for structural and/or biochemical and biophysical analyses.

1.5.5 CDH^2

An extension of the CDH methodology, termed CDH^2, was developed and is used for determining the minimal interacting regions between two protein targets (Littler, 2010, Maclagan et al., 2011). For CDH^2, a competent *E. coli* strain is created, transformed with a construct expressing the first of the two target proteins, and a DNA fragment library is created for the second protein, using the standard CDH method (using different affinity tags). The resulting expression vectors are transformed into the previously created *E. coli* strain. A colony blot is used to identify colonies that overexpress both recombinant proteins, taking

advantage of two different affinity tags that are encoded by each expression vector (probed with antibodies against both tags). Subsequent affinity chromatography purification (for the tag of the fragment protein) and a western blot for the two affinity tags, confirms interaction of the expressed proteins. The technique is then reversed, in order to ascertain the minimal interaction regions / interfaces of the two proteins. A proof-of-concept study, used the previously characterised interaction between the molecular chaperone Hsp90β (Heat shock protein 90β), and the Hsp90β co-chaperone Cdc37. CDH[2] successfully identified the known minimal interacting regions for both proteins (Maclagan et al., 2011). Subsequently, this technique was used to investigate a number of further PPIs (Littler, 2010).

1.5.6 Assessment of the recombinant library

Subsequent to the identification of the 'initial hits', these require verification and assessment to ascertain whether the individual constructs are viable (Savva et al., 2007). The protein can be identified by Edman degradation (N-terminal sequencing) and by mass spectrometry; this verifies the recombinant protein is expressed from the clone sequence and identifies if there are any further modifications to the sequence (e.g. protein truncations). Recombinant protein can be treated with protease to determine the protease resistance, which can indicate the stability of folding. Globular proteins are typically protease resistant, whilst unstructured regions are typically protease sensitive; flexible regions can cause difficulties for the determination of the atomic resolution structure of the protein. If smaller protein fragment are identified as being protease resistant, these can be identified and the expression level of the smaller recombinant protein can be tested. Thermal denaturation of the recombinant protein can determine the 'folded-ness' and the thermal stability of the protein. Circular dichroism (CD), used for assessment of the secondary structure of proteins, can be used to identify folded proteins and the relative secondary structure content. Size exclusion chromatography (SEC), analytical ultracentrifugation (AUC) and/or dynamic light scattering (DLS), can be used to determine the molecular mass of the recombinant protein and the sample dispersity. Additionally, NMR can be used to determine the 'folded-ness' of the recombinant protein in solution. These analyses aim to identify

the most optimal clones to take forward for further investigation, such as drug discovery efforts and structural studies, but these are dependent on the target.

1.6 Protein Kinases

Protein kinases are traditionally difficult to express in the heterologous host *E. coli* and have been the subject to much study by domain hunting technologies to identify optimal expression constructs. In recent years, structures of a large number of protein kinases, from different kinase families, have been determined. The SGC alone have solved over 75 unique protein kinase structures, using their bioinformatics-led approach to expression construct design (section **1.3**).

Protein kinases are the key mediators of signal transduction pathways in eukaryotic cells, regulating a large number of essential cellular processes. Careful control is therefore required in order to maintain cellular stability, with alteration or aberrations of kinase activity in particular networks linked to cancer and a number of other diseases. Protein kinases are defined by their catalytic activity; the reversible transfer of the γ-phosphate from adenosine triphosphate (ATP) to the hydroxyl side chain of serine, threonine or tyrosine residues of a target protein (substrate). This can cause a change in the behaviour of the modified substrate protein, either through conformational alterations, changes to its enzyme activity or changes in interaction with its cognate macromolecular partner(s) (Endicott et al., 2012).

Sequencing and subsequent analysis of the human genome, identified 518 genes which express a protein kinase, compromising ~1.7% of the genes in the human genome; this is referred to as the 'kinome'. The kinase domain or the 'catalytic core' of protein kinases (~290 amino acids) is highly conserved both in amino acid sequence and in tertiary structure. The majority of these, based on amino acid sequence homology, can be classified into seven different groups (AGC, CAMK, CK1, CMGC, STE, TK and TKL). The remaining proteins kinases are classified as 'atypical' and whilst having low amino acid sequence homology to other protein kinases, these retain activity (Manning et al., 2002). N- and/or C-terminal extensions to the

kinase domain are however not highly conserved in amino acid sequence (Endicott et al., 2012).

The kinase domain is 'bi-lobal', and is comprised of a small N-terminal lobe connected by a linker, the so-called 'hinge-region', to a larger C-terminal lobe; both lobes are essential for catalytic activity. The N-terminal lobe is typically composed of a β-sheet and the α-helix (α-C helix). The C-terminal lobe is predominantly α-helical in nature and is more rigid in structure (refer to **Figure 7.9A**) (Knighton et al., 1991, Hanks and Quinn, 1991, Hanks and Hunter, 1995). Within the kinase domain, there are also a number of conserved features including the glycine-rich loop (GxGxxG) and α-C helix, found in the N-terminal lobe, and in the C-terminal lobe, the RD (Arg-Asp acid) motif and the activation segment with its conserved DFG (Asp-Phe-Gly) and APE (Ala-Pro-Glu) motifs. The 'hinge-region' connecting the two lobes is flexible in nature, and enables the movement of the two lobes relative to each other, between the so-called "open" and "closed" conformations, facilitating ATP binding and turnover. ATP binds to a deep cleft formed at the interface of the two lobes (ATP-binding pocket). The binding of ATP promotes closure of the kinase lobes, but the orientation of the activation segment is required for protein kinase activity; this often requires phosphorylation on a residue within the activation loop (although not all kinases require this phosphorylation event, such as Chk1). The interaction of the kinase domain with a protein substrate is localised predominantly in the C-terminal lobe (Endicott et al., 2012).

A number of human diseases, such as cancer, are linked with dysregulation of signalling pathways through aberrant protein kinase activity; therefore, inhibition of kinase activity presents itself as an ideal form of therapeutic intervention. Typically, small molecule inhibitors of protein kinases are designed to bind to the ATP binding pocket, thus preventing ATP binding and turnover, and resulting in decreased or no enzyme activity (Zhang et al., 2009b). However, the active site ("closed" conformation) of protein kinases is highly conserved, making the development of specific kinase inhibitors difficult, due to the large number of potential off-target effects, which may result in undesirable side effects. However,

determination of the tertiary structure of a particular kinase domain, has enabled a more rational design-led approach to developing inhibitors, tailoring particular molecular interactions that are highly-specific to a particular kinase or kinase-family (Grant, 2009). However, for a number of desirable protein kinase targets, determination of a tertiary structure remains elusive.

1.6.1 The 'pan-kinase' inhibitor Staurosporine

Staurosporine (STU), is a microbial alkaloid originally isolated from *Streptomyces staurosporeus* in 1977 (Omura et al., 1977). STU is a highly potent broad-spectrum protein kinase inhibitor, which binds with an affinity in the nanomolar range to the ATP-binding site of many different protein kinases; this renders it as highly unsuitable for use in disease treatment, due to the large number of off-target effects resulting in cellular toxicity (Ruegg and Burgess, 1989). STU binding occurs through an induced-fit model, resulting in a number of structural alterations to the ATP binding site; there is concomitant movement of the activation segment and rigidification of the normally flexible glycine-rich loop upon binding (Lamers et al., 1999).

This following part of this introduction focuses on the human protein Claspin, a target selected for the aforementioned CDH methodology.

1.7 Cell cycle

The eukaryotic cell cycle can be sub-divided into four distinct phases. The first, G_1-phase is a so-called 'gap' phase. The second is S-phase where the majority of DNA replication occurs resulting in the duplication of the DNA, forming sister chromatids. Next is a second gap phase, G_2-phase, which ensures that DNA replication has been completed and any residual DNA damage repaired. The final phase is M-phase, in which sister chromatids are separated, and distributed equally into two daughter cells, before cytokinesis separates them (Nurse, 1997).

1.7.1 DNA replication

Eukaryotic DNA replication is divided into three phases: initiation of replication (**Figure 1.4**), replication of DNA by the replisome (**Figure 1.5**) and replication

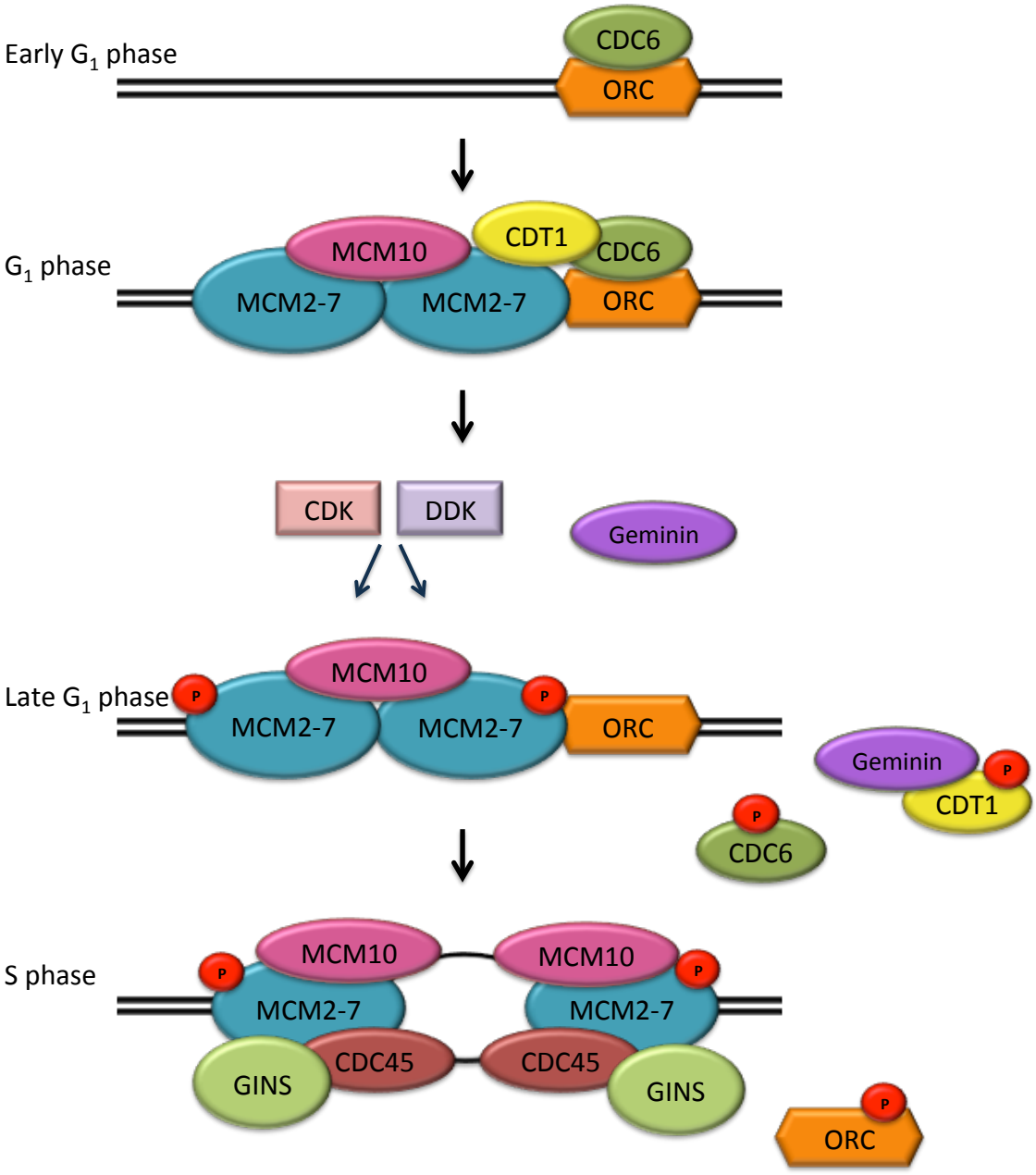


Figure 1.4 Eukaryotic DNA replication initiation.
A schematic for the initiation of eukaryotic DNA replication. DNA origins are bound by ORC, which recruits CDC6 during early G_1-phase. This complex recruits CDT1 and enables loading of the MCM2-7 helicase and MCM10 during G_1-phase. During late G_1-phase DDK and cyclinE-CDK2 phosphorylation removes CDC6 and CDT1; where Geminin sequesters CDT1. Cdc45 and GINS are recruited resulting in origin firing and the unwinding of the origin DNA during S-phase ready for replisome formation and DNA replication. 'P' denotes phosphorylation. Adapted from Leman and Noguchi, (2013).

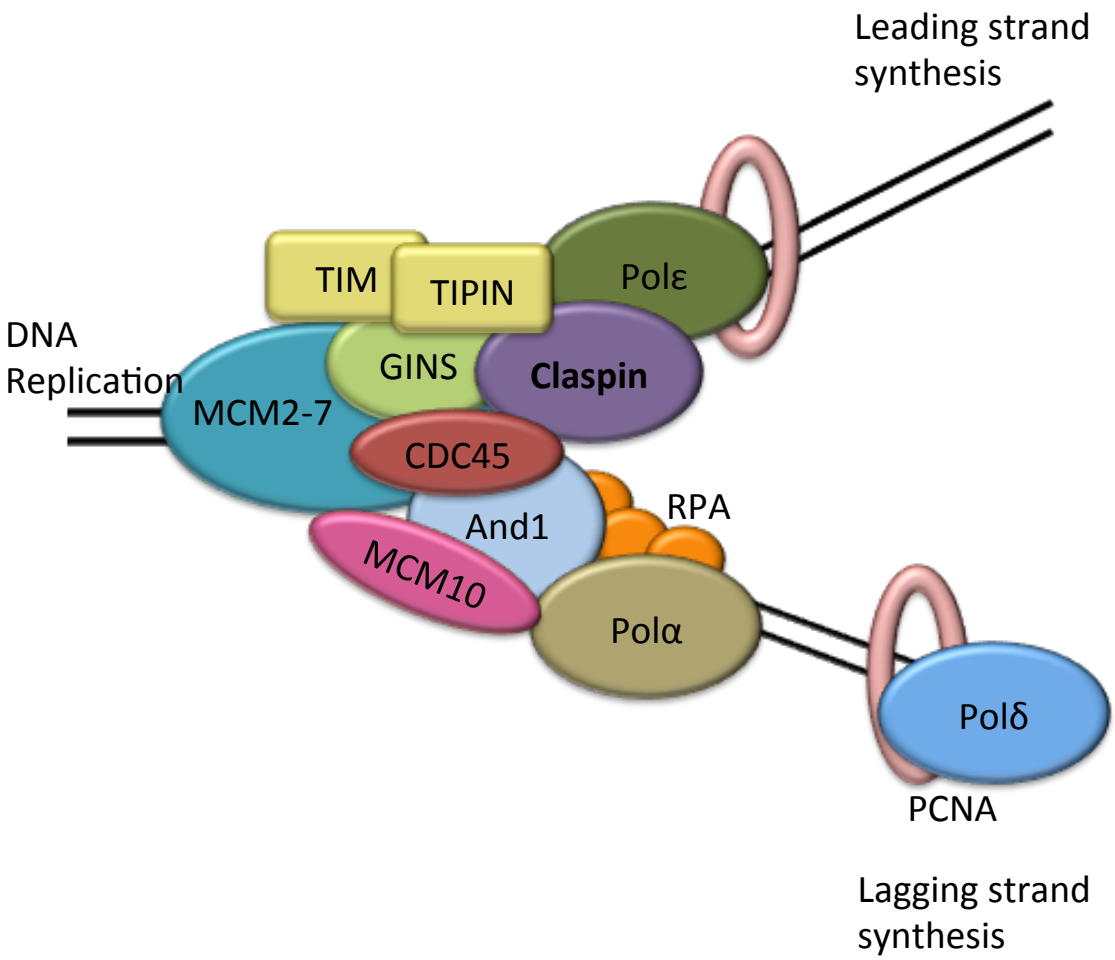


Figure 1.5 The eukaryotic replisome.
Replisome assembly required for DNA synthesis. DNA is unwound by the MCM2-7 helicase with Cdc45 and GINS (CMG complex) and the DNA polymerase α / RNA primase synthesises a short RNA primer to enable DNA replication. On the 'leading strand', DNA replication is catalysed by DNA polymerase ε, whilst on the 'lagging strand', replication requires DNA polymerase δ. RPA coats any exposed ssDNA after DNA unwinding and prior to DNA synthesis. The FPC (Claspin-Tim-Tipin-And1) is required for the stability of the replication fork and to link the MCM2-7 helicase with the DNA polymerase ε and α. Adapted from Errico and Costanzo, (2012).

termination, each of which are coordinated with stages of the cell cycle, as reviewed by Bell and Dutta, (2002), Aze et al., (2012), and Leman and Noguchi (2013).

DNA replication is initiated from 'origins', regions of DNA that are recognized by the hexameric ORC (Origin Recognition Complex), where origins are bound by ORC during late M- and early G_1-phase. During late G_1-phase there is step-wise recruitment of additional replication factor proteins to the ORC complex, initially the proteins CDC6 (cell division cycle, protein 6) and CDT1 (chromatin licensing and DNA replication factor 1) are recruited. This complex enables the recruitment and loading of the MCM2-7 (mini-chromosome maintenance 2-7) helicase and MCM10, to form the pre-replicative complex (pre-RC). Pre-RCs are distributed throughout genomic DNA, but remain inactive until the onset of S-phase, when a sub-set of pre-RCs are 'fired'. At the onset of S-phase, a large number of phosphorylation events are driven by two protein kinases, DDK (Dbf4/Drf1-dependent CDC7 kinase) and cyclin E-CDK2 (cyclin-dependent kinase 2). CDC6 and CDT1 are removed by phosphorylation, and Geminin sequesters CDT1 preventing replication re-initiation. As a result of these phosphorylations, both Cdc45 and GINS (go-ichi-ni-sans) are recruited to the MCM2-7 helicase, forming the CMG (Cdc45-MCM-GINS) complex, leading to the concomitant activation of helicase activity. This complex serves to unwind DNA at the origin, in a process termed 'origin firing', to form a 'replication-bubble'. Subsequently, deposition of DNA polymerases occurs, and in the presence of a number of additional proteins including the RFC (replication factor C) -clamp loader and the PCNA (proliferating cell nuclear antigen) clamp, an active replication fork is produced, which is able to start replicating DNA.

Upon DNA unwinding, DNA polymerase α, in association with a RNA primase, synthesises a short RNA primer. On the 'leading strand', DNA replication is continuous and is catalysed by DNA polymerase ε. On the 'lagging strand', replication is instead discontinuous, requiring multiple sites of priming (again by DNA polymerase α / RNA primase) and extension thought the catalytic activity of DNA polymerase δ, to form so-called Okazaki fragments. The DNA is processed at a

later point, to remove the RNA primers, and seal the nicks in the DNA backbone. The enzymatic action of the MCM2-7 helicase (for unwinding) and the DNA polymerases are highly coupled, in order to promote efficient replication, contained within the replisome. Progressive DNA replication requires a number of additional proteins that include Claspin (refer to section **1.10.1**), with the obligate Tim (Timeless) -Tipin (Timeless-interacting protein) complex with And1 (Acidic nucleoplasmic DNA-binding protein 1) (refer to section **1.10.5**). DNA replication terminates on the meeting of two replication bubbles on a length of DNA.

1.8 Repair of DNA damage and checkpoints

The integrity of genetic material in cells is continuously challenged by both endogenous and exogenous sources of damaging agents (Lindahl et al., 1993). In order to maintain faithful replication of DNA, and to monitor the completion of each cell cycle phase, several different 'DNA-damage checkpoints' have evolved; defined as G_1-S, Intra-S-phase, S-M, and spindle checkpoints, each correlating with a particular point in the cell cycle. These checkpoints respond to the presence of DNA lesions, single-stranded DNA (ssDNA) or DSB in DNA (**Figure 1.6**). During the process of DNA replication, there is additional protection in the form of the 'replication checkpoint', which is further described in section **1.8.1**. Activation of a checkpoint results in cell cycle arrest, providing time for the cell to repair the damage. If arrest of the cell cycle is prolonged, the fate of the cell can be directed down one of programmed cell death (apoptosis).

Each of these pathways and their associated proteins are highly conserved from yeasts to mammals, but the focus here will be on Metazoa [as reviewed in Nyberg et al., (2002), Sancar et al., (2004), Kastan and Bartek, (2004), Bartek and Lukas, (2007), Smith et al., (2010)]. Both the DNA-damage and replication checkpoint pathways are activated by a highly regulated signal-transduction cascade, acting from the site of DNA damage or replicative stress, and are formed of different proteins called 'sensors', 'transducers', 'mediators' and 'effectors'. Sensors directly detect the DNA damage and these communicate to, and result in the subsequent activation of transducers; which include the PIKK (phosphatidylinositol 3' kinase-related) family of protein kinases including ATM (ataxia-telangiectasia mutated)

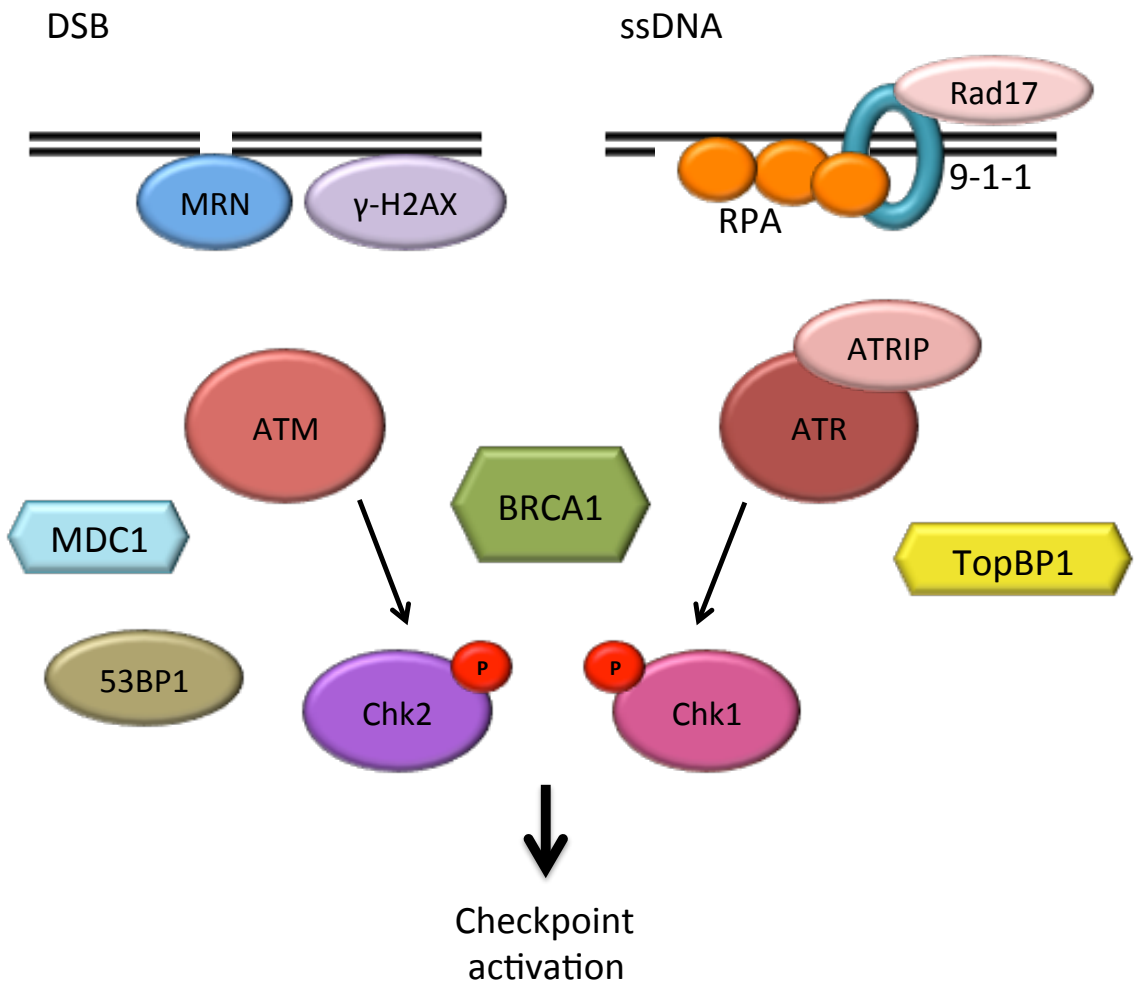


Figure 1.6 Activation of the DNA damage checkpoint.
An overview of DNA damage checkpoint signalling pathway in response to both DSB and ssDNA, resulting in the activation of Chk2 and Chk1 respectively. 'P' denotes phosphorylation. Adapted from Yoshiyama et al., (2013).

and ATR (ataxia-telangiectasia mutated and RAD3-related). ATM is predominantly activated at DSBs, whilst ATR is activated as a result of the detection of regions of exposed ssDNA, including those caused directly by DNA damage and/or by replication fork stalling. The transducer signal is carried forward through the activities of the downstream effector protein kinases, Chk1 and/or Chk2 (Checkpoint kinase 1 / 2). This action requires the so-called 'mediators', which physically connect the transducers and effectors together, often through complex sets of highly-regulated PPIs. Such mediators include 53BP1 (Tumour suppressor p53-binding protein 1), MDC1 (Mediator of DNA damage checkpoint protein 1) and BCRA1 (Breast cancer type 1 susceptibility protein), TopBP1 (topoisomerase-binding protein-1), and Claspin specifically for ATR mediated-checkpoint signalling during the activation of the replication checkpoint. Signalling from the downstream effector kinases results in arrest of the cell cycle.

1.8.1 The replication checkpoint

The replication checkpoint [reviewed in Sancar et al., (2004), Bartek and Lukas, (2007), Smith et al., (2010), Osborn et al., (2002)] is specifically activated by stalling of replication forks; caused by replicative stress, DNA lesions or by proteins bound to DNA in the path of a progressing replication fork. Checkpoint activation serves to stabilise the replication fork, preventing its collapse, suppress further origin firing, and to enable DNA repair or time for the alleviation of replicative stress conditions. A collapsed replication fork would activate the DNA damage checkpoint, leading to genome instability.

Activation of the replication checkpoint is dependant on the ATR-Chk1 signalling cascade (**Figure 1.7**). During normal DNA replication RPA (Replication Protein A), a ssDNA-binding protein, coats any exposed ssDNA, thus protecting it. Activation of ATR-Chk1 signalling is thought to occur when excessive stretches of RPA-coated ssDNA are formed, generated by uncoupling of the MCM2-7 helicase and DNA polymerases by replicative stress. The ATR kinase is localised to a stalled replication fork through its obligate partner ATRIP, which can bind directly to RPA-coated ssDNA tracts. However, this localisation is not sufficient for full activation of ATR kinase activity — this requires the presence of the Rad9-Rad1-

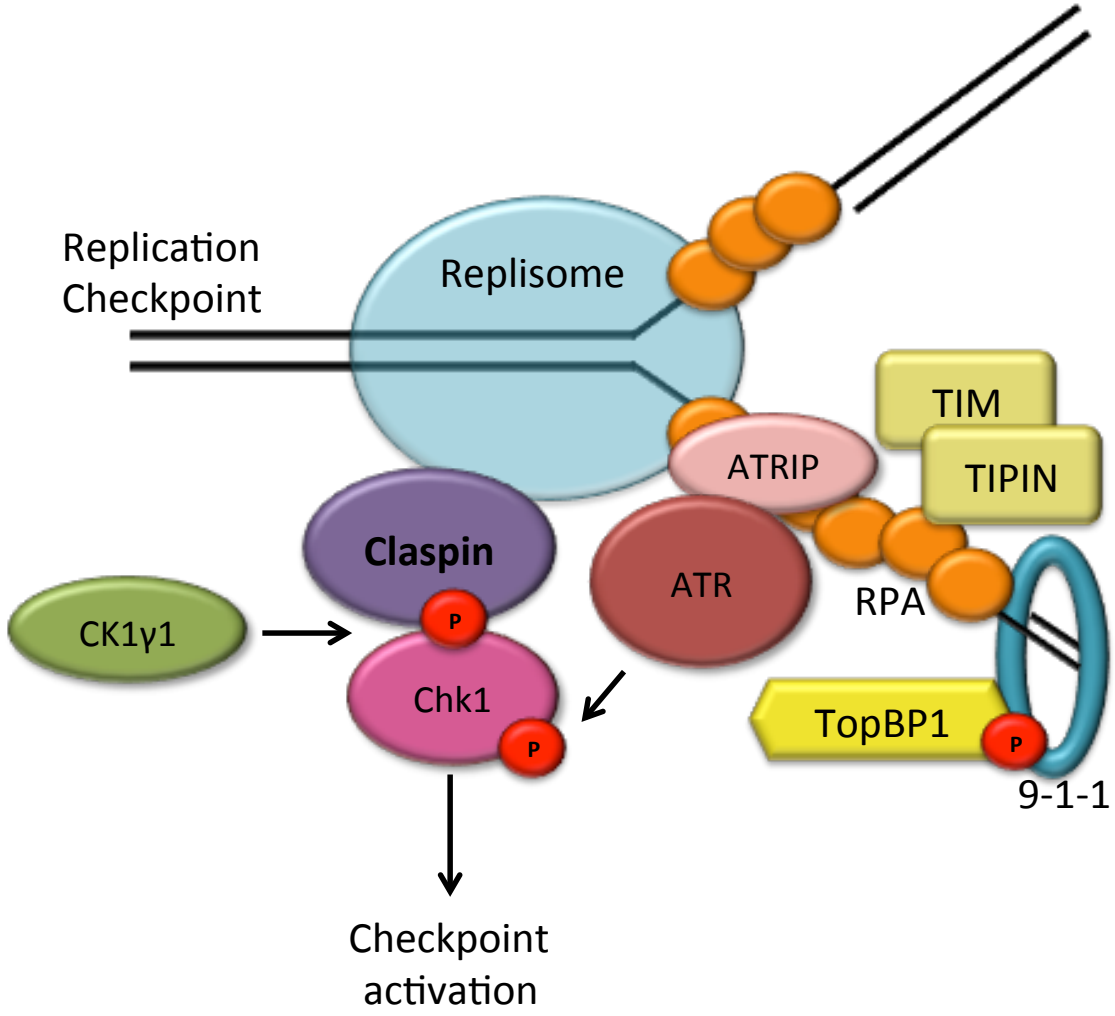


Figure 1.7 Activation of the replication checkpoint.
Upon replication stress, excessive stretches of RPA-coated ssDNA are formed. ATR is localised to a stalled replication fork through ATRIP, which binds directly to RPA. Activation of ATR kinase activity requires localisation of the 9-1-1 complex and TopBP1. The Tim-Tipin complex is required for localisation of Claspin at the stalled replication fork and to enhance the activation of Chk1. Claspin is phosphorylated by CK1γ1, enabling Chk1 interaction with Claspin. Subsequently Chk1 is activated by phosphorylation by ATR. 'P' denotes phosphorylation. Adapted from Smith et al., (2010).

Hus1 (9-1-1) complex, loaded onto DNA by the Rad17-RFC clamp-loader, in an event independent of ATR function; and TopBP1, which is directly recruited by the phosphorylated C-terminus of Rad9 (9-1-1 complex). This complicated set of requirements is thought to be necessary, in order to prevent unnecessary signalling from the low levels of RPA-coated ssDNA found at normal (non-stalled) DNA replication forks.

Downstream signalling, after ATR activation, is initiated through the phosphorylation of the effector protein Chk1 (section **1.9**), which requires the mediator protein Claspin (Zhao and Piwnica-Worms, 2001, Kumagai and Dunphy, 2000, Kumagai et al., 2004). The function(s) of Claspin during the replication checkpoint is described in section **1.10.3,** while the interaction between Claspin and Chk1 is described in section **1.10.4**. Once activated, Chk1 phosphorylates the phosphatase Cdc25A (thereby targeting it for degradation) (Sanchez et al., 1997) and the protein kinase Wee1 (resulting in its activation) (O'Connell et al., 1997); thereby the activation of Chk1 inhibits the CDK (cyclin dependant kinase) required for entry into mitosis, and results in a slowing of S-phase progression.

1.9 Chk1 protein kinase

The protein kinase Chk1 was originally identified in *Schizosaccharomyces pombe* (*S. pombe*; fission yeast) as being essential for cell cycle arrest in response to DNA damage (Walworth et al., 1993). Subsequent to this, homologues of Chk1 were identified in other species (Kumagai et al., 1998). The roles of Chk1 and Chk2 have swapped over from yeasts to metazoans, where Chk1 in *S. pombe* (Cds1) is the sequence homologue of metazoan Chk2 but functional homologue of Chk1 (Rhind and Russell, 2000), the same is also true for Chk1 in *S. cerevisiae* (Rad53); for clarity these are labelled as: $spCds1^{Chk1}$ and $scRad53^{Chk1}$ throughout.

The regulation of Chk1 kinase activity *in vivo* is complex, with interplay between auto-inhibition, phosphorylated and de-phosphorylated states, and protein interactions is not yet fully understood [reviewed in Tapia-Alveal et al., (2009)]. However, its function is critical in ensuring correct cell cycle regulation.

Chk1 consists of a N-terminal protein kinase domain and a C-terminal regulatory domain that modulates kinase activity. The structure of the kinase domain of human Chk1-KD has been determined (PDB 1IA8), which found the activation segment of Chk1 to be in the 'active' conformation, despite not being phosphorylated (Chen et al., 2000). *In vivo*, Chk1 is held in an inactive form by the regulatory domain, but *in vivo*, the loss or deletion of this domain does not result in an active protein kinase (Ng et al., 2004). However, stringent washing of immunoprecipitated Chk1 showed increased kinase activity (Walker et al., 2009). Interesting, recombinant full-length Chk1 or the kinase domain are both functionally active *in vitro* (Jackson et al., 2000, Chen et al., 2000). The C-terminus of Chk1 has been identified as auto-inhibitory to kinase activity (Katsuragi and Sagata, 2004, Kosoy and O'Connell 2008), where two small highly conserved regions (aa 376-382 and 446-455) within this domain are thought to interact with specific locations at both the N- and C-terminal lobes of the kinase domain (Caparelli and O'Connell, 2013). This is thought to hold the kinase domain in an inactive state, and thus prevent either ATP or substrate binding. Mutation / deletion of this region and mutation of specific residues within the conserved sequences was found to both positively and negatively affect the kinase activity of Chk1 (Kosoy and O'Connell 2008, Caparelli and O'Connell, 2013). The regulatory domain has recently been predicted to be a KA1 (kinase associated 1) domain (Caparelli and O'Connell, 2013); a known auto-inhibitory domain found in other protein kinases.

ATR is known to phosphorylate Chk1 at multiple SQ sites, in a region between the kinase and regulatory domain, with phosphorylation of Ser345 (by ATR) is critical for Chk1 activation (Lopez-Girona et al., 2001, Capasso et al., 2002); this is thought to remove the inhibitory interaction of the regulatory domain. However, for full kinase activity, phosphorylation of both Ser317 and Ser345 is necessary for Chk1 autophosphorylation on Ser296 (Zhao and Piwnica-Worms, 2001, Gatei et al., 2003, Katsuragi and Sagata, 2004, Okita et al., 2012). The autophosphorylation event is enhanced by the interaction with Claspin (Kumagai et al., 2004).

1.10 Claspin

The protein Claspin was originally identified through a replication stress-induced Chk1 interaction screen, using *Xenopus* egg extracts (Kumagai and Dunphy, 2000). Human Claspin is a large, highly acidic protein, 1339 amino acids in length (151 kDa), which is highly conserved amongst Metazoans (Kumagai and Dunphy, 2000), but apparently absent in prokaryotes. Homologues of Claspin, called Mrc1 (Mediator of replication checkpoint 1) have been also been identified in the yeasts, although the overall amino acid sequence conservation is relatively poor (Alcasabas et al., 2001, Tanaka and Russell, 2001). Mrc1 was itself identified in *S. cerevisiae* (scMrc1) screens for mutants under chronic replication stress conditions (Alcasabas et al., 2001), and in *S. pombe* (spMrc1) by Hydroxyurea (HU; DNA replication inhibitor) sensitivity with replication stress (Tanaka and Russell, 2001).

Claspin contains no known amino acid sequence motifs or regions of sequence homology to any previously determined structures (Kumagai and Dunphy, 2000). It is predominantly acidic in nature (pI = 4.5), but it does contain four conserved basic patches (BP), termed BP1 (aa 265-331), BP2 (aa 470-600), BP3 (aa 721-783) and BP4 (aa 1157-1285) (Lee et al., 2005). There are also a number of SQ/TQ sites located throughout the protein, which are typically substrates of ATR and ATM (Kumagai and Dunphy, 2000). The only structural information available to date for the human protein is a low-resolution Electron Microscopy (EM) image, using rotary shadowing, which indicates that Claspin has a ring-shaped structure when bound to branched DNA (**Figure 1.8**) (Sar et al., 2004). The N-terminal part of the protein interacts with DNA (Sar et al., 2004, Lee et al., 2005, Serçin and Kemp, 2011, Uno and Masai, 2011, Yilmaz et al., 2011), which is described in section **1.10.2**, whilst the C-terminus contains a Chk1 binding site that is required for the activation of Chk1 (Kumagai and Dunphy, 2003, Jeong et al., 2003), which is described in section **1.10.4**.

1.10.1 Claspin at the replication fork

Claspin is known to be a cell cycle regulated protein, with levels of the protein peaking during S-phase (DNA replication) (Chini and Chen, 2003), where Claspin

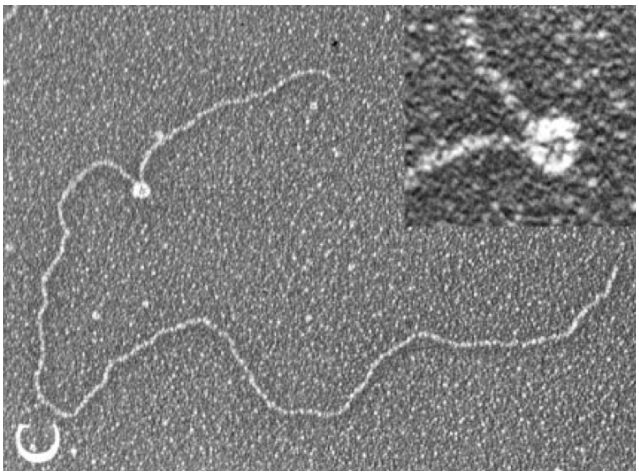


Figure 1.8 Electron micrograph of Claspin bound to DNA.
Rotary shadowing EM, which indicates that Claspin has a ring-shaped structure when bound to branched DNA. Adapted from Sar et al., (2004).

localises to the nuclei (Chini and Chen, 2003, Sierant et al., 2010). Claspin is thought to be loaded on to chromatin during replication initiation, which is dependent on both bound Cdc45 and pre-RC formation, but not on RPA; potentially indicating that Claspin first interacts with chromatin during unwinding of DNA origins (Lee et al., 2003). The interaction between Claspin and Cdc45 is mediated by the N-terminal region of Claspin and the C-terminal region of Cdc45, and may enable the observed recruitment of Claspin to the pre-RC (Serçin and Kemp, 2011, Broderick et al., 2013). The maximum level of Claspin-Cdc45 complex is found at S-phase, but in response to ultraviolet (UV) damage the interaction between the two proteins is reduced (Broderick et al., 2013).

Claspin has been shown to be required for 'normal' rates of replication fork progression during S-phase. It has also been found to be required for both replication fork stability and density during unperturbed replication (Lee et al., 2003, Lin et al., 2004, Petermann et al., 2008, Scorah and McGowan, 2009, Yoshimura et al., 2011). Depletion of Claspin from *Xenopus* egg extracts resulted in the slowing of DNA replication (Lee et al., 2003), and this was confirmed in human cells by DNA-fibre analysis and DNA combing (Petermann et al., 2008, Scorah and McGowen, 2009, Yoshimura et al., 2011). Furthermore, knockdown of Claspin resulted in increased spontaneous cell death events (Yoshimura et al., 2011). For human cells, a greater effect on DNA replication was seen when both Claspin and Chk1 were co-depleted (Petermann et al., 2008). In contrast, overexpression of Claspin in human cells was reported to result in increased cellular proliferation (Lin et al., 2004).

Claspin is described as an 'adapter' protein, and a number of potential interacting proteins have been identified for Claspin from interaction screens including a number of replisome-critical proteins such as MCM2-7, DDK (Cdc7), Cdc45, and the DNA polymerases (Chini and Chen, 2003, Lee et al., 2005, Brondello et al., 2007, Kim et al., 2008, Gold and Dunphy, 2010, Nakaya et al., 2010, Uno and Masai, 2011, Serçin and Kemp, 2011, Rainey et al., 2013, Broderick et al., 2013). Additionally, Claspin forms a complex with Tim-Tipin and And1 (Gotter et al., 2007), which is described in **1.10.5**. However, whilst these interactors have been

identified the vast majority of interactions have not been mapped to Claspin, and only a few interactions have been reconstituted *in vitro* (**Figure 1.9**).

In Metazoa, Claspin has been identified as a direct phosphorylation target of DDK (Cdc7 kinase) in *Xenopus* egg extracts (Gold and Dunphy, 2010), human cells (Kim et al., 2008), and in an *in vitro* kinase assay (Rainey et al., 2013). Cdc7 has also been found to phosphorylate Claspin during unperturbed DNA replication and during replicative stress (Kim et al., 2008). Interaction with the MCM2-7 helicase appears to be dependent on phosphorylation of Claspin by DDK (Rainey et al., 2013). In *Xenopus*, DDK (through its Drf1 subunit) has been found to interact directly with amino acids 856-867 of Claspin, where Asp861 and Gln866 are essential for interaction (Gold and Dunphy, 2010). Interestingly, the same region of Claspin (aa 856-867) is also required for the phosphorylation-dependent interaction with Chk1 (Kumagai and Dunphy, 2003, Jeong et al., 2003), as described in section **1.10.4**. Mutant Claspin, that couldn't interact with DDK, still was able to activate the replication checkpoint, but DNA replication was slower (Gold and Dunphy, 2010). Furthermore, the loss or inhibition of Cdc7 is found to reduce Claspin protein levels and also results in impaired Chk1 activation although there was no effect on Chk1 activity once the replication checkpoint had been activated (Kim et al., 2008, Rainey et al., 2013).

The yeast homologue, Mrc1 is also regulated by the cell cycle, again with an observed increase in protein levels during S-phase (Tanaka and Russell, 2001, Osborn and Elledge, 2003). Mrc1 has been identified as binding to early firing replication origins, where DNA replication initiation is required for the association of Mrc1 with chromatin. This is loaded following similar steps as identified for Claspin-chromatin association in Metazoa (Osborn and Elledge, 2003, Gambus et al., 2006, Hayano et al., 2011). Mrc1 is not an essential protein, but as in Metazoa, scMrc1 has been shown to be required for 'normal' rates of replication fork progression during unperturbed replication (Alcasabas et al., 2001, Osborn and Elledge, 2003, Szyjka et al., 2005, Tourriere et al., 2005, Hodgson et al., 2007). Mrc1 travels with the replication fork and makes direct associations with the replication fork itself, and replication fork-associated proteins including; Cdc45, DNA

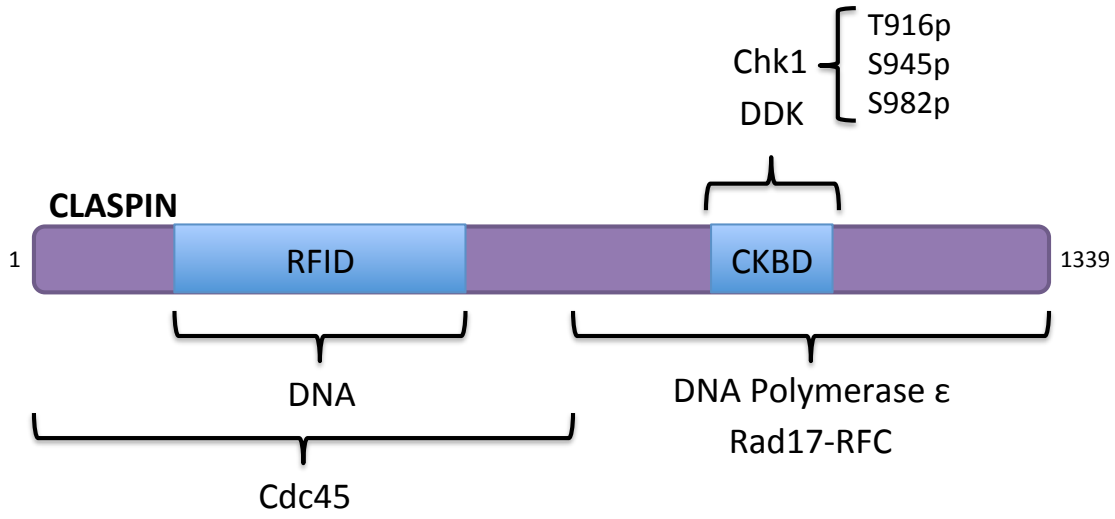


Figure 1.9 Interactions of human Claspin.
A summary of the identified functional sequence regions (blue), and protein interactions that have been located on human Claspin. Phosphorylated sites on Claspin, which are required for interaction with Chk1 are listed. RFID=Replication Fork Interacting Domain, CKBD=Chk1 Binding Domain. Adapted from Serçin and Kemp, (2011).

Polymerase ε and the MCM2-7 helicase (Katou et al., 2003, Osborn and Elledge, 2003, Calzada et al., 2005 Gambus et al., 2006, Komata et al., 2009, Lou et al., 2008, Tsai et al., 2015). The central region of scMrc1 is required for interaction with the MCM2-7 helicase (Komata et al., 2009), whilst two regions (at the N and C-terminus) of scMrc1 are required for an interaction with the scPol2 subunit of DNA polymerase ε during unperturbed replication (Lou et al., 2008). The interaction of scMrc1 with scPol2 is changed during replication stress, by phosphorylation of the N-terminus of scMrc1, resulting in loss of the N-terminal interaction (Lou et al., 2008). Furthermore, this interaction with scPol2 appears to have a role in regulating cellular replicative senescence from telomere shortening (Deshpande et al., 2011). Additionally, as with Claspin, Mrc1 makes interactions with the yeast homologues of Tim and Tipin (Katou et al., 2003, Hodgson et al., 2007, Tourriere et al., 2005, Calzada et al., 2005, Tanaka et al., 2010), described further in section **1.10.5**.

As with Metazoan Claspin, a number of post-translational modifications (PTM) have been identified on Mrc1. scMrc1 and spMrc1 are regulated by $Hsk1^{Cdc7}$ phosphorylation, the orthologue of Cdc7 kinase, and this can be reconstituted *in vitro* (Shimmoto et al., 2009, Matsumoto et al., 2010). Furthermore, scMrc1 interacts with $Hsk1^{Cdc7}$, and the interaction has been identified as being mediated by the central segment of scMrc1 (Shimmoto et al., 2009).

1.10.2 Claspin interacts with DNA

Claspin has been identified as having DNA interacting functionality. A region within the N-terminus of Claspin (aa 265-605) termed the Replication Fork Interaction Domain (RFID), has been shown to interact with chromatin in *Xenopus* egg extracts (Lee et al., 2005); this region is flanked by the aforementioned basic patches BP1 and BP2. Mutation of selected conserved residues or deletion of the entire BP1 and BP2 affected chromatin binding, but removal of the amino acids between BP1 and BP2 (aa 376-425) was found to have no effect on chromatin association (Lee et al., 2005). In addition, the N-terminal region of Claspin has been shown to directly interact with DNA *in vitro*, using both gel-shift (EMSA) and pull-

down assays. A region termed the DNA Binding Domain (DBD; aa 149-340) was identified as being sufficient for the DNA binding functionality (Sar et al., 2004).

spMrc1 was identified as interacting with chromatin, where phosphorylation of Ser604 was found to enhance chromatin association (Zhao et al., 2003). Similarly, spMrc1 has been shown to have DNA-binding activity, which has been mapped to a 150 aa region at the N-terminus of the protein (aa 160-317). It was found that deletion of this region, lead to HU hypersensitivity, and an associated impairment in the activation of the replication checkpoint in response to DNA damage (Zhao and Russell, 2004). Furthermore, the DNA binding activity of spMrc1 was diminished with mutation of Lys235 and Lys236 (Zhao and Russell, 2004, Tanaka et al., 2010).

Both Claspin and spMrc1*, in vitro,* have a consistent preference for double-stranded DNA (dsDNA) over ssDNA, and an overall preference for branched DNA structures including overhangs, Y-shaped, and fork-like structures, whilst having no sequence specificity (Sar et al., 2004, Zhao and Russell, 2004, Tanaka et al., 2010, Serçin and Kemp, 2011, Uno and Masai, 2011 and Yilmaz et al., 2011). It has also been reported that there is a preference for binding of Claspin to DNA containing bulky adducts (Yilmaz et al., 2011). Sequence conservation analysis of Claspin and spMrc1 identified a potential HTH (Helix-turn-helix) motif (human Claspin; aa 279-313, spMrc1; aa 204-244) within the identified DBD of both Claspin and spMrc1 (Zhao and Russell, 2004); these motifs are commonly associated with DNA binding functionality (Brennan and Matthews, 1989).

1.10.3 Claspin and the DNA replication checkpoint

As well as being required during DNA replication, Claspin has been shown to be required for activation of the DNA replication checkpoint, through mediating the activation of Chk1 in response to replicative stress; this interaction is described in section **1.10.4**. In *Xenopus* egg extract, the immuno-depletion of Claspin prevented Chk1 activation and the subsequent cell cycle arrest in response to DNA damaging agents (Kumagai and Dunphy, 2000, Kumagai and Dunphy, 2003). This was subsequently shown in human cells and human cell free systems (Chini and Chen,

2003, Lin et al., 2004, Clarke and Clarke, 2005, Chini and Chen, 2006, Liu et al, 2006b), and has been reconstituted *in vitro* using purified proteins (Lindsey-Boltz et al., 2009, Lindsey-Boltz and Sancar, 2011). The depletion of Claspin induced DNA damage in human cells (Liu et al., 2006b). During DNA replication stress, the presence of Claspin on chromatin was found increased (Lee et al., 2003, Lee et al., 2005), and this localisation, but not retention on chromatin, was required for the efficient activation of Chk1 (Lee et al., 2005). However, Claspin maintained a pan-nuclear localisation during replication stress, as was identified during DNA replication (Liu et al., 2006b). Claspin recruitment to stalled replication forks has been identified to occur via an interaction with Tipin, as part of the obligate Tim-Tipin complex (Errico et al., 2007, Yoshizawa-Sugata and Masai, 2007, Kemp et al., 2010), which is described in section **1.10.5**. Furthermore, Claspin interaction with Rad9 (9-1-1 complex) (Sierant et al., 2010, Liu et al., 2012b) or Rad17 (-RFC clamp loader complex) (Wang et al., 2006), have also been implicated in the localisation of Claspin during replication stress. Additionally, phosphorylation by Cdc7 has been shown to be required for the chromatin association of Claspin during replication stress (Kim et al., 2008).

In response to replicative stress, Claspin becomes hyper-phosphorylated (Chini and Chen, 2003), in particular a region at the C-terminus of Claspin termed the Chk1 binding domain (CKBD), which enables the transient association of Chk1 (Kumagai and Dunphy, 2003, Lee et al., 2005, Jeong et al., 2003, Clarke and Clarke, 2005, Chini and Chen, 2006, Lindsey-Boltz et al., 2009); this interaction is required for the efficient phosphorylation of Chk1 by ATR, and is described in section **1.10.4**. Phosphorylation of Claspin is also thought to regulate additional PPIs at stalled replication forks, and to enable efficient fork restart when the checkpoint is switched off (Rainey et al., 2013). A number of PTMs have been mapped on Claspin, including additional phosphorylation and ubiquitination sites. In *Xenopus*, Claspin phosphorylation has been identified on Thr817 and Ser819 (S/TQ motifs) in response to DSBs (Yoo et al., 2006). Ubiquitination by BRCA1 E3 ligase has been identified on human Claspin residues Lys60, Lys89, Lsy96 and Lys105 in response replication stress (Sato et al., 2012). These modifications may provide some

discrimination in the repair of a subset of genotoxic lesions, or for the activation of specific alternative DNA repair pathways (Yoo et al., 2006, Sato et al., 2012).

As within Metazoa, Mrc1 is required for activation of the DNA replication checkpoint. In *S. pombe*, spMrc1 was indicated in sensing DNA replication stress and for the activation of the replication checkpoint (Osborn and Elledge, 2003). Mutation of scMec1^{ATR} S/TQ phosphorylation sites in scMrc1 resulted in a checkpoint deficient mutation with no replication defect, indicating separate DNA replication and replication checkpoint functions of the protein (Osborn and Elledge, 2003). Phosphorylation of Mrc1 is required for the activation of the effector kinases scRad53^{Chk1} / spCds1^{Chk1} (Zhao et al., 2003, Osborn and Elledge, 2003). Furthermore, the central region of scMrc1 has been identified as being required for the mediator functionality of scMrc1 and for interacting with the replisome during DNA replication stress (Naylor et al., 2009, Shimmota et al., 2009). As in Metazoa, the yeast homologues of Tim-Tipin, spSwi1^{Tim} and spSwi3^{Tipin}, have also been shown to be required for the chromatin association of spMrc1 during replication stress (Shimmota et al., 2009).

Additionally, PTMs have been identified on Mrc1 in response to replication stress, including phosphorylation by spRad3^{ATR}, scMec1^{ATR}, scRad53^{Chk1}, Hsk1^{Cdc7} and spTel2^{Clk-2}; these are thought to regulate PPIs, the activation of the replication checkpoint and the stability of the stalled replication fork (Zhao et al., 2003, Shikata et al., 2007, Lou et al., 2008, Naylor et al., 2009, Shimmoto et al., 2009, Matsumoto et al., 2010). Furthermore, scMrc1 appears to also be regulated by extracellular stimuli from osmostress, through phosphorylation by the stress-activated protein kinase Hog1 (SAPK), in an action independent of scMec1^{ATR} and scRad53^{Chk1}, and results in delayed origin firing and slowing of DNA replication (Duch et al., 2012).

The adaptation from the replication stress checkpoint requires the release of Claspin from chromatin and the degradation of Claspin, as described in section **1.10.7**; this is also true for Mrc1 in yeasts.

1.10.4 Claspin-Chk1 interaction

The effector kinase Chk1 acts downstream from ATR, through an interaction mediated by Claspin (Kumagai and Dunphy, 2000). Previously it has been shown that Chk1 binds to Claspin, in a phosphorylation dependent manner, to a sequence region termed the CKBD (aa 847-903) (Kumagai and Dunphy, 2003, Lee et al., 2005, Jeong et al., 2003, Clarke and Clarke, 2005, Chini and Chen, 2006, Lindsey-Boltz et al., 2009). The CKBD was identified from experiments in *Xenopus* egg extracts and was shown to be both necessary and sufficient for Chk1 binding and its activation by ATR (Kumagai and Dunphy, 2003). However, for recovery from the replication stress checkpoint a somewhat larger 129 amino acid region termed the Chk1 activation domain (CKAD; aa 776-905) was required (Lee et al., 2005). The basic patch BP2 has also been shown to enhance the activation of Chk1 in *Xenopus*, but through an unknown mechanism (Kumagai and Dunphy, 2003, Lee et al., 2005). The CKBD is composed of a highly conserved region of amino acids, containing a 10 aa repeat termed a Chk1 binding motif (CKB), conforming to the motif: ExxxLC(S/T)GxF (where x is any amino acid) (Kumagai and Dunphy, 2003). The CKB motif is repeated twice in the *Xenopus* protein, and is located between residues 858-867 (xCKB motif 1) and 889-898 (xCKB motif 2), whilst in the human protein, there are three CKB motifs, located between residues 910-919 (CKB motif 1), 939-948 (CKB motif 2) and 976-985 (CKB motif 3) (Kumagai and Dunphy, 2003). This motif is found in other Claspin homologues, but interestingly it is not found in the yeast proteins (Kumagai and Dunphy, 2003).

Phosphorylation of the CKBD is essential for the interaction of Claspin with Chk1 (Kumagai and Dunphy, 2003, Chini and Chen, 2003, Clarke and Clarke, 2005, Chini and Chen, 2006, Lindsey-Boltz et al., 2009); where phosphorylation has been identified for *Xenopus* Claspin at Ser864 and Ser895 for xCKB motif 1 and 2 respectively, and for human Claspin at Thr916, Ser945 and Ser982 for CKB motif 1, 2 and 3 respectively (Kumagai and Dunphy, 2003, Clarke and Clarke, 2005). This was subsequently shown in human cells and human cell free systems (Clarke and Clarke, 2005, Chini and Chen, 2006) and was reconstituted *in vitro* (Lindsey-Boltz et al., 2009). In *Xenopus* egg extracts, mutation of the phosphorylation sites to either non-phosphorylatable or phospho-mimetic residues prevented interaction

with Chk1 and its subsequent activation (Kumagai and Dunphy, 2003, Jeong et al., 2003), this was also found in a human cell-free system (Chini and Chen, 2006, Lindsey-Boltz et al., 2009). In the human cell-free system, it was shown a single peptide corresponding to a phosphorylated composite CKB motif was able to directly interact with Chk1, but could not activate it (Clarke and Clarke, 2005). For Chk1 activation and autophosphorylation, it was found that two phosphorylation sites where required, specifically at Thr916 and Ser945, in CKB motif 1 and 2 respectively (Clarke and Clarke, 2005). Whilst the third CKB motif also has the propensity to be phosphorylated at Ser982, this region of Claspin has the lowest conservation at the amino acid sequence level and its phosphorylation does not appear essential for activation of Chk1 (Clarke and Clarke, 2005). Recently, the Drosophila protein kinase Gish was identified from an RNA interference screen from cell extracts that had reduced levels of Chk1 phosphorylation – this was verified by the specific inhibition of Gish kinase. The human equivalent of Gish kinase is CK1γ1 kinase, and this was shown to phosphorylate Claspin on the CBK motifs; however this does not rule out the possibility that other kinases may also be responsible (Meng et al., 2011). CK1γ1 kinase has a strong preference for acidic residues around the phosphorylation site, as found in the Claspin CKB motifs (Meng et al., 2011).

Chk1 can interact with the phosphorylated CKB motifs through its N-terminal kinase domain, whilst the C-terminal region is thought to enhance the interaction, but does not seem to make a direct interaction itself (Jeong et al., 2003). The X-ray crystal structure of the Chk1 kinase domain (PDB: 1IA8) (refer to **Figure 7.9A**) has a positively charged pocket formed from four highly conserved residues Lys54, Arg129, Thr153 and Arg162, close to the catalytic centre of the kinase, which is bound to a sulphate ion (Chen et al., 2000). Sulphate ions are often found bound to phosphate binding sites in crystal structures, as they are iso-structural with phosphate ions, and ammonium sulphate is a common component of mother liquors used in crystallisation. In *Xenopus*, individual mutation of each of the four residues of Chk1 prevented its stable association with phosphorylated Claspin, and there was a concomitant reduction in the level of phosphorylated, active Chk1 and associated downstream phosphorylation events (Jeong et al., 2003). The bound

sulphate ion is found to orientate the side chain of Arg162 (Chen et al., 2000) and therefore binding of Chk1 to the phosphorylated CKBD may result in slight conformational change to Chk1 (Jeong et al., 2003).

The interaction of Chk1 with Claspin brings Chk1 into close proximity to ATR for phosphorylation, followed by Chk1 autophosphorylation for full activation (Zhao and Piwnica-Worms, 2001, Kumagai et al., 2004) (refer to section **1.9**). A number of proteins have been identified as enhancing Chk1 activation including; TopBP1 (Liu et al., 2006b), Rad9 (Sierant et al., 2010), Rad17 (Wang et al., 2006), Cdc7 (Kim et al., 2008, Rainey et al.,2008) and RHINO (Rad9, Hus1, Rad1-interacting nuclear orphan protein) (Lindsey-Boltz et al., 2015). Furthermore, the Tim-Tipin complex has been shown to be required for efficient Chk1 activation at stalled replication forks (Chou and Elledge, 2006, Gotter et al., 2007, Unsal-Kacmaz et al., 2007, Yoshizawa-Sugata and Masai, 2007, Errico et al., 2007, Kemp et al., 2010), described in section **1.10.5**. Additionally the PIKK DNA-PKcs has been implicated in maintaining the complex of Claspin-Chk1 during replication stress ATR-Chk1 signalling activation (Lin et al., 2014). Once activated, Chk1 has been shown to have a lower affinity for binding Claspin, resulting in its dissociation from Claspin and dispersal throughout the nucleus, thus enabling a nuclear-wide replication checkpoint response (Jeong et al., 2003, Smits et al., 2006).

Worthy of mention is a region of sequence homology between Claspin and APE2 [Apurinic/apyrimidinic (AP) endonuclease 2] has been identified. APE2 is an endonuclease that generates ssDNA in response to oxidative stress and thus activates ATR-Chk1 signalling. APE2 contains a CKB-motif like region, which when phosphorylated was found bind to Chk1 and result in Chk1 activation (Willis et al., 2013).

The replication checkpoint activation pathway is conserved in the yeasts, where phosphorylated Mrc1 is necessary for the recruitment of the effector kinases spCds1^{Chk1} or scRad53^{Chk1} (mammalian Chk2 homologues), which are then phosphorylated by spRad3^{ATR} or scMec1^{ATR} respectively (Alcasabas et al., 2001, Tanaka and Russell, 2001, Zhao et al., 2003). In *S. pombe* the interaction between

the Claspin and Chk1 orthologues, spMrc1 and spCds1^{Chk1} respectively, occurs via a FHA (forkhead-associated) domain, N-terminal to the kinase domain of spCds1^{Chk1}. spRad3^{ATR} phosphorylates the C-terminus of spMrc1 at Thr645 and Thr653 in a replication stress-dependent manner, which facilitates the binding of spCds1^{Chk1} via its FHA domain (Tanaka and Russell, 2001, Zhao et al., 2003, Xu et al., 2006). This brings spCds1^{Chk1} into close proximity of spRad3^{ATR} for phosphorylation (Tanaka and Russell, 2004). spCds1^{Chk1} is phosphorylated on Thr11 by spRad3^{ATR}, resulting in spCds1^{Chk1} dimerisation, mediated by the FHA domain, and the kinase is fully activated by trans-autophosphorylation on Thr328, enabling downstream signalling (Xu et al., 2006, Xu and Kelly, 2009, Yue et al., 2011). As with Metazoan Chk1, a small C-terminal region of spCds1^{Chk1} is inhibitory to its kinase activity, thus preventing inappropriate spontaneous activation of the protein kinase (Xu and Kelly, 2009).

As in *S. pombe*, the activation of the replication checkpoint in *S. cerevisiae* is almost identical. The central region of scMrc1 is required for mediator functionality, with the C-terminus of scMrc1 required for interaction with scRad53^{Chk1}, however unlike spMrc1 or Claspin, scMrc1 does not contain a specific docking motif for interaction with scRad53^{Chk1} (Xu et al., 2006, Chen and Zhou, 2009, Naylor et al., 2009, Chen et al., 2014). scMec1^{ATR} has been shown to phosphorylate scMrc1, enabling the FHA domain of scRad53^{Chk1} to bind to scMrc1, which brings scRad53^{Chk1} into close proximity of scMec1^{ATR} for phosphorylation (Chen et al., 2014). scMrc1 phosphorylation by scMec1^{ATR} at a stalled replication fork is required for the retention of and continued activation of scMec1^{ATR} at the stalled replication fork (Naylor et al., 2009). The *S. cerevisiae* replication checkpoint was reconstituted *in vitro*, which identified scMrc1 stimulated the activity of scMec1^{ATR} for scRad53^{Chk1} by enzyme-substrate interaction (Chen and Zhou, 2009). Furthermore, co-localisation of scMec1^{ATR} and scMrc1 was identified as being sufficient for the phosphorylation and activation of scRad53^{Chk1} (Berens and Toczyski, 2012). Recently it was reported in *S. cerevisiae,* on the activation of the replication checkpoint, scMrc1 was required for the suppression of repair by homologous recombination (HR) (repair by strand crossover), which can result in genome instability (Alabert et al., 2009). However, for the recovery from the

replication checkpoint, for replication restart, HR is required. scMrc1 release from DNA may enable the assembly of HR centres to complete repair (Alabert et al., 2009, Prado, 2014).

1.10.5 The Fork Protection Complex

Claspin has been shown to interact with Tipin, Tim and And1, forming the 'fork protection complex' (FPC) [also termed the replication pausing complex (RPC)]. The FPC complex is found associated with chromatin, and is required for replication fork stability during unperturbed replication and during replication stress, thereby preventing replication fork collapse, as reviewed in Errico and Costanzo, (2012). The obligate Tim-Tipin complex, travels with the replisome and is required for 'normal' DNA replication progression. Additionally, this complex stabilises stalled replication forks, enhances the activation of Chk1 signalling, and furthermore enables the resumption of DNA synthesis after inactivation of the DNA replication checkpoint (Gotter et al., 2007, Errico et al., 2007, Yoshizawa-Sugata and Masai, 2007, Unsal-Kacmaz et al., 2007, Errico et al., 2009, Smith et al., 2009, Urtishak et al., 2009, Kemp et al., 2010, Leman et al., 2010). And1 is required for 'normal' progression of DNA replication (Errico et al., 2009, Yoshizawa-Sugata and Masai, 2009, Tanaka et al., 2009). However, as with Claspin, the proteins Tim and Tipin have no apparent enzymatic functions or sequence homology to any other structurally characterised proteins – this has made it particularly difficult to dissect and/or assign individual functions to the protein within the FPC complex (Leman and Noguchi, 2012, Smith-Roe et al., 2013).

The FPC has shown to interact directly with the replisome and during unperturbed replication has been found to translocate with the replication fork, and has been proposed to link the replicative CMG helicase complex with the DNA polymerases (Chou and Elledge, 2006, Zhu et al., 2007, Errico et al., 2009, Cho et al., 2013); whereby Claspin and Tim interact with the leading strand DNA polymerase ε (Cho et al., 2013, Aria et al., 2013), And1 and Tipin interacts with RNA primase DNA polymerase α (Zhu et al., 2007, Errico et al., 2009, Cho et al., 2013), and Tim interacts with the lagging strand polymerase DNA polymerase δ (Cho et al., 2013). These interactions are thought to coordinate the activities of DNA unwinding by

the CMG helicase, with DNA polymerase activity through modulation of their activities (Cho et al., 2013, Aria et al., 2013).

The loss of Tim or Tipin resulted in replication stress, prevented the accumulation of Claspin at a stalled replication fork, the inability to efficiently activate Chk1 and a down regulated replication checkpoint activation (Chou and Elledge, 2006, Gotter et al., 2007, Errico et al., 2007, Unsal-Kacmaz et al., 2005, Unsal-Kacmaz et al., 2007, Yoshizawa and Masai, 2007 Smith et al., 2009). An interaction between the FPC and RPA was shown to occur via Tipin, which may recruit Claspin to stalled and/or stressed replication forks with RPA coated ssDNA (Unsal-Kacmaz et al., 2007, Gotter et al., 2007, Nakaya et al., 2010, Kemp et al., 2010).

Additionally, Claspin and the Tim-Tipin complex also appear to have roles in the stability of difficult to replicate DNA tracks or fragile sites during DNA replication. Knockdown of the Claspin protein in cells, caused an increase in fragile site instability, resulting in a number of chromosomal aberrations; this was especially prevalent at tri-nucleotide repeats (Focarelli et al., 2009, Chou and Elledge, 2006, Liu et al., 2012a).

The FPC complex is highly conserved, and in yeasts is composed of the orthologues spMrc1, spSwi1^{Tim}, spSwi3^{Tipin} and spMcl1^{And1} in *S. pombe* and scMrc1, scTof1^{Tim}, scCsm3^{Tipin} and scCtf4^{And1} in *S. cerevisiae*. In the yeasts, the proteins of the FPC have similar identified functionality as identified in Metazoa (Katou et al., 2003, Noguchi et al., 2004, Calzada et al., 2005, Tourriere et al., 2005, Hodgson et al., 2007). These proteins associate with the replication fork (Nedelcheva et al., 2005, Bando et al., 2009, Shimmoto et al., 2009, Uzunova et al., 2014, Tanaka et al., 2010), and a similar organisation is found to connect the replicative helicase and the DNA polymerases, and translocate with the replication fork (Katou et al., 2003, Noguchi et al., 2004, Gambus et al., 2009, Tanaka et al., 2009, Nedelcheva et al., 2005, Bando et al., 2009). scMrc1 associates with scTof1^{Tim} and scCsm3^{Tipin} in the nucleus, and this is required for the chromatin association of scMrc1, and this complex has been reconstituted *in vitro* (Bando et al., 2009, Uzunova et al., 2014, Tanaka et al., 2010). spMrc1 makes a direct interaction with spSwi1^{Tim} and spSwi3^{Tipin} *in vitro*, and this

is required for a high affinity interaction with DNA; specifically to complex DNA structures (i.e. arrested fork structures) (Tanaka et al., 2010). Furthermore, mutation of Lys235 and Lys236, which are required for DNA binding functionality (refer to section **1.10.2**) (Zhao and Russell, 2004, Tanaka et al., 2010) also show a reduced interaction with spSwi1^{Tim} and spSwi3^{Tipin} *in vitro* (Tanaka et al., 2010). Additionally, as found in Metazoa, Mrc1 and the FPC are identified as being required to prevent instability during DNA replication through fragile sites such as tri-nucleotide repeats, microsatellite sites and secondary structure forming sequences (Freudenreich and Lahiri, 2004, Razidlo and Lahue, 2008, Voineagu et al., 2009, Alver et al., 2012, LeClere et al., 2012).

1.10.6 Transcriptional regulation of Claspin

The promoter of the Claspin gene has been identified as being under the control of the E2F transcription factor family; these factors are known to control a number of replication-associated genes (Iwanaga et al., 2006). It has been reported, that gene transcription of Claspin in G_1-phase is prevented by the ubiquitination and degradation of the E2F transcription faction by the APC (Anaphase Promoting Complex)^{cdh1} (APC activator protein) (Gao et al., 2009). Furthermore, the Claspin promoter has also shown to be regulated by the NF-κB (nuclear factor κB) transcription factor family and their regulatory kinase IKK (IκB kinase) (Kenneth et al., 2010). Additionally DNA-PKcs (DNA-dependent protein kinase catalytic subunit) has also been found to be important for transcriptional regulation of the Claspin gene (Lin et al., 2014). However, the levels of mRNA encoding Claspin have been found to be consistent throughout the cell cycle (Kenneth et al., 2010). Furthermore, human Claspin is expressed as two isoforms: full-length Claspin and a slightly truncated isoform that does not contain the C-terminal 7 amino acids. Investigation of the functionality of these two isoforms identified the final 7 amino acids to be required for interaction with Rad9 (9-1-1 complex), and to enable the efficient activation of ATR-Chk1 signalling (Liu et al., 2012b).

1.10.7 Degradation of Claspin

Levels of the Claspin protein are known to fluctuate with the cell cycle and change in response to DNA damage (Chini and Chen, 2003, Mamely et al., 2006, Bennett

and Clarke, 2006). Levels of the Claspin protein, are at their lowest during G_1-phase, but peak during the S- and G_2-phase of the cell cycle, and decrease again at the G_2- to M-phase transition (Chini and Chen, 2003). The protein is continually degraded by the ubiquitin-proteasome pathway (Chini et al., 2006); by both the APC^{Cdh1} (Bassermann et al., 2008, Faustrup et al., 2008, Gao et al., 2009) and SCF (Skp, Cullin, F-box containing complex) $^{\beta\text{-TrCP}}$ (β-transducin repeat containing E3 ubiquitin protein ligase) E3 ubiquitin ligases (Peschiaroli et al., 2006, Mailand et al., 2006, Mamely et al., 2006, Studach et al., 2010); but Claspin is transiently stabilized during S-phase by a number of ubiquitin proteases, with further stabilization during replication checkpoint activation (Mamely et al., 2006, Bennett and Clarke, 2006, Gao et al., 2009).

The protein contains a conserved degron motif (aa 328-340), more specifically the LLK (Leu-Leu-Lys) sequence (aa 333-335) (Gao 2009). Low levels of Claspin protein are maintained during G_1-phase by targeting to the proteasome by the APC^{Cdh1} (Bassermann et al., 2008, Faustrup et al., 2008, Gao et al., 2009). Mutation of the LLK sequence or inhibition of Cdh1 activity, results in increased levels of Claspin protein and the subsequent activation of Chk1 in G_1-phase (Gao et al., 2009).

During DNA replication and replication checkpoint activation, Claspin is stabilised in order to enable efficient DNA replication and the activation of the effector kinase Chk1 respectively. The USP7 (ubiquitin carboxyl-terminal hydrolase), an ubiquitin protease, interacts with and stabilises Claspin during the processes of DNA replication and during replicative stress; it deubiquinates Claspin, thus preventing it being targeting to the proteasome by $SCF^{\beta\text{-TrCP}}$ (Faustrup et al., 2009). Interestingly USP7 has no effect on the degradation of Claspin in G_1-phase, but it is re-activated in M-phase, with DNA damage, which results in Claspin protein stabilisation (Faustrup et al., 2009). Other ubiquitin proteases have also been shown to have roles in preventing the degradation of Claspin during DNA replication and the replication checkpoint, and are required for the activation of ATR-Chk1 signalling; these include USP20 (Yuan et al., 2014, Zhu et al., 2014), USP28 (Zhang et al., 2006) and USP29 (Martin et al., 2014). USP20 and USP28 were

identified as being activated by ATR phosphorylation (Zhu et al., 2014, Zhang et al., 2006), and with knockdown of these proteins, cells had S-phase defects and impaired the activation of the replication checkpoint, whilst overexpression increased the levels of Claspin protein (Zhu et al., 2014, Martin et al., 2014). Furthermore, during replication stress, Claspin has been identified as a phosphorylation target of Chk1 and this has been proposed to enhance the stability of Claspin, thereby enabling further Chk1 activation (Chini et al., 2006, Mamely et al., 2006, Bennett and Clarke, 2008).

For replication checkpoint recovery and/or for entry into G_2-phase of the cell cycle, the levels of the Claspin protein are again reduced (Mailand et al., 2006, Mamely et al., 2006, Studach et al., 2010). In this phase of the cell cycle, Claspin is phosphorylated by Plk1 (Polo-like kinase) protein kinase and targeted for degradation by the $SCF^{\beta\text{-}TrCP}$ ubiquitin ligase (Peschiaroli et al., 2006, Mailand et al., 2006, Mamely et al., 2006, Studach et al., 2010). The Polo-box domain (PBD) of Plk1 has been found to interact with the C-terminus of Claspin (Mamely et al., 2006). Claspin contains a DSGxxS phospho-degron sequence at the N-terminus (aa 29-34), which is phosphorylated by Plk1 kinase (Peschiaroli et al., 2006, Mailand et al., 2006). The phosphorylated degron provides a binding site for the β-TrCP subunit of the $SCF^{\beta\text{-}TrCP}$ ubiquitin ligase, which ubiquitinates Claspin, and targets it for proteasomal degradation (Mailand et al., 2006, Mamely et al., 2006, Peschiaroli et al., 2006). Mutation of the Claspin phospho-degron motif, to inhibit Plk1 phosphorylation, or depletion of the Plk1 kinase in human cells, prevented phosphorylation and resulted in stabilisation of the Claspin protein, as well as prolonged activation of Chk1 and delays in cell cycle progression (Mailand et al., 2006, Mamely et al., 2006, Peschiaroli et al., 2006). Similar findings were reported for experiments in *Xenopus* (Yoo et al., 2004), where additional phosphorylation of *Xenopus* Claspin by Plx1 (*Xenopus* Plk1 homologue) resulted in disassociation of Claspin from chromatin; required for adaptation to the replication checkpoint (Yoo et al., 2004). Furthermore, modulation of the APC^{Cdh1} complex has also been shown to result in decreased levels of the Claspin protein on the exit of S-phase; Cdh1 levels in the cell rise during S-phase (a process dependent on CDK activity), and

Claspin is targeted for degradation by ubiquitination, which results in decreased Chk1 activity and thus enabling progression of the cell cycle (Oakes et al., 2014).

In response to DNA damage in G_2-phase, the proteolysis of Claspin is prevented, in order to establish the replication checkpoint, resulting in G_2-M arrest and providing time for DNA damage repair. The APC^{Cdh1} ubiquitin ligase was identified as being re-activated, and as in G1-phase this targets both Plk1 and Claspin for degradation; but the action on Claspin is countered by Usp28, thereby preventing Claspin degradation and enabling Chk1 signalling activation (Bassermann et al., 2008).

When a cell is unable to exit the replication checkpoint, the cell undergoes apoptosis; Claspin is cleaved at Asp1072 by Caspase-7, resulting in a decrease of Chk1 activity (Clarke et al., 2005, Semple et al., 2007). Caspase-3 cleavage sites have also been described: including Asp25 and also a number of sites at the C-terminus of the protein that were not identified (Semple et al., 2007). Of a worthy note, expression of Caspase-7 can be modulated by STAT3 (Signal transducer and activator of transcription 3) in Epstein Barr Virus (EBV) infected cells. This was found to result in decreased levels of Claspin protein and to inhibit activation of the replication checkpoint during EBV infection (Koganti et al., 2014).

As in Metazoa, spMrc1 protein levels are controlled during the cell cycle (Tanaka and Russell, 2001). scMrc1 has been identified as a target of the SCF^{Dia2} ubiquitin ligase complex, which ubiquitinates scMrc1 *in vivo* and *in vitro* targeting this for proteasome-mediated degradation; this has been identified as being required for recovery from the replication checkpoint (Mimura et al., 2009, Fong et al., 2013). The SCF^{Dia2} complex has been found tethered to the replisome during DNA replication, by interactions with scMrc1 and $scCtf4^{And1}$, and this appear to be required during chronic replication fork stalling (Morohashi et al., 2009). Furthermore, the interaction between scMrc1 and SCF^{Dia2} partially overlaps interaction sites scMrc1 makes with scPol2, which is lost during replication stress (Lou et al., 2008, Mimura et al., 2009), and this could indicate competition for interaction between the two proteins and scMrc1.

Of note, in a recent study in *S. cerevisiae*, scMrc1 was identified as being localised to an 'intranuclear quality control compartment' (INQ), which required Dia2 interaction. This compartment may play a role in the recovery from genotoxic stress through the sequestering and turnover of PTM replication stress proteins (Gallina et al., 2015).

1.10.8 Claspin and Cancer

There have been a number of studies that have mentioned Claspin in regards to cancer, in particular: where increased levels of the protein or mRNA have been detected (Lin et al., 2004, Tsimaratou et al., 2007, Verlinden et al., 2007, Benevolo et al., 2012, Choi et al., 2014, Ganzinelli et al., 2011, Allera-Moreau et al., 2012); a few mutations of the Claspin gene have been found (Erkko et al., 2008, Zhang et al., 2009b); and Claspin has been used as a proliferative marker for specific tumour types (Tsimaratou et al., 2007, Verlinden et al., 2007, Benevolo et al., 2012, Zhang et al., 2009b, Allera-Moreau et al., 2012). However, the relationship of Claspin with cancer has yet to be determined, but speculation would indicate S-phase progression, which would enable cancer progression.

1.11 Biochemical and Biophysical Techniques

The subsequent part of this introduction culminates with a description of a number of biochemical and biophysical techniques (used in this thesis) for the characterisation and/or determination of the structure of a recombinant protein.

1.11.1 Fluorescence polarisation

Fluorescence polarisation (FP) or fluorescence anisotropy (Perrin, 1926), enables the study of an interaction in solution and determination of an equilibration dissociation constant (K_d) for that interaction (Lea and Simeonov, 2011); however, the stoichiometry of the interaction is not determined as FP assays are carried out in sub-stoichiometric conditions, in order to enable accurate K_d determination (Huang, 2003).

FP requires one macromolecule (such as DNA or a synthetic peptide) to be labelled with a fluorophore. Excitation of this fluorophore causes an electron to move up to

a higher energy state, and after a short period the electron drops back to the lower energy state and the absorbed energy is re-emitted as a photon but at a longer (emission) wavelength. In FP, the fluorophore is selectively excited with plane-polarised light; only molecules with their fluorophores correctly orientated with respect to the incoming light absorb a photon. Un-liganded or 'free' fluorophores tumble rapidly and randomly in solution and therefore any re-emitted photons are randomly emitted in all directions (loss of coherence). However, when bound by an interacting macromolecule, the rate of tumbling is decreased (larger hydrodynamic radius) and photons are re-emitted in the same plane as the incoming light (coherent). The use of two photomultiplier tubes to detect emitted photons in the same plane (parallel) or perpendicular planes, allows the use of the following formula to calculate FP in millidegrees; as shown in **Equation 1.1**; where P = fluorescence polarisation, and $F_{\parallel}$ and $F_{\perp}$ = fluorescence intensity parallel and perpendicular to the excitation plane respectively (Lea and Simeonov, 2011).

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}}$$

Equation 1.1

1.11.2 Circular Dichroism

Circular Dichroism (CD) is a spectroscopic technique, which measures differences in the absorption of left- and right-circularly polarised light (Kelly et al., 2005, Greenfield, 2006). This method can be used to study chiral molecules (such as proteins and nucleic acids), which have differential absorption of the polarised light at particular wavelengths. CD provides a reliable method to examine the 'folded-ness' of a protein and estimate its secondary structure composition. CD spectra are typically recorded at wavelengths between 180 and 260 nm; the resultant spectrum provides distinct signatures that can be classified by type such as α-helix, β-sheet or random coil (**Figure 1.10**). CD spectra can be readily compared in order to identify conformation changes as a result of temperature, additives or various types of interactions (including PPIs and protein-nucleic acid interactions). The percentage of each secondary structure element can be

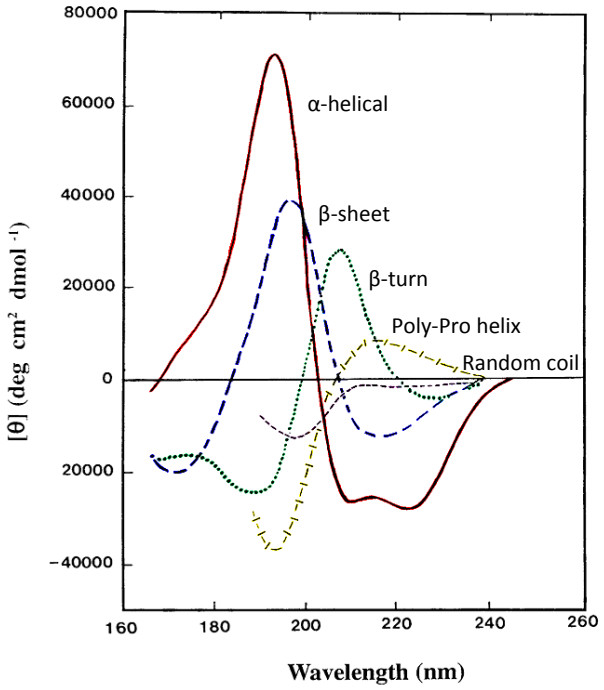


Figure 1.10 CD spectra for defined protein secondary structures.
α-helical (red line), β-sheet (blue line), β-turn (green line), poly-proline helix (extended helix) (yellow line) and random coil (purple line) signatures. Adapted from Kelly et al., (2005).

estimated by deconvolution of the measured spectrum. This process uses algorithms that match the experimental spectrum to those of proteins with known tertiary structure. The web-based server DichroWeb (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) uses a number of separate algorithm-based programs such as SELCON (Sreerama and Woody, 1993), CONTIN (Provencher and Glockner, 1981) and CDSSTR (Johnson, 1999) to deconvolute the experimental data using specific reference datasets for spectra comparison (Sreerama et al., 2000a, Sreerama, et al., 2000b, Sreerama and Woody, 2004).

1.11.3 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) enables the analysis of macromolecules and macromolecular interactions, in their native states, in solution, without need for calibration or approximation (Ralston, 1993, Howlett et al., 2006, Cole et al., 2008). AUC enables the analysis of hydrodynamic and thermodynamic properties of macromolecules across a wide range of masses, from peptides to viruses. Briefly, analytes are subjected to a centrifugal force, which causes any macromolecules present in the sample to migrate through the solution, away from the central axis. This movement can be recorded (typically by UV absorption at a given wavelength) as a function of time. Centrifugation forms a concentration gradient and a solvent / macromolecule boundary whose movement is dependent on the mass of the macromolecule and the inverse of its frictional ratio (shape of the macromolecule). Spreading of the solvent / macromolecule boundary is also observed with time due to diffusion effects, this opposes the centrifugal movement, which is independent of the particle density but is dependent on the particle size.

Sedimentation velocity (SV) and sedimentation equilibrium (SE) are two types of AUC experiment. SV is typically carried out at high velocities (40-60 000 rpm) for short durations (3-4 hours), and measures the movement of solvent / macromolecule boundaries with time (**Figure 1.11A**). These measurements enable the determination of hydrodynamic properties of the macromolecule(s) and the calculation of molecular mass, stoichiometry and the shape of the macromolecule (spherical or elongated). SE is carried out at much lower velocities for longer

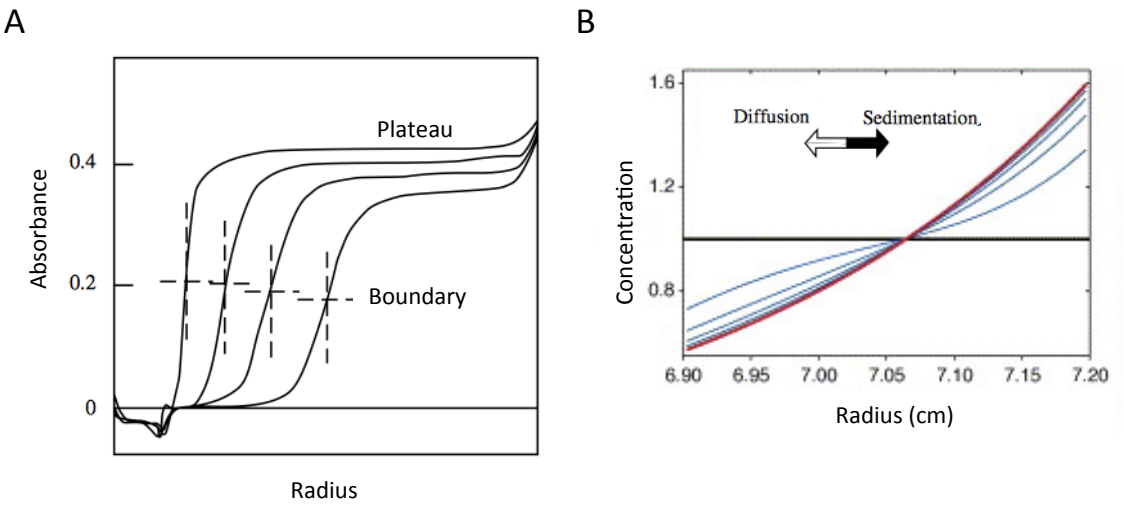


Figure 1.11 Analytical ultracentrifugation.
(**A**) SV-AUC example plot, showing the movement of the sample boundary upon centrifugation (as a function of time), with the radius of the cell, as indicated on the plot. Adapted from Ralston, (1993). (**B**) SE-AUC example plot, showing the sedimentation (black arrow) and the diffusion (white arrow) across a cell. This shows the absorbance gradient changing (blue lines) upon centrifugation and the final concentration gradient (red line). Adapted from Howlett et al., (2006).

durations and measures the sample distribution after reaching equilibrium (**Figure 1.11B**). This method also enables the determination of hydrodynamic properties for a macromolecule as well as the stoichiometry and strength of interaction of any complexes in solution.

1.11.4 Biological Small Angle X-ray Scattering

Biological Small Angle X-ray Scattering (bioSAXS) is a method that can be used to analyse biological macromolecules in solution (Jacques and Trewhella, 2010, Mertens and Svergun, 2010). For bioSAXS, a highly collimated and focused x-ray beam (typically from a synchrotron source) is used to irradiate a sample containing a given macromolecule and any scattered x-rays (at small angles, typically 0.1 – 10 degrees) are detected. As the X-rays pass through the sample, some will interact with the macromolecules and result in elastic scattering, which is detected as a 'ring' of varying intensity. bioSAXS is intrinsically a low-resolution structural technique; macromolecules are randomly orientated in solution, which results in a rotational averaging of the scattered X-rays and thereby limits the amount of information gathered from an experiment. However, the data collected can be used to analyse the size of the macromolecule, including the molecular mass, the radius of gyration (Rg), the volume of the macromolecule, and maximum diameter of the macromolecule. Furthermore, shape and conformational alterations of a macromolecule can be monitored.

In recent years a number of *ab initio* algorithms have been developed to model bioSAXS data as low-resolution 'bead-models', these include DAMMIN (Dummy Atom Model Minimization) (Svergun, 1999) and DAMMIF (Dummy Atom Model Minimization Fast) (Franke and Svergun, 2009). These data can enhance or clarify the details of structures determined by X-ray crystallography or NMR, especially where crystal artefacts, structural anomalies or flexible regions are suspected (Jacques and Trewhella, 2010, Mertens and Svergun, 2010). bioSAXS can also provide insights into the structures of large proteins and protein complexes, in addition to monitoring conformation changes as a result of interactions with a partner molecule or protein.

1.11.5 Nuclear Magnetic Resonance

NMR Spectroscopy is an 'in solution' structural technique that enables the determination of the tertiary structure of both molecules and macromolecules (Poulson 2002, Marion, 2013). Protein NMR studies typically requires isotope labelling of the protein, i.e. those that have detectable nuclear spin (NMR-active) when placed in magnetic fields. Such isotopes include those in natural occurrence; such as ^{1}H and ^{31}P, or those that can be incorporated into a protein by labelling methods; ^{13}C and ^{15}N (note: ^{12}C and ^{14}N isotopes are not active by NMR). NMR-active isotopes have two energy levels and the nuclei flip to the higher energy state when exposed to electromagnetic radiation (radio-wave frequency). Return of the nuclei to the lower energy state, results in in the emission of a pulse of electromagnetic radiation, which can be detected. The resonance frequency of the emitted pulse (the chemical shift) is dependent on the chemical environment of the nuclei from the surrounding electrons. For large molecules such as proteins, the recorded chemical shifts can overlap, and therefore become complicated to interpret. For a ^{1}H- and ^{15}N-labelled protein, there is a coupled transfer of magnetism between ^{1}H and ^{15}N nuclei, which can be recorded in a hetero-nuclear single-quantum correlation spectroscopy (HSQC) experiment. Each observed peak represents a coupled ^{1}H-^{15}N nuclei from an amino acid in the backbone of a protein sample (with the exception of proline). HSQC plots can also provide an initial indication as to the 'folded-ness' of a protein. The tertiary structure of a protein can be determined by the assignment of all the chemical shift peaks on an HSQC plot; however full and correct assignment of a spectrum can take many months. Typically proteins with a molecular mass below 20 kDa are suitable for study by NMR, although proteins up to 30 kDa can also be studied with protocol modifications, such as using three isotope labelling and stronger magnetic fields. As the molecular mass of a protein increases the tumbling rate decreases, which causes broadening of the NMR signal, resulting in difficulties in spectrum assignment (Kay, 2005).

1.11.6 X-ray crystallography

X-ray crystallography is the most commonly used technique for the determination of the tertiary structure of proteins and their complexes (Rhodes, 2006). The

method requires crystals; orderly repeating lattices of molecules, which are not held together by covalent interactions. These are typically grown using purified proteins (in aqueous buffer) by vapour diffusion (controlled evaporation in a closed system) in either a sitting- or hanging-drop format. Matrix solutions (buffer, salt, precipitant and/or additives) are used for the controlled precipitation of the protein, which if the conditions are optimal, result in crystal growth. X-ray diffraction uses monochromatic X-ray electromagnetic radiation produce by either an 'in-house' rotating anode or at a synchrotron source. Protein crystals; specifically the electron cloud surrounding each nuclei of the protein molecule; interact with and diffract the incident x-ray beam, resulting in a diffraction pattern (reflections) that can be detected and measured. Each reflection results from constructive interference to produce an 'in-phase' wave (termed coherent scattering) from the crystal lattice at angles which satisfy Bragg's law (**Equation 1.2**); where λ is the incident wavelength (nm), d_{hkl} is the interplanar spacing (nm), n is an integer and sinθ is the angle of scatter (degrees).

$$n\lambda = 2d_{hkl}\sin\theta$$

Equation 1.2

After measurement of the intensities of each reflection, from a given crystal, and assignment of the correct space group (the arrangement of molecules within the crystal), the resulting data can be used to determine an electron density map, which is constructed from Fourier transforms of so-called structure factors (F_{hkl}). The structure factor is composed of the amplitude of the reflection (measured and related to the intensity of the reflection) and the phase of the reflection, which is not measured directly and must be separately determined; this is termed the crystallographic 'Phase Problem'. The Phase Problem must be simultaneously solved for every reflection as they each have a unique phase term. The phase problem can be solved computationally by molecular replacement; whereby the phases for the tertiary structure of a homologous protein/s can be used to solve the structure of the new protein. Alternatively, for a protein with no similar tertiary structure, the phase problem can be solved experimentally, by techniques such as isomorphous replacement and anomalous scattering – which rely on

measurable differences in reflection pairs (Friedel / Bijvoet pairs). With both phase and intensity measurements, the resulting electron density maps can then be used to build the tertiary arrangement of amino acids for the crystallised protein. Typically this involves many rounds of manual building and refinement in order to generate the final experimental model.

Chapter 2

Materials and methods

The chemicals used in this thesis were purchased from Sigma or Fisher Scientific unless otherwise stated.

2.1 Bacterial medium

Bacterial medium used for the culturing of *E. coli*, was supplemented with 100 µg/ml ampicillin sodium salt, unless otherwise stated.

Luria-Bertani (LB) medium

1% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl.

Luria-Bertani Agar (L-agar plates)

LB medium containing 1.2% (w/v) agar.

Super Optimal broth with Catabolite repression (S.O.C.) medium

S.O.C (Invitrogen).

Turbo Broth (TB) medium

4.76% (w/v) Turbo Broth powder (Molecular Dimensions), 0.4% (v/v) glycerol.

Auto-induction medium

6% (w/v) powered Overnight Express Instant TB Medium (Novagen), 1% (v/v) glycerol, pH 6.9 ± 0.2.

NMR non-inducing starter (NMR NI) medium

0.5% (v/v) OnEx solution 3, 1% (v/v) OnEx solution 4, 1% (v/v) OnEx NMR solution 5, 1% (v/v) OnEx NMR solution 6, 1% (v/v) OnEx solution 2 [Overnight Express Autoinduction NMR Medium kit (Novagen)].

^{15}N-labeled NMR autoinduction inducing (NMR auto-induction) medium

1% (v/v) OnEx Solution 3, 1% (v/v) OnEx solution 4, 1% (v/v) OnEx NMR solution 1, 1% (v/v) OnEx NMR solution 2, 5 mM [^{15}N] ammonium chloride [Overnight Express Autoinduction NMR Medium kit (Novagen)].

2.2 Molecular cloning

2.2.1 Plasmid and construct DNA

Protein expression plasmids and construct DNA were sourced from: MEK1 4F11 (MEK1-KD), NF-kB p65, NF-kB p65 RHR (Rel homology region) and, C/EBPβ (CCAAT/enhance binding protein beta) containing plasmids were from Dr Stefanie Reich (Domainex); Chk2-KD containing plasmid was from Dr Antony Oliver (University of Sussex); full-length Chk1 (His-Chk1) construct was from Prof Keith Caldecott (University of Sussex); GST was from the 'in house' p3E vector created by Dr Antony Oliver (University of Sussex); His-BRaf V600E containing plasmid from Dr Sigrun Polier (University of Sussex); His-p50 (Hsp90 co-chaperone Cdc37) containing plasmid from Dr Chrisostomos Prodromou (University of Sussex); full-length Claspin, Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His constructs, and the GST-NIK (kinase active / kinase dead) and His-NIK (kinase active / kinase dead) constructs were synthesized (GenScript); the Claspin fragment constructs were created from the full-length Claspin construct.

2.2.2 Polymerase chain reaction and amplified product purification

PCR for standard cloning was performed using KOD hot start DNA polymerase kit (EMD Millipore). Typically, PCR reactions contained 1.5 mM $MgSO_4$, 0.8 mM dNTP mix, 0.6 μM 3'- and 5'-primers (section **2.2.2.1**), 20 ng template DNA, 1 Unit (U) KOD Polymerase in 1x KOD hot start DNA polymerase buffer, in a final volume of 50 μl. Typical PCR conditions were; 95 °C for 5 minutes for pre-denaturation, and 30 cycles of; 95 °C for 30 seconds for denaturation, 65 °C for 30 seconds for primer annealing and 72 °C for 1 minute/kb DNA to be amplified; with a final step of 72 °C for 10 minutes for complete extension. PCR products were purified using the illustra GFX PCR DNA and gel band purification kits (GE Healthcare), following the manufacturers instructions.

2.2.2.1 DNA cloning primers

All DNA primers were purchased from MWG Biotech. These were all dissolved in dH_2O and are listed in **Table 2.1**.

Name	Sequence
Claspin Fw	5’-ATGACGGGTGAGGTGGGTTC
Claspin Rv	5’-GGATTCGAGGTACTTGAAAATG
Claspin NdeI Fw	5’-CATATGACGGGTGAGGTGGGTTC
Claspin HindIII Rv	5’-AAGCTTGGATTCGAGGTACTTG
A1G6D NdeI Fw	5’-TCGCATATGTACCAGTCGTCGCACCATAAAG
A1G6D Rv His	5’-TCGAAGCTTTTATTAATGGTGATGGTGATGGTGCCCTTCG
A1G12 NdeI Fw	5’-CGCATATGCCCTTCTACAAAACTGTAGCTGACTCTG
A1G12 HindIII Rv His	5’-CGAAGCTTTTATTAATGGTGATGGTGATGGTGCCCTTGTTTT
Claspin MPENK NdeI Fw	5’-CGCATATGCCGGAGAACAAAACCATCCACG
Claspin YQS NdeI Fw	5’-CGCATATGTACCAGTCGTCGCACCATAAAGAA
Claspin R1 HindIII Rv His	5’-CGAAGCTTTTATTAATGGTGATGGTGATGGTGCCCCTGTTT TTCCTGAATTTCC
Claspin R2 HindIII Rv His	5’-CGAAGCTTTTATTAATGGTGATGGTGATGGTGCCCTTCGCTGTGATTGTTACCG
Chk1 FL NheI Fw	5’-CGGCTAGCATGGCAGTGCCCTTTG
Chk1 FL HindIII Rv	5’-CGAAGCTTTTATTATGTGGCAGGAAGCC
Chk1 NdeI Fw	5’-CATATGGGATCCATGGCTGTTCCGTTTG
Chk1 273 BamHI Fw	5’-CGGGATCCATGGCTGTGCCGTTTGTGG
Chk1 273 HindIII Rv	5’-CGAAGCTTTTATTACCCTTTCTTGAGGGG
Chk1 NdeI Fw	5’-CGCATATGGCAGTGCCCTTTGTGG
Chk1 XmaI Fw	5’-CGCCCGGGATGGCAGTGCCC
Chk1270 HindIII Fw	5’-CGAAGCTTTTATTAATGGTGATGGTGATGGTGTTTCAGCGGTTTATTATACC
Chk1 270 BamHI Rv	5’-ATGGATCCTTATTATTTCAGCGGTTTATT
Chk1 289 Rv His	5’-GAATTCAAGCTTTCATCAGTGGTGGTGGTGGTGGTGACCTG

Table 2.1 DNA cloning primers.
Where the target and restriction enzyme are listed and Fw=forward, Rv=reverse, His=His_6 affinity tag.

2.2.3 Restriction endonuclease digest

Restriction endonucleases were used to create complementary ends for subsequent ligation of the insert into the vector. Typically, 40 μl of purified PCR DNA product or 40 μl of <5 μg of vector DNA was cleaved in double digest reactions containing 2.5 μl each restriction endonuclease (New England Biolabs) in 1x NEB Buffer, to a final volume of 50 μl. These were incubated at 37 °C for 2 hours, and the reactions were terminated by a brief incubation at 60 °C. The DNA was purified by agarose gel electrophoresis.

2.2.4 Agarose DNA electrophoresis and purification

DNA samples were resolved by size by agarose gel electrophoresis on a 1.0% (w/v) agarose 1x TAE (40 mM Tris acetate, 1 mM EDTA, pH 8.2-8.4) gel containing 0.04 μl/ml SYBR Safe DNA gel stain (Invitrogen), at 100 V for 30 minutes. DNA samples were resuspended with ~2x DNA Loading Buffer Blue (Bioline), and

resolved against HyperLadder 1kb I DNA marker (Bioline). Agarose gels were visualized and imaged using a UV transilluminator station (Syngene) with GeneSnap software (Syngene InGenious system). For DNA extraction, agarose gels were visualized using a UV transilluminator and the respective bands were extracted. In-gel DNA samples were purified using the illustra GFX PCR DNA and gel band purification kits, following the manufacturers instructions.

2.2.5 DNA ligation

Ligation of the digested PCR DNA inserts and vectors was by Clonables 2x ligation premix (Novagen). Ligation reactions were carried out typically at a 3:1 and 4:1 of insert:vector ratio (to 5 μl) as determined from the DNA concentration (section **2.4**). Ligation reactions of 5 μl ligation mix with 5 μl Clonables 2x ligation premix, incubated at room temperature for between 15-60 minutes, prior to *E. coli* transformation (section **2.3.1**). Claspin DNA fragments were ligated to pDXV4 vector (Domainex), His-Chk1 were ligated into 'in house' p2B vector based on pFastBac created by Dr Antony Oliver (University of Sussex), and Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His were ligated into pFastBac1 (Life Technologies).

2.3 DNA plasmid propagation

2.3.1 Transformation of *E. coli* XL10 Gold competent cells

Ultracompetent *E. coli* XL10 Gold cells (Agilent) were used for the propagation of plasmid DNA. Typically, 5 μl of the ligation reaction or 50-200 ng/μl (1-2 μl) of plasmid DNA was incubated on ice for 30 minutes with 12-25 μl *E. coli* XL10 Gold cells that had thawed on ice. Cells were heat-shocked for 45 seconds at 42 °C, returned to ice for 2 minutes, resuspended with 150 μl LB medium (no antibiotics), and incubated shaking at 37 °C for 30-60 minutes. The culture was plated onto a L-agar, and incubated overnight at 37 °C, then inspected for colonies.

2.3.2 Colony polymerase chain reaction

Colony PCR to confirm a DNA insert from DNA ligation was performed using PuReTaq Ready-To-Go PCR beads (GE Healthcare). Typically, a PCR bead was resuspended in 25 μl dH_2O containing 3.5 μM 3'- and 5'-primer (section **2.3.4**).

Individual colonies were picked, resuspended in the bead solution, and re-streaked on a L-agar plate. For purified DNA, ~20 ng DNA was resuspended in the bead solution. PCR conditions were as in section **2.2.2**, but with an extension of 2 minutes/kb DNA to be amplified. DNA was resolved by agarose gel electrophoresis (section **2.2.4**).

2.3.3 Amplification and purification of plasmid DNA

Plasmid DNA was amplified by the inoculation of 5 ml LB medium with a single colony, and incubated shaking overnight at 37 °C. Cells were harvested the next day by centrifugation at 4000 rpm for 5 minutes. Plasmid DNA was extracted and purified from the cell pellet using the QIAprep Spin MiniPrep kit (QIAgen), following the manufacturers instructions.

2.3.4 DNA sequencing and analysis

DNA sequencing was carried out by Source Bioscience (Oxford). DNA sequencing primers (**Table 2.2**) were purchased from MWG Biotech and dissolved in dH_2O. DNA sequencing was analysed used 4Peaks software (Nucleobytes) and Blast (Basic Local Alignment Search Tool) alignment.

Name	Sequence	DNA
T7F	5'-TAATACGACTCACTATAGGG	All *E. coli*
T7R	5'-GCTAGTTATTGCTCAGCGG	All *E. coli*
PolHF	5'-ACCCGGCAAGAACCAAAAACTCACT	p2B
PolHR	5'-TCCAAGTTTCCCTGTAGAACTCTTTTCCTT	p2B
SV40pA-F	5'-TATTTATGCAGAGGCCGAGG	pFastBac1
SV40pA-R	5'-GAAATTTGTGATGCTATTGC	pFastBac1
pUC/M13F	5'-GTTTTCCCAGTCACGAC	Bacmid
pUC/M13R	5'-CAGGAAACAGCTATGAC	Bacmid

Table 2.2 DNA sequencing primers.
Where F=forward, R=reverse.

2.4 Determination of DNA and protein concentrations

The concentration of both DNA and protein were measured using a NanoDrop 2000C spectrophotometer (Thermo Scientific). Absorbance readings for DNA and proteins were measured at 260nm (A_{260}) and 280nm (A_{280}) respectively.

2.5 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to resolve protein samples, using pre-cast 10- or 15-well 1 mm NuPAGE Novex 4-12% Bis-Tris gels (Invitrogen). Protein samples were prepared in 4x NuPAGE Lithium Dodecyl Sulphate (LDS) Sample Buffer (Invitrogen) with 5% (v/v) β-mercaptoethanol (LDS-LB), at a 2:1 ratio (protein:LDS-LB). These were boiled at 95 °C for 5 min and briefly centrifuged. Gels were resolved in a XCell SureLock Mini-Cell Electrophoresis System (Invitrogen) in 1x NuPAGE MES SDS Running Buffer (Invitrogen), at 200 V for 40 minutes. Protein samples were resolved against SeeBlue Plus2 Pre-stained Protein Standard (Invitrogen). For **Figures 6.9B**, **7.1A**, **7.11**, and **Appendix 3.4C** these used Unstained Protein Ladder, Broad Range (10-250 kDa) (New England Biolabs). For **Figures 6.9B** and **6.9C**, these used 10-well RunBlue Precast Gels 4-20% Tricine gels (Expedeon) in 1x RunBlue SDS-PAGE running buffer (Expedeon), at 160 V for 60 minutes. For **Figure 7.11** this used a 12-well TruPAGE Precast Gels 4-20% gels (Sigma) in Tris-MOPS SDS running buffer [1.2 M Tris, 0.6 M MOPS, 2% (w/v) SDS], at 180 V for 45 minutes.

2.5.1 SDS-PAGE Staining

Resolved SDS-PAGE gels were stained in Instant Blue protein stain (Expedeon), a Coomassie base gel stain, by microwaving in stain for 30 seconds and rocking for 10 minutes. Gels were destained in dH_2O until the background stain was minimal. Protein gels were imaged using GeneSnap software (Syngene InGenious system).

2.6 Western blotting

Resolved SDS-PAGE gels were transferred onto Hybond ECL 0.45 nitrocellulose membrane (GE Healthcare) or polyvinylidene difluoride (PVDF) membrane (Amersham) by electro-blotting, via semi-dry transfer using the TransBlot Turbo Transfer System (Bio-Rad). PVDF membrane was pre-soaked in 100% methanol for 5 minutes for activation and washed in dH_2O. The membranes and 0.8 mm blotting paper (Whatman) were soaked in western Blot transfer buffer (0.025 M Tris pH 8.3, 0.192 M glycine, 20% (v/v) methanol) for 5 minutes, prior to transfer sandwich assembly. Transfer occurred at 25 V for 30 minutes. Blotted membranes

were blocked with Phosphate Buffered Saline (PBS) with 0.1% (v/v) Tween-20 (PBST) and 5% (w/v) milk powder (Marvel), for 5 minutes. The 6xHis Monoclonal Antibody (Clontech) primary antibody (1/3000 dilution) was added and incubated rocking for 30 minutes at room temperature. The membranes were rinsed twice with PBST. Subsequently, Anti-Mouse IgG (H+L) Alkaline phosphatase Conjugate (Promega) secondary antibody (1/5000 dilution) in PBST was incubated with the membrane rocking for 30 minutes rocking at room temperature. The membranes were rinsed three times with PBST. These were developed using SIGMA*FAST* BCIP (5-bromo-4-chloro-3-indolyl phosphate) / NBT (nitroblue tetrazolium) alkaline phosphatase substrate (Sigma), until purple bands were visualized (~5 minutes), and the membranes were washed with dH_2O. For this colourmetric reaction, the alkaline phosphatase conjugate secondary antibody dephosphorylates BCIP to an "indoxyl" intermediate, and NBT oxidizes this intermediate to produce a purple precipitate. NBT is reduced forming a blue diformazan precipitate. Both coloured products precipitate at the secondary antibody site, forming a visible blue spot.

2.7 Bioinformatics

Gene sequences were accessed from the UniProt database (Consortium, 2014), and alignments were prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010). Sequence homology searches were carried out using a web-based Blast search (Altschul et al., 1990). Secondary structure predictions were performed using the web-based services Phyre^2 (Kelley and Sternberg, 2009) and PsiPred (Jones, 1999, Buchan et al., 2013). Secondary structure disorder predictions were carried out using DisoPred (Buchan et al., 2013). Claspin template-based models were produced by ModBase (Pieper et al., 2011) and DomSerf, annotated on Genome3D (Lewis et al., 2013, Buchan et al., 2013). Homology modelling for full-length Claspin and Claspin protein fragments (Uniprot ID: Q9HAW4) was performed using SwissModel (Schwede, 2003). Protein threading used the Raptor X server (Kallberg et al., 2012). Physical and chemical properties from the primary sequence were determined using ExPASy ProtParam (Gasteiger et al., 2005). Tryptic digest predictions were carried out using the ExPASy PeptideCutter tool (Gasteiger et al., 2005).

2.8 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) was used to measure the interaction between STU and MEK1-KD protein using an ITC microcalorimeter (Microcal), with a 1.4 ml cell volume, at 30 °C. Dialysed MEK1-KD protein was supplemented with 2% (v/v) DMSO (dimethyl sulfoxide). STU was prepared in MEK1 GF buffer [20 mM HEPES pH 8.0, 300 mM NaCl, 6% (v/v) glycerol, 0.5 mM TCEP (tris(2-carboxyethyl)phosphine)] used in dialysis, with 2% (v/v) DMSO. Nine 13.5 μl aliquots of 231 μM MEK1-KD were injected into 20 μM STU. Heats of dilution were determined separately by titration of MEK1-KD (as above) into MEK1 GF buffer with 2% (v/v) DMSO. Corrected data were fitted using a non-linear least-squares curve-fitting algorithm (Microcal Origin), with three floating variables, stoichiometry, binding constant, and change of enthalpy of interaction.

2.8.1 Protein dialysis

MEK1-KD protein was dialyzed using a 3.5K MW cut-off Slide-A-Lyzer dialysis cassette (Pierce) for buffer exchange and to remove buffer impurities, following the manufacturers instructions. MEK1-KD protein was injected into the cassette, and incubated rotating in 2 litres MEK1 GF buffer overnight, at 4 °C, subsequently MEK1-KD protein was extracted and used for ITC.

2.9 Affinity probe for the Octet

2.9.1 Modelling binding of the affinity probe

The structure of the affinity probe (AP) was drawn using the PRODRG server (Schuttelkopf and van Aalten, 2004), and modelled in PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrodinger, LLC). The interaction of the AP with streptavidin (SA) (PDB: 3RY2) and a protein kinase (PDB: 1AQ1), was modelling using Coot (Emsley and Cowtan, 2004).

2.9.2 Affinity probe synthesis and analysis

The synthesis of the AP was by Michael Paradowski (University of Sussex). Staurosporine (STU) (LC Laboratories) and Fmoc (Fluorenylmethyloxycarbonyl

chloride) protected acid linker (Fmoc-NH-PEG_4-COOH) [PEG_4 (Polyethylene Glycol)] (Biomatrik) were treated with HBTU [*O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate] and DIFEA (*N*,*N*-diisopropylethyl amine) in DCM (dichloromethane), to give the required intermediate obtained in 63% yield after silica flash chromatography. Standard Fmoc de-protection of the intermediate at room temperature in the presence of piperidine led to the deprotected amine intermediate, which was obtained in 70% yield after silica flash chromatography. Acylation of the primary amine moiety intermediate was accomplished by treatment with NHS-Biotin (Acros Organics) in THF (tetrahydrofuran) and presence of Et3N (triethylamine), to give the final product (STU-PEG_4-Biotin) as a colourless oil, in 68% yield after silica flash chromatography. This was analysed by NMR using a VNMRS Varian 500 MHz instrument by Michael Paradowski (University of Sussex). The AP was provided as a yellow solid (21.653 mg). The AP was dissolved in 100% DMSO and furthered diluted with the required buffer.

2.10 Affinity probe protein pull-downs

Interaction between the AP and purified proteins, was shown using Dynabeads MyOne Streptavidin T1 (Dynabeads) (Invitrogen). Resuspended Dynabeads as a 50% slurry (~50 μl Dynabeads/tube) were washed twice with 100 μl dH_2O and twice with 100 μl TM buffer [MEK1 GF buffer with 0.05 % (v/v) Tween-20], by pelleting with a magnetic rack. The AP was prepared in AP binding buffer [20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 50% (v/v) DMSO, 2% (v/v) Tween-20], and 150 μl of 50 μM AP was added to Dynabeads, and incubated rotating overnight at room temperature. Subsequently, the beads were pelleted, the supernatant was removed, and the Dynabeads were washed four times with 200 μl DTM buffer [TM buffer with 2% (v/v) DMSO]. The AP-Dynabead were incubated with 200 μl of 25 μM MEK1-KD, Claspin, GWL and Chk2-KD in DTM buffer, rotating at 4 °C overnight. Subsequently the Dynabeads were pelleted, the supernatant removed, and the Dynabeads were washed four times with 200 μl DTM buffer. After the final wash step, the Dynabeads were resuspended in 10 μl DTM buffer,

mixed with 40 µl LDS-LB, and boiled at 95 °C for 5 minutes. The Dynabeads were pelleted and the supernatant was resolved by SDS-PAGE (section **2.5**).

2.11 ForteBio Octet

2.11.1 Assay programming

For assay development of the AP screen the ForteBio Octet RED96 (ForteBio) was used with streptavidin (SA) biosensors (ForteBio), using kinetics characterization software (ForteBio). Location of the biosensors, experimental steps, and timing were defined in the assay set-up software (ForteBio) prior to running any assay. The plate rotation speed was 1000 rpm, and all experiments were carried out at room temperature. Biosensors were hydrated in the AP assay buffer for 30 minutes prior to use. Experiments were prepared in a 96-well black plate (Greiner), with 200 µl sample volume in each well. Purified MEK1-KD, Kinase X and NF-kB p65 proteins, and STU were prepared in the appropriate assay buffer. For STU inhibition of MEK1-KD, 1 µM MEK1-KD protein was incubated with 1.2 µM STU for 15 minutes on ice. The precise conditions for each assay are described in the individual figure legends.

2.11.1.1 Octet cellular lysate preparation

Bacterial lysates were prepared from *E. coli* BL21 (DE3) cells transformed with the MEK1-KD or an empty expression vector in IPTG induced expression cultures (section **2.13**); were lysed by lysozyme lysis (section **2.12.13**), and clarified (section **2.15.2.1**). The supernatant in CDH lysis buffer supplemented with 2% (v/v) DMSO and 0.05% (v/v) Tween-20, was probed using an AP-biosensor on the Octet. For detergent screening, these lysates were supplemented with 2% (v/v) DMSO and varying concentrations of Tween-20. CDH lysates were prepared from *E. coli* BL21 (DE3) cells transformed with selected plasmid DNA [BRaf V600E, MEK1-KD and Chk2-KD, His_6-NIK (kinase active and dead) or GST-NIK (kinase active and dead), p50, GST, NF-κB p65, NF-κB p65 RHR and C/EBPβ), and expressed in 4 ml auto-induction expression cultures (section **2.12.12**). These were lysed by either lysozyme lysis in CDH lysis buffer (section **2.12.13**), or sonication lysis (section **2.15.2.1**) in CDH equilibration buffer, and clarified

(section **2.12.13**). Lysates were supplemented to a final concentration of 2% (v/v) DMSO and 0.05% (v/v) Tween-20.

2.11.1.2 Octet one-step purified proteins

For testing the Octet using one-step purified proteins, the proteins were expressed in auto-induction medium (section **2.11.1.1**) and purified in a 96-well plate format as in section **2.12.14**. For Strep- or GST-tagged proteins, these used 10 µl Neutravidin resin (Pierce) or 10 µl GS (Glutathione sepharose) resin (GE Healthcare) respectively. The elution for all proteins was in two 50 µl and a 110 µl steps, where the elution buffer for; Ni-NTA resin was CDH low elution buffer [50 mM HEPES pH 7.5, 100 mM NaCl, 250 mM imidazole, 0.5 mM TCEP, 10% (v/v) glycerol], for Neutravidin resin was Strep elution buffer [50 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 10% (v/v) glycerol, 2.5 mM desthiobiotin] and for GS resin was GST elution buffer [50 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 10% (v/v) glycerol, 25 mM L-Glutathione reduced]. All one-step purified proteins were supplemented with 2% (v/v) DMSO and 0.05% (v/v) Tween-20.

2.12 Combinatorial Domain Hunting

2.12.1 DNA recoding

The human CLSPN (Claspin) was recoded, to ensure optimal fragmentation, using recoding software Astuce (Domainex proprietary software); and: DNA translation using ExPASy Translate [ExPASy server; (Gasteiger et al., 2005)]; sequence alignment using ExPASy SIM [ExPASy server; (Gasteiger et al., 2005)]; identifying specific restriction endonuclease site using NEBcutter V2.0 (New England Biolabs; Vincze et al., 2003); determining DNA codon usage using EMBOSS (Rice et al., 2000); identifying bacterial promoter sequences using BPROM (SoftBerry); identifying RNA secondary structure using MFold (Zuker, 2003) and identifying difficult to synthesise sequences using GenePerfect (Gene Oracle).

2.12.2 Polymerase chain reaction for the incorporation of dUTP for CDH

This section is redacted for reasons of confidentiality.

2.12.3 S1 nuclease titration

This section is redacted for reasons of confidentiality.

Table 2.3 S1 nuclease titration.

This Table is redacted for reasons of confidentiality.

2.12.4 CDH dUTP DNA fragmentation reaction and purification

This section is redacted for reasons of confidentiality.

2.12.5 CDH fragmentation DNA electrophoresis, extraction and purification

This section is redacted for reasons of confidentiality.

2.12.6 CDH fragment end processing

This section is redacted for reasons of confidentiality.

2.12.7 TOPO cloning ligation and transformation of *E. coli* XL10 Gold cells

This section is redacted for reasons of confidentiality.

2.12.8 CDH colony polymerase chain reaction

This section is redacted for reasons of confidentiality.

2.12.9 CDH library transformation of *E. coli* BL21 (DE3)

This section is redacted for reasons of confidentiality.

2.12.10 Colony lifting and protein expression

This section is redacted for reasons of confidentiality.

2.12.11 Colony lysis and blotting

This section is redacted for reasons of confidentiality.

2.12.12 Colony selection and small-scale expression cultures

This section is redacted for reasons of confidentiality.

2.12.13 Harvesting and lysis of the small-scale expression cultures

This section is redacted for reasons of confidentiality.

2.12.14 96-well recombinant protein purification

This section is redacted for reasons of confidentiality.

2.13 Recombinant *E. coli* protein expression

2.13.1 Transformation of E. *coli* BL21 (DE3) competent cells

Competent *E. coli* BL21 DE3 cells (Novagen) were used for recombinant protein expression, performed as described in section **2.3**, but using 50-200 ng/μl (in 1-2 μl) of plasmid DNA and 12-25 μl *E. coli* BL21 DE3 cells.

2.13.2 Small-scale *E. coli* BL21 (DE3) recombinant protein expression

Small-scale expression tests of the Claspin protein fragments used 40 ml auto-induction medium were inoculated with a colony streak from a transformed L-agar

plate. Cultures were incubated and harvested as described in sections **2.13.4** and **2.12.12**.

2.13.3 Overnight pre-culture of transformed *E. coli* BL21 DE3 cells

Pre-cultures were prepared for large-scale expression culture growth. For each protein, a colony streak from a transformed L-agar plate was resuspended in 50-100 ml LB medium prior to incubation shaking at 37 °C overnight.

2.13.4 Large-scale *E. coli* BL21 (DE3) recombinant protein expression

Large-scale recombinant protein expression, used pre-cultures to inoculate between 1 and 4 litres of LB or TB medium, split 1 litre per 2 litre flask. Cultures were incubated shaking at 37 °C at 240 rpm until the cultures reached OD_{600} of 0.6-0.8 for LB or 1.5-1.8 for TB medium (where OD_{600} of 1.0 = 8 x 10^8 cells/ml), after which the temperature of the incubator was reduced to 20 °C to gradually cool the cultures for 1 hour. Subsequently, the cultures were induced with 1 mM IPTG for protein expression at 20 °C for between 16-18 hours. All *E. coli* expression cultures were harvested by centrifugation at 4-5000 rpm for 12 minutes at 4 °C, and the cell pellets were frozen at -20 °C.

2.14 *Sf*9 insect cell baculovirus expression

2.14.1 Bacmid transposition and purification

For expression in insect cells, the Chk1 constructs (section **2.2.1**) were transposed into a bacmid, using *E. coli* DH10 MultiBac competent cells (Invitrogen). Roughly, 1 μg (in 5 μl) of Chk1 plasmid DNA was incubated on ice for 30 minutes with 100 μl *E. coli* DH10 MultiBac cells thawed on ice. These were heat-shocked for 45 seconds at 42 °C, returned to ice for 2 minutes, and resuspended with 900 μl S.O.C. medium (no antibiotic). These were incubated shaking at 37 °C at 225 rpm for 4-6 hours. The cultures were diluted with S.O.C medium, and plated onto bacmid transposition plates [L-agar plates with 50 μg/ml kanamycin, 10 μg/ml tetracycline, 7 μg/ml gentamycin, 100 μg/ml Bluo-gal, 40 μg/ml IPTG], and incubated for 48 hours at 37 °C, followed by 24 hours at 4 °C; positive transposition was identified by blue-white screening. Individual white colonies,

were used to inoculate 4 ml bacmid transposition medium [LB medium with 50 µg/ml kanamycin, 10 µg/ml tetracycline, 7 µg/ml gentamycin], and incubated shaking at 37 °C at 250 rpm for up to 24 hours. The bacmid cultures were harvested by centrifugation at 4000 rpm for 10 minutes, and the bacmid DNA was purified (section **2.3.3**). Cell pellets were resuspended in 0.4 ml Buffer P1 and 0.4 ml Buffer P2 [both from QIAprep Spin MiniPrep kit], gently inverted and incubated at room temperature for 5 minutes. After, this was slowly mixed with 0.4 ml 3 M potassium acetate pH 5.5, and incubated on ice for 10 minutes. Subsequently, this was centrifuged at 13,000 rpm for 10 minutes at 4 °C. The supernatant was resuspended in 0.8 ml ice-cold isopropanol, gently inverted and incubated at -20 °C for 10 minutes. This was then centrifuged at 13,000 rpm for 15 minutes at room temperature and the supernatant was aspirated. The DNA pellet was washed in 0.7 ml 70% (v/v) ethanol and centrifuged at 13,000 rpm for 10 minutes, and the supernatant was aspirated. The DNA pellet was air-dried for 10 minutes, and dissolved in 40 µl dH_2O. Transposition was verified by colony PCR (section **2.3.2**). Transposition occurs by Tn7 elements in the plasmid containing the insert and the bacmid in the *E. coli* DH10MultiBac cell strain, this requires transposition proteins expressed on an additional plasmid in the *E. coli* cell strain. Transposition into the bacmid disrupts the *lacZ*α gene, and prevents expression of the *LacZ* fragment required for active β-galactosidase enzyme to hydrolyses Bluo-gal forming a blue product; therefore white colonies indicate a successful transposition. The bacmid contains the polyhedrin (polH) promoter sequence, which is an insect cell promoter sequence for cellular protein expression, for recombinant protein expression.

2.14.2 Insect cell transfection and baculovirus amplification

*Sf*9 cells were selected for Chk1 recombinant protein expression. *Sf*9 is a clonal derivative from the *Sf*21 cell line that was originally isolated from pupal ovarian tissue of the fall army worm (O'reilly and Miller, 1992, Vaughn et al., 1977). *Sf*9 cell transfection involves incubating the *Sf*9 cells with the recombinant bacmid DNA, for the production of recombinant baculovirus particles containing the recombinant bacmid DNA; used for subsequent *Sf*9 cell infection. The protocols for Sf9 cell transfection and baculovirus amplification are as described in Manual,

(2002). During baculovirus amplification, the cell pellet was harvested and analysed by western blotting (section **2.6**). The viral titre [plaque-forming units per ml (pfu/ml)] of the Passage (P)3 baculovirus for each construct was determined by a plaque assay, as described in Manual, (2002). The highest P3 viral titres for Chk1 were obtained for P3 baculoviruses harvested between 70 and 75% cellular viability.

2.14.3 *Sf*9 recombinant protein expression and harvesting

Sf9 recombinant protein expression followed the protocol described in Manual, (2000). Typically, between 2 and 9 litres of *Sf*9 culture (between 1.8 x 10^5 and 2.4 x 10^6 cell/ml) in *Sf*-900 II SFM medium (pen/strep) (Invitrogen), was used for recombinant protein expression split as 500 ml in 2 litre roller bottles. Infected *Sf*9 cultures were incubated shaking at 28 °C, at 150 rpm for 62 hours or until the cellular viability dropped to 92 to 95% viability. Cultures were harvested by centrifugation at 1200 rpm for 10 minutes at 4 °C. Cell pellets were flash frozen in liquid nitrogen and stored at -80 °C.

2.15 Recombinant protein purification

During lysis and all purification steps, the proteins and buffers were all incubated at 4 °C or on ice throughout. Fractions throughout all of these purification steps were resolved by SDS-PAGE (section **2.5**).

2.15.1 ÄKTAFPLC purification

An ÄKTAFPLC (Fast Protein Liquid Chromatography) system (GE Healthcare), controlled using UNICORN system control software (GE Healthcare), was used to run 1 or 5 ml HiTrap Q FF columns (GE Healthcare), 5 ml HiTrap Heparin HP columns (GE Healthcare) and Size Exclusion Chromatography (SEC) columns (all Superdex (SD) columns; GE Healthcare), at 4 °C. The system was flushed with the appropriate buffer prior to the equilibration of the columns. The eluted volume was collected by fractionation, and the A_{280} chromatograph enabled identification of fractions containing protein.

2.15.2 Purification of Claspin protein fragments

2.15.2.1 *E. coli* cellular lysis and clarification

For the lysis of *E. coli* pellets, the frozen pellets were defrosted under running water, and the pellet was resuspended in roughly 40 ml CDH equilibration buffer per 10 g cell paste, with an EDTA free protease inhibitor tablet. These were sonicated on ice, in 5 second pulses at 30-40% amplitude, interspersed with 5 seconds rest, until cells were disrupted. Cellular lysates were clarified by centrifugation at 20,000 rpm for 40 minutes at 4 °C. The soluble supernatant fraction was harvested from the insoluble fraction and filtered using a Millex-SV 5.0 µm filter (Millipore) prior to further purification.

2.15.2.2 IMAC purification for all Claspin fragments

The Claspin supernatant was firstly purified using in batch format using Talon-IMAC resin (Clontech). Talon resin was pre-equilibrated in five column volumes (CV) of CDH equilibration buffer. The supernatant was incubated with 1-8 ml Talon resin (dependent on the culture volume), by rotation for 40 minutes at 4 °C. The unbound volume was allowed to flow-through by gravity flow, and the resin was washed with 10 CV of CDH equilibration buffer to remove all unbound material. The proteins were eluted by resuspending the resin in two CV of CDH elution buffer and collecting the flow-through, repeated six times.

2.15.2.3 IEX purification for C-terminal fragments

A second purification for the C-terminal Claspin fragments was by Ion Exchange (IEX) chromatography. For IEX purification, IMAC-eluted proteins were diluted by 50% in NaCl free buffer (50 mM HEPES pH 7.5, 0.5 mM TCEP), and samples were loaded onto either a 1 or 5 ml HiTrap Q FF column, pre-equilibrated in IEX A buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 0.5 mM TCEP). The column was washed with two CV of IEX A buffer and the proteins were purified by ionic charge, by elution over a salt gradient from 0% to 100% IEX B buffer (20 mM HEPES pH 7.5, 2 M NaCl, 0.5 mM TCEP) over 20 CV.

2.15.2.4 Heparin chromatography purification for N-terminal fragments

A second purification for all N-terminal Claspin fragments was by heparin chromatography. For Heparin purification, IMAC-eluted proteins were loaded

directly onto two 5 ml Heparin columns, pre-equilibrated in Heparin A buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP); if the specific fragment displayed weak binding to Heparin these were diluted by 50% in NaCl free buffer prior to loading. The column was washed with two CV of Heparin A buffer, and the proteins were purified by ionic charge, by elution over a salt gradient from 0% to 30-80% IEX B buffer over 20 CV.

2.15.2.5 SEC purification for Claspin fragments

The final purification step for all Claspin protein fragments was by SEC. The fractions containing recombinant protein from the previous purification step were pooled, and were concentrated using VivaSpin concentrators if required (section **2.15.4**). C-terminal protein fragments (8-10 ml) were resolved on a 26/60 SD75 column, and N-terminal protein fragments (3-5 ml) were resolved on a 16/60 SD200 column; both pre-equilibrated in Claspin GF buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP). N-terminal Claspin fragments used for CD, AUC, bioSAXS, and NMR, were purified in Claspin GF low buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP).

2.15.3 Purification of *Sf*9 expressed Chk1

2.15.3.1 *Sf*9 cellular lysis

For the lysis of *Sf*9 cell pellets, SC equilibration buffer [50 mM HEPES pH 7.5, 10 mM Imidazole, 1 M NaCl, 0.5 mM TCEP, 10% (v/v) Glycerol] with EDTA free protease inhibitor tablet and 250 U Benzonase, was added to 1-2 litre frozen cell paste (~5 g), to a total of 45 ml, and were defrosted under H_2O. The cell paste was resuspended and the cells were lysed using a pestle homogeniser (Sigma), with 8-10 repeats, prior to lysate clarification by centrifugation (section **2.15.2.1**).

2.15.3.2 IMAC purification of Chk1 proteins

Clarified Chk1 protein cell lysates were purified using Talon-IMAC, described in section **2.15.2.2**, but using SC equilibration buffer for Talon resin equilibration and wash steps, and SC elution buffer [50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole, 0.5 mM TCEP, 10% (v/v) Glycerol] for protein elution.

2.15.3.3 Size Exclusion Chromatography for Chk1 proteins

The final purification step for the Chk1 proteins was by SEC. Proteins were prepared as described in section **2.15.2.5**; 3-5 ml for a 16/60 SD75 column or 8-10 ml for a 26/60 SD75 column; pre-equilibrated in Chk1 GF buffer [20 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM TCEP, 10% (v/v) glycerol].

2.15.4 Protein concentration

Protein elution fractions containing the recombinant protein were pooled and concentrated using VivaSpin concentrators (GE Healthcare). Concentrators of an appropriate molecular weight (MW) membrane cut off (either 10 kDa or 30 kDa) and volume (0.5, 6 or 20 ml) were selected. The VivaSpin concentrators were equilibrated in protein specific buffer by brief centrifugation. For protein concentration these were centrifuged at 4 °C, and were regularly resuspended.

2.15.5 Purified proteins

Chk2-KD was purified as described in (Oliver et al., 2006). MEK1-KD was purified as described in (Meier et al., 2012) in a final buffer of MEK1 GF buffer. Chk2-KD and MEK1-KD were both flash frozen in aliquots in liquid nitrogen and stored at -80 °C. NF-kB p65, Kinase X and MEK1-KD (for ITC; section **2.8**) were gifted by Dr Stefanie Reich (Domainex), and GWL kinase was gifted by Dr Mohan Rajasekaran (University of Sussex). The N-terminal Claspin protein fragments were either immediately used or were stored at 4 °C. For bioSAXS, concentrated N-terminal Claspin protein fragments were flash frozen in liquid nitrogen and stored at -80 °C. *Sf*9 expressed Chk1-KD was used immediately for biochemical assays or for protein crystallography. For use of frozen proteins, these were rapidly defrosted and immediately incubated on ice. For use of purified proteins these were hard spun by brief centrifugation to pellet any insoluble material and the supernatant was carefully aspirated. For sensitive techniques, the purified proteins were filtered with brief centrifugation using Pierce Spin Cups Cellulose Acetate Filter (Pierce), which were pre-equilibrated in the purification buffer. The protein concentration was accurately determined prior to sample analysis (section **2.4**).

2.15.6 Edman degradation

Edman degradation (N-terminal sequencing) was used to identify the N-terminus of specific proteins identified by SDS-PAGE. Gel samples were electro-blotted onto PVDF membrane (section **2.6**), and stained with Ponceau S [0.1% (w/v) Ponceau S in 5% (v/v) acetic acid] until bands were visible, then destained in dH_2O. The bands of interest were analysed by Edman sequence analysis by either Protein and Nucleic Acid Chemistry Facility (PNAC; Cambridge) or AltaBiosciences (Birmingham). Edman analysis uses Edman reagent (phenylisothiocyanate), which in alkaline conditions reacts with the amine group of the N-terminal residue resulting in a phenylthiocarbamyl-amino acid (Edman, 1960). The cleaved amino acid is identified by high-performance liquid chromatography (HPLC).

2.15.7 Mass spectrometry

Mass spectrometry was used to identify breakdown protein products identified by SDS-PAGE. In-gel tryptic digest reactions were performed as described by Zee and Garcia, (2013). The peptides were analysed using LTQ-OrbitrapXL (Thermo Scientific) by Dr Steve Sweet (University of Sussex), with peptide analysis on the Mascot server (Perkins et al., 1999) by Abubakar Hatimy (University of Sussex).

2.16 DNA oligonucleotides

All DNA oligonucleotides used in DNA binding assays were purchased from MWG Biotech; DNA oligonucleotides are listed in **Table 2.4** and the nomenclature is listed in **Table 2.5** These were dissolved in 1x oligo annealing buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$), and the DNA concentration was determined (section **2.4**). For dsDNA oligonucleotides, complementary oligonucleotides were mixed with a 1:1.2 ratio for labelled:unlabelled DNA oligonucleotides, or a 1:1 ratio for two unlabelled DNA oligonucleotides. The DNA oligonucleotides were annealed by; incubation at 95 °C for 5 minutes, 70 °C for 10 minutes and temperature reduction by 0.02 °C s^{-1}, until at 10 °C.

DNA Oligonucleotide	Modification	Sequence
DS1_35_Cy5	5’-Cy5	CATGGTCTCACGTGTAGAGGGTAAACTTGGTCAGC
DS2_35		GCTGACCAAGTTTACCCTCTACACGTGAGACCATG
DS3_30_Cy5	5’-Cy5	CATGGTCTCACGTGTAGAGGGTAAACTTGG
DS4_30		CCAAGTTTACCCTCTACACGTGAGACCATG
DS5_25_Cy5	5’-Cy5	CATGGTCTCACGTGTAGAGGGTAAA
DS6_25		TTTACCCTCTACACGTGAGACCATG
DS7_20_Cy5	5’-Cy5	CATGGTCTCACGTGTAGAGG
DS8_20		CCTCTACACGTGAGACCATG
DS3_30_FITC	5’-FITC	CATGGTCTCACGTGTAGAGGGTAAACTTGG
DS4_30		CCAAGTTTACCCTCTACACGTGAGACCATG
DS7_20_FITC	5’-FITC	CATGGTCTCACGTGTAGAGG
DS8_20		CCTCTACACGTGAGACCATG
DS33_19_FITC	5’-FITC	CATGGTCTCACGTGTAGAG
DS37_19		CTCTACACGTGAGACCATG
DS34_18_FITC	5’-FITC	CATGGTCTCACGTGTAGA
DS38_18		TCTACACGTGAGACCATG
DS35_17_FITC	5’-FITC	CATGGTCTCACGTGTAG
DS39_17		CTACACGTGAGACCATG
DS36_16_FITC	5’-FITC	CATGGTCTCACGTGTA
DS40_16		TACACGTGAGACCATG
DS9_15_FITC	5’-FITC	CATGGTCTCACGTGT
DS10_15		ACACGTGAGACCATG
DS11_10_FITC	5’-FITC	CATGGTCTCA
DS12_10		TGAGACCATG
DS3_30_C		CATGGTCTCACGTGTAGAGGGTAAACTTGG
DS4_30_C		CCAAGTTTACCCTCTACACGTGAGACCATG
DS7_20_C(T)		CATGGTCTCACGTGTAGAGG
DS8_20_C(T)		CCTCTACACGTGAGACCATG
DS9_15_C		CATGGTCTCACGTGT
DS10_15_C		ACACGTGAGACCATG
DS11_10_C		CATGGTCTCA
DS12_10_C		TGAGACCATG
DS7_20C		CATGGTCTCACGTGTAGAGG
DS8_20C		CCTCTACACGTGAGACCATG
CR1_20C		CATGGTCTCACGTGTAGAGGCC
CR2_20C		CCTCTACACGTGAGACCATGGG

Table 2.4 DNA oligonucleotides.
Where Cy5 and FITC represents cyanine-5 and fluorescein isothiocyanate respectively.

Assay	Nomenclature	Oligonucleotide 1	Oligonucleotide 2
EMSA	35 bp dsDNA	DS1_35_Cy5	DS2_35
	30 bp dsDNA	DS3_30_Cy5	DS4_30
	25 bp dsDNA	DS5_25_Cy5	DS6_25
	20 bp dsDNA	DS7_20_Cy5	DS8_20
	35 nt ssDNA	DS1_35_Cy5	-
	30 nt ssDNA	DS3_30_Cy5	-
	25 nt ssDNA	DS5_25_Cy5	-
	20 nt ssDNA	DS7_20_Cy5	-
FP	40 bp dsDNA	AA1_40_FITC	AA2_40
	30 bp dsDNA	DS3_30_FITC	DS4_30
	20 bp dsDNA	DS7_20_FITC	DS8_20
	19 bp dsDNA	DS33_19_FITC	DS37_19
	18 bp dsDNA	DS34_18_FITC	DS38_18
	17 bp dsDNA	DS35_17_FITC	DS39_17
	16 bp dsDNA	DS36_16_FITC	DS40_16
	15 bp dsDNA	DS9_15_FITC	DS10_15
	10 bp dsDNA	DS11_10_FITC	DS12_10
	40 nt ssDNA	AA1_40_FITC	-
	30 nt ssDNA	DS3_30_FITC	-
	20 nt ssDNA	DS7_20_FITC	-
	19 nt ssDNA	DS33_19_FITC	-
	18 nt ssDNA	DS34_18_FITC	-
	17 nt ssDNA	DS35_17_FITC	-
	16 nt ssDNA	DS36_16_FITC	-
	15 nt ssDNA	DS9_15_FITC	-
	10 nt ssDNA	DS11_10_FITC	-
ASEC and tryptic digests	30 bp dsDNA	DS3_30_C	DS4_30_C
	20 bp dsDNA	DS7_20_C(T)	DS8_20_C(T)
	15 bp dsDNA	DS9_15_C	DS10_15_C
	10 bp dsDNA	DS11_10_C	DS12_10_C
Crystallography	20 bp dsDNA	DS7_20C	DS8_20C
	20 bp dsDNA + 2 nt ssDNA	CR1_20C	CR2_20C

Table 2.5 DNA oligonucleotide nomenclature.
Composition of DNA oligonucleotide used for assessment of Claspin DNA binding functionality, showing nomenclature breakdown by assay. See **Table 2.4** for the individual oligonucleotide sequences.

2.17 Claspin Peptides

All peptides were purchased from Peptide Protein Research Ltd, and were dissolved in peptide buffer [10 mM HEPES pH 7.5, 0.5 mM TCEP]; refer to **Table 2.6**. For preparation with purified proteins, a dilution of PhosStop phosphatase inhibitor tablet (Roche) was included to prevent peptide dephosphorylation.

Peptide	Abbreviation	Sequence	Assay			
			FP	Thermal Denaturation	ADP-Glo	Protein Crystallography
Flu-Claspin-T916p	(Flu-T916p)	Flu-GGMDELLDLC[pT]GKFTSQ	Y	-	-	Y
Flu-Claspin-S945p	(Flu-S945p)	Flu-GGMEELLNLC[pS]GKFTSQ	Y	Y	Y	Y
Flu-Claspin-S982p	(Flu-982p)	Flu-GGMEEALALC[pS]GSFPTD	Y	-	-	-
Flu-Claspin-T916	(Flu-T916)	Flu-GGMDELLDLCTGKFTSQ	Y	-	-	-
Flu-Claspin-S945	(Flu-S945)	Flu-GGMEELLNLCSGKFTSQ	Y	-	Y	-
Flu-Claspin-S982	(Flu-S982)	Flu-GGMEEALALCSGSFPTD	Y	-	-	-
Flu-Claspin-M937-	(Flu-M937-)	Flu-GGEELLNLC[pS]GKFTSQ	Y	-	-	-
Flu-Claspin-L941G	(Flu-L941G)	Flu-GGEELGNLC[pS]GKFTSQ	Y	-	-	-
Flu-Claspin-L943G	(Flu-L943G)	Flu-GGEELLNGC[pS]GKFTSQ	Y	-	-	-
Flu-Claspin-C944A	(Flu-C944A)	Flu-GGEELLNLA[pS]GKFTSQ	Y	-	-	-
Flu-Claspin-F948A	(Flu-F948A)	Flu-GGEELLNLC[pS]GKATSQ	Y	-	-	-
Claspin-T916p	(T916p)	MDELLDLC[pT]GKFTSQ	-	-	-	Y
Claspin-S945p	(S945p)	MEELLNLC[pS]GKFTSQ	-	-	-	Y

Table 2.6 Claspin CKB peptides.
Where Flu represents fluorescein, and [pT] and [pS] represent a phospho-threonine and a phospho-serine respectively. Y=used in the specified assay.

2.18 Analytical Size Exclusion Chromatography

2.18.1 Calibration of the 10/300 SD 200 Increase column

Calibration of the ASEC column enables the accurate determination of the molecular mass of proteins and complexes, by extrapolation for the elution volume from the calibration curve. The 10/300 SD200 Increase column was calibrated using the molecular mass standard from the low and high molecular weight gel filtration calibration kits (GE Healthcare); prepared with 1 mg/ml thyroglobulin, 1 mg/ml ferritin, 1 mg/ml aldolase, 1 mg/ml conalbumin, 1 mg/ml ovalbumin and 1 mg/ml ribonuclease A, to a final volume of 100 µl in Calibration buffer (20 mM HEPES pH 7.5, 150 mM NaCl). The calibration curve was determined by plotting the partition coefficient (K_{av}) against the logarithm of the molecular mass. K_{av} was calculated for each protein using **Equation 2.1** using the elution volume for each

protein (V_e), the column void volume (determined from the start of void peak) (V_o) and the geometric column volume (V_c). The molecular mass for the Claspin protein fragments was calculated by using the linear regression equation on the calibration curve (**Figure 2.1**).

$$K_{av} = \frac{V_e - V_o}{V_c - V_o}$$

Equation 2.1

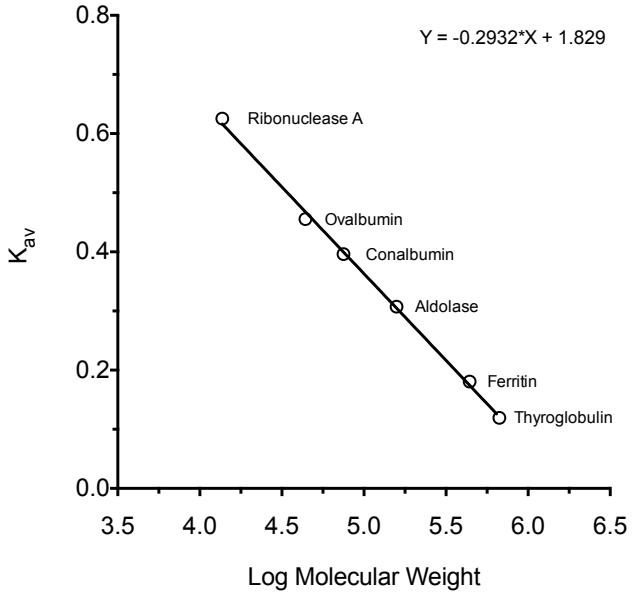


Figure 2.1 10/300 SD200 Increase column calibration curve.
Calibration curve generated using calculated values for K_{av} and log molecular mass for the molecular mass standards (as indicated). The equation displayed was used for the calculation of the molecular mass of eluted unknown proteins.

2.18.2 ASEC for the Claspin protein fragments

ASEC was used to investigate the molecular mass of the Claspin protein fragments and their interaction with DNA. Claspin protein samples of 5 and / or 10 mg/ml with a sample loading volume of 100 µl, were resolved on a 10/300 SD200 Increase column, pre-equilibrated in Claspin GF buffer. The molecular mass was estimated using the calibration curve for the column. For assessment of DNA binding functionality, Claspin protein fragments at 5 mg/ml were mixed with a

1:1.2 (protein:DNA) ratio with either 10, 15, 20 or 30 bp dsDNA (section **2.16**), in a final volume of 100 µl, and incubated on ice for 30 minutes prior to resolution by ASEC. Selected fractions were resolved using both SDS-PAGE (section **2.5**) and Native-PAGE (section **2.26.1**).

2.19 Circular dichroism

CD is a spectroscopic technique, which enables the analysis of the secondary structure within a protein. CD analysis was performed for Claspin protein fragments in Claspin GF low buffer at concentrations between 30 and 150 µM. CD spectra (for proteins and their respective buffers) were measured using a JASCO J-715 spectropolarimeter (JASCO), using JASCO spectra manager software (JASCO). The temperature was controlled using a JASCO PTC-384W peltier temperature control system (JASCO) at 20 °C. Samples were analysed in a 0.1 mm quartz cuvette (Starna). For sample analysis normal sensitivity was used with a bandwidth of 1 nm, a response time of 4 seconds and with a continuous scan speed of 50 nm/minute. Three scans per sample were recorded from 260 nm to 180 nm with a data pitch of 0.1 nm, which were automatically averaged and the buffer baseline was subtracted from the scans. Data was eliminated when the High Tension Voltage (HTV) was above 600 mV, a measure of the signal-to-noise ratio at which signals become unreliable. For CD thermal denaturation, 90 µM protein samples were incubated in the sample holder of the spectropolarimeter at 80 °C for 3 minutes, controlled by the peltier temperature control system, and then CD spectra were recorded. For CD measurement inducing the co-solvent 2,2,2-trifluoroethanol (TFE), 30 µM protein samples were incubated with 40% (v/v) TFE for 3 minutes prior to loading into the cuvette and CD spectra measurement. All CD spectra are reported in mean residue ellipticity to enable CD spectra comparison.

2.19.1 Circular dichroism spectra deconvolution

Estimation of the secondary structure content in proteins can be obtained through CD spectra deconvolution using the DichroWeb server (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008). This used the CDSSTR algorithm (Sreerama et al., 2000a) and protein reference data sets 3, 4, 6, 7, SP175 and SMP180 (Sreerama et al., 2000a, Sreerama et al, 2000b). Secondary structure content estimation was

the average percentage from deconvolution using each protein reference data set, when the Normalized Root Mean Squared Deviation (NRMSD) was less than 0.1 (Mao et al., 1982).

2.20 ThermoFluor

ThermoFluor was used to investigate the stability of recombinant proteins, using SYPRO orange dye 5000x concentrate (Sigma) with the LightCycler 480 (Roche), in 96-well plates (Roche). SYPRO orange is a fluorescent dye that interacts with the hydrophobic sites of a protein, which are exposed during protein unfolding. This results in a significant increase in the fluorescence signal of the dye, but the fluorescence signal is rapidly quenched with water. The manufacturer does not reveal the exact concentration of the SYPRO orange dye solution. SYPRO orange dye was diluted to a 20x stock in the protein sample buffer, where 15 μl was gently resuspended with 45 μl protein samples at 1, 2.5 and 5 μM in concentration. The final concentrations were 0.75, 1.9 and 3.75 μM in 5x SYPRO orange. Triplicate 20 μl samples were dispensed into a 96-well plate (including controls with no protein), and the plate was sealed with ThermalSeal RTS sealing film (Roche). Thermal denaturation used the LightCycler 480, the excitation and emission wavelengths were set at 465 and 580 nm respectively, and the temperature was increased from 20 to 80 °C by 0.03 °C s^{-1}, acquiring 20 data points per second. The melting temperature (*T*m) was determined by the minima of the negative peak.

2.20.1 Chk1-KD thermal shift optimization screening

Optimal buffers and/or additives can enhance the thermal stability of a protein, and these can be identified using thermal denaturation. Chk1-$KD^{1-270/289}$-His proteins at 1.9 μM, were incubated with 100 mM additive buffers [MES pH 6.5, HEPES pH 7.0 and pH 7.5 and Tris pH 7.5, pH 8.0 and pH 8.5 (all Molecular Dimensions)], or additives including; ADP (Sigma), STU (Sigma), EDTA or a Claspin peptide (Flu-S945p) (section **2.17**) at ratios of 1:1.2, 1:5, 1:10 and/or 1:20 (protein:additive). Samples were incubated on ice for 10 minutes prior to analysis.

2.21 Chemical crosslinking

Chemical crosslinking results in specific amino acids in close proximity become chemically linked, which is irreversible by boiling or SDS denaturation. Using the homobifunctional chemical crosslinker BS3 (Bis(sulfosuccinimidyl)suberate) (Pierce), crosslinking occurs between primary amine side groups of lysine or the N-terminal amino acid. BS3 crosslinker was dissolved in dH_2O and diluted with Claspin GF buffer. Crosslinking reactions contained 30 or 60 μM Claspin protein fragments with 0.5, 5 and 50 μM, and 0.5, 5 and 50 mM BS3 crosslinker, in a final volume of 20 μl. Reactions were incubated at room temperature for 60 minutes, and quenched with 2 μl 2 M Tris pH 8.0 with incubation for 15 minutes at room temperature. Reactions were analysed by SDS-PAGE (section **2.5**).

2.22 Analytical ultracentrifugation

AUC is an optical centrifugation technique, which enables the analysis of macromolecules and macromolecular interactions in their native states in solution. This enables the analysis of hydrodynamic and thermodynamic properties of macromolecules in solution. Sedimentation Velocity (SV)-AUC was preformed using the Beckman Optima XL-A analytical ultracentrifuge (Beckman Coulter) equipped with an absorption (ABS) optical scanner (Beckman Coulter), and a An50 Ti analytical rotor (Beckman Coulter); with the assistance of Dr John McGeeham and Dr Anna Swiderska (University of Portsmouth). 400 μl of purified Claspin protein fragments (0.6 mg/ml B2D9 and 0.3 mg/ml A1G12 and B2C2), and 425 μl respective buffer samples were loaded into the appropriate sections of a 12 mm optical path length double-sector cell (Beckman Coulter) sealed with 140 inch-pounds of torque. Samples were centrifuged under vacuum at 23 °C, with scans recorded every 10 minutes. B2D9 was centrifuged at 15,000 rpm and scans were recorded at 280 nm. A1G12 and B2C2 were centrifuged at 35,000 rpm and scans were recorded at 230 nm. Buffer subtractions were automatically calculated and the raw data was analysed using $DCDT^{+2}$ (Philo, 2006).

2.23 Biological Small Angle X-ray Scattering

bioSAXS is an in solution X-ray diffraction technique, of which scattering data can be used to study a number of parameters of a protein, and data can be used to generate low-resolution three-dimensional bead models of the protein. Scattering data from Claspin protein fragments was collected on the BM29 beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, allocated on the MX-1603 application. Purified Claspin protein fragments were concentrated slowly using VivaSpin concentrators (section **2.15.4**) with regular inversion, with 100 µl aliquots taken during concentration at roughly 1, 2.5, 5, 7.5 and / or 10 mg/ml. Protein samples were flash frozen in liquid nitrogen and stored at -80 °C, and subsequently prepared as in section **2.15.5**. For bioSAXS, protein and respective buffer samples were loaded into the automated sample changer, and 20 µl of each sample was aspirated, and under continuous flow irradiated using 100% transmission of the X-ray beam collecting 10 frames, in a total duration of 15 seconds, at 4 °C. Automatic beamline processing (ESRF) averaged the 10 frames and scattering from the buffer was subtracted. Data was presented using PRIMUS software (Konarev et al., 2006) using the Kratky and Guinier plot functions.

2.24 Nuclear Magnetic Resonance

NMR can be used to determine the in solution structure of a protein. For analysis by NMR proteins must be chemically labelled with an isotopic label. A1G12 and B2C2 transformed *E. coli* strain BL21 (DE3) (section **2.13.1**) were grown in auto-induction NMR medium-^{15}N (Novagen). A 50 ml pre-culture (section **2.13.3**) was grown in NMR NI medium at 30 °C overnight, and 40 ml was used to inoculate 1 litre NMR auto-induction medium split in two 2 litre flasks. Cultures were incubated shaking at 37 °C, at 280 rpm for 5 hours. Subsequently the temperature was reduced to 25 °C, and cultures were harvested after 20 hours by centrifugation. Cellular harvesting, lysis and protein purification were as described in section **2.15.2**. Concentrated B2C2 at 260 µM, was resuspended with 10% (v/v) deuterium oxide (Sigma) prior to analysis. NMR spectra data was collected for ^{15}N-protein using the Varian VNMRS 600 MHz (Thermo Scientific) at 600 MHz for 3

hours at 20 °C. Data collection and analysis ($^{1}H–^{15}N$ HSQC spectrum) was performed by Dr Iain Day (University of Sussex).

2.25 Tryptic digest

Tryptic digest analysis of proteins aims to identify globular regions of proteins, through the identification of trypsin insensitive fragments. Tryptic analysis used a final concentration of 10 μM Claspin protein fragments, resuspended with trypsin (prepared in acidic buffer) (Promega) at a final concentration between 0.01 and 10 nM, in a total volume of 35 μl. Samples were incubated at room temperature and 10 μl was removed from each reaction after 5, 10 and 15 minutes, and resuspended with LDS-LB. Samples were analysed by SDS-PAGE (section **2.5**). For tryptic digests with DNA, proteins A1G12 and B2C2 (10 μM) were mixed with 20 nt / bp of ss- or dsDNA (section **2.16**), with a 1.2:1 ratio (DNA:protein), and incubated at room temperature for 20 minutes, prior to tryptic digest.

2.26 Electrophoretic Mobility Shift Assay

EMSAs can be used to identify an interaction between protein and DNA and the stoichiometry of the interaction. EMSA binding reactions were prepared with final concentrations of 2 μM purified Claspin protein fragments and 0.2 μM 20, 25, 30 and 35 nt / bp ss- or dsDNA oligonucleotides (section **2.16**), prepared in 1x EMSA binding buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$) to a total volume of 10 μl. These were incubated at room temperature for 20 minutes, prior to resolution by native polyacrylamide gel electrophoresis (Native-PAGE).

2.26.1 Native Polyacrylamide Gel Electrophoresis

Native-PAGE was used to resolve protein-DNA samples, in non-denaturing conditions, using cast 10-well 1 mm 5% (v/v) acrylamide gels. Native-PAGE gels were cast using glass plates (Bio-Rad), resolving gel mix [5% (v/v) 30% acrylamide/bis solution (29:1) (Bio-Rad), 0.5x TBE buffer (8.9 mM Tris pH 8.3, 8.9 mM boric acid, 2 mM EDTA), 5% (v/v) glycerol, 0.067% (v/v) APS (ammonium persulphate), 0.067% (v/v) TEMED (N,N,N'N'-tetramethylethylenediamine)], using a 10-well 1 mm comb (BioRad) to create wells. When set, Native-PAGE gels were

pre-run for 60 minutes at 25 mA at 4 °C, in 0.5x TBE buffer, using a Mini-PROTEAN Tetra cell system (Bio-Rad). Wells were flushed with 0.5x TBE buffer, and 4x loading buffer [50 mM HEPES pH 7.5, 0.2 % (w/v) bromophenol blue, 40% (w/v) sucrose] was loaded into one lane of the gel for dye front detection. Samples were prepared resuspended in EMSA loading buffer [20 mM HEPES pH 7.5, 50% (v/v) glycerol], at a 5:1 ratio (sample:buffer). Native-PAGE gels were resolved in 0.5x TBE running buffer at 12 mA at 4 °C until the dye front had resolved over 50% of the gel (~90 minutes). EMSAs using Cy5 fluorescent oligonucleotides were imaged using the FLA-5100 fluorescent imager (FujiFilm), using the 670 nm wavelength at 600 V (imaging the Cy5 emission spectra). Native-PAGE gels of ASEC samples (section **2.18**) were visualized by ethidium bromide (EtBr) (Fisher Scientific) staining. Gels were removed from the glass casing and placed in 0.5x TBE buffer containing 50 μg/ml EtBr, and incubated rocking for 15 minutes. Gels were rinsed in 0.5% TBE buffer prior to visualization of the DNA bands (section **2.2.4**).

2.27 Fluorescence Polarization

FP was used to investigate the interaction between purified proteins and a fluorescently labelled binding partner peptide or DNA. Two-fold serial dilutions of Claspin protein fragments were prepared in Claspin GF buffer and incubated with 100 nM FITC-labelled ss- and dsDNA of varying lengths (section **2.16**). Two-fold serial dilutions of $Chk1^{1-270/289}$-His and Chk2-KD proteins were prepared in Chk1 GF buffer and incubated with 100 nM Flu-labelled CKBD peptides (section **2.17**). Chk1-KD was incubated with STU at a 1:1.2 ratio (protein:STU), for 10 minutes prior to preparation of the serial dilutions. All FP samples were incubated for 15 minutes at room temperature prior to analysis. Samples of 50 μl were transferred in triplicate to black 96-well plates (Thermo Scientific, Nunc). The plate fluorescence was measured using the excitation and emission wavelengths of 485 nm and 520 nm respectively, using the POLARstar Omega multimode microplate reader (BMG Labtech) with the Fluorescence Polarization optic set (BMG Labtech). For each well 50 flashes were collected and averaged. Triplicate readings were averaged and the background fluorescence was subtracted. The FP data was plotted using GraphPad Prism V6.0f for Mac OS X, using non-linear fitting with a one site, specific binding with Hill slope.

2.28 Chk1 activity ADP-Glo assay

To determine the kinase activity of recombinant Chk1 protein kinase the Chk1 ADP-Glo assay (Promega) was used in a white 384-well plate (Corning), using full-length Chk1 (Promega) and Chk1-KD$^{1-289}$-His proteins. The ADP-Glo assay was prepared as described in the ADP-Glo Kinase Assay Technical Manual (Promega, 2015); using 50 μM ATP (Promega) and a Chktide substrate (Promega) for phosphorylation. Chk1 protein kinases (12.5 ng/μl full-length Chk1 or 6.25 ng/μl Chk1-KD$^{1-289}$-His) were incubated with Flu-S945p or Flu-S945 CKB peptides (section **2.17**) at a 2:1 ratio (peptide:protein), and incubated on ice for 20 minutes prior to preparation of 2-fold serial dilutions. Luciferase luminescence was detected using a PHERAstar FS (BMG Labtech) using the luminescence optic (BMGLabtech). For each well 50 flashes were collected and averaged. Triplicate results were averaged, background luminescence was subtracted, and data plotted using GraphPad Prism V6.0f for Mac OS X. A standard curve representing ATP-to-ADP turnover was performed to ensure the assay was working within the linear range (i.e. signal not saturated at 100% ADP) (**Figure 2.2**).

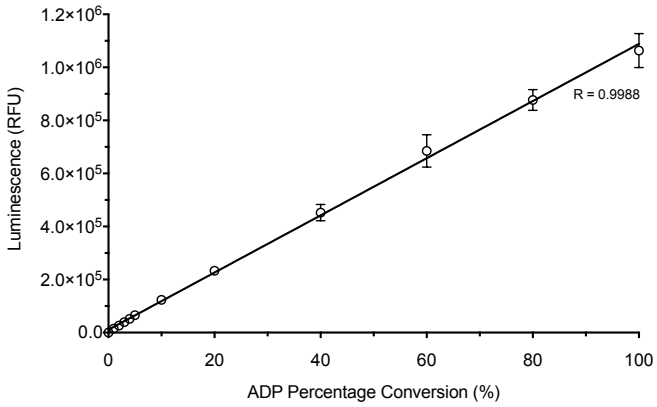


Figure 2.2 ADP-Glo conversion curve.
Linear regression analysis for the ATP-to-ADP conversion curve, for the percentage conversion of ATP-to-ADP, prepared in accordance with the ADP-Glo Kinase Assay Technical Handbook (Promega, 2015). Luciferase luminescence was measured using a PHERAstar FS with a luminescence optic. Each data point is the mean of three individual experiments and error bars represent one standard deviation.

2.29 Protein crystallography

2.29.1 Sparse matrix screening

For protein crystallization screening, purified proteins (section **2.15.5**) were used for sparse matrix crystallisation screening. Crystallization screens were set up as a 1:1 drop ratio (protein to well solution; final drop volume of 0.4 μl) in sitting-drop MRC 2-well crystallization plates (Swissci; Hampton Research), with a reservoir volume of 50 μl, using the ARI Crystal Phoenix Liquid Handling System (Art Robbins Instruments) or the Oryx8 (Douglas Instruments). Crystal screening plates were sealed using ClearVue sheets, and incubated at either 4, 14 or 20 °C. Plates were inspected 3-5 days after being set up, were routinely checked for the first 2 months, and were periodically checked for a further 4 months. Sparse matrix commercial crystallization screens included PEG/Ion, SaltRx, Index and Natrix (Hampton Research) and Structure screen I + II, PACT premier, JCSG-plus, ProPlex, Morpheus, and MIDAS (Molecular Dimensions) screens.

2.29.1.1 Claspin protein fragment crystallography

Crystallization studies of the Claspin protein fragments were carried out using the purified protein fragments B2C4, A1G12, B2C2 and B1F8 at concentrations between 5 and 60 mg/ml. Claspin-DNA co-crystallization studies used purified A1G12 or B2C2 protein at concentrations of 5, 10 and 15 mg/ml, mixed with either 20 bp + 2 nt dsDNA or 20 bp dsDNA DNA templates (section **2.16**) at a 2:1 or 5:1 ratio (DNA:protein), and incubated on ice for 20 minutes prior to setting up trays.

2.29.1.2 Chk1-KD – Claspin CKB phosphopeptide crystallography

Co-crystallization studies of Chk1-KD$^{1-270}$-His or Chk1-KD$^{1-289}$-His with Claspin CKB phosphopeptides: Flu-T916p, Flu-S945p, T916p or S945p (section **2.17**). For pre-concentrated Chk1-KD proteins, this was mixed with phosphopeptides at between a 1:2 and 1:8 (protein:peptide) ratio and incubated on ice for 2 hours or overnight at 4 °C, prior to setting up crystallization trials. For dilute Chk1-KD proteins, Claspin phosphopeptides were added to the protein with a 1:5 (protein:peptide) ratio and incubated on ice for 2 hours or overnight at 4 °C, prior to sample concentration (section **2.15.4**). Additives including 1:2.5 MgADP and/or 1:1.2 STU (Chk1:additive) ratio, were sampled with phosphopeptide addition.

Crystal screens were set up at concentrations between 2.5 and 6.0 mg/ml, at a 1:1, 2:1 or 3:1 drop ratio (protein to well solution; final drop volume of 0.4, 0.6 and 0.8 µl respectively). Crystals were identified for Chk1-KD$^{1-270}$-His with T916p phosphopeptide at 20 °C and these were optimized.

2.29.2 Protein crystal optimisation screening

Crystal optimisation screens were prepared manually as a 1:1 drop ratio (protein to well solution; final drop volume of 1 µl) in sitting-drop 48-well plates (MRC; Molecular Dimensions), with a reservoir volume of 100 µl. Optimisation crystallisation conditions were prepared around the original identified conditions.

2.29.2.1 Chk1-KD$^{1-289}$-His protein crystallography

Chk1-KD$^{1-289}$-His protein at 6.2 mg/ml, was used for apo crystallisation studies. Crystals grew at 14 °C, in the condition of 8% (v/v) PEG 8K, 0.1 M HEPES pH 7.5, 36% (v/v) ethylene glycol; based on optimization screen for PDB 3JVR (Vanderpool et al., 2009). The resultant crystals were harvested (section **2.29.5**). This screen was also set up with Chk1-KD$^{1-289}$-His with the S945p CKB phosphopeptide at a 5:1 (peptide:protein) ratio, with and without STU at a 1.2:1 (STU:protein) ratio; incubated as described in section **2.29.1.2**.

2.29.3 Chk1-KD$^{1-270}$-His sparse matric optimization

Optimization screens were prepared for Chk1-KD$^{1-270}$-His with the T916p phosphopeptide at a 5:1 (peptide:protein) ratio and incubated on ice for 2 hours. Screening was around conditions PEG/ion A2 (0.2 M Potassium fluoride pH 7.3, 20% (v/v) PEG 3350), PEG/ion C9 (0.2 M Sodium sulphate decahydrate pH 6.7, 20% (v/v) PEG 3350) and Structure Screen I + II D1 (0.2 M Sodium acetate pH 8.0, 20% (v/v) PEG 4000), and screens were set up at 20 °C. The resultant crystals were harvested (section **2.29.5**).

2.29.4 Chk1-KD$^{1-289}$-His protein crystal soaking

Chk1-KD$^{1-289}$-His protein crystals (section **2.29.2.1**) were used for peptide soaking. Harvested Chk1-KD$^{1-289}$-His crystals were added to 5 µl reservoir solution from the

crystallization condition, containing 0.8 mM S945p or Flu-S945p peptides (section **2.17**), incubated at 4 °C for 16 hours, and then harvested (section **2.29.5**).

2.29.5 Crystal harvesting

Protein crystals were harvested using a CrystalCap SPINE HT 0.1-0.2 mm loop (Hampton Research) by Dr Mark Roe (University of Sussex). If required, the protein crystals were cryoprotected by increasing the concentration of glycerol in the crystal, by transfer and soaking in reservoir solution containing increasing concentrations of glycerol [up to 30% (v/v) glycerol]. The crystals were flash frozen and stored in liquid nitrogen prior to data collection. Cryoprotection, protects the protein crystal in a vitreous glass-like nature, which enhances the diffraction of the crystal (improves the molecular order), prevents the formation of ice on freezing (can destroy the crystal), reduces the radiation damage to the protein crystal by X-rays (slowing free radical movement), and also enables the storage and movement of the protein crystals.

2.30 Structural determination by X-ray diffraction

Diffraction data was collected from a single crystal of Chk1-KD$^{1-289}$-His at a temperature of 100K on beamline IO2 at DLS, Didcot. Diffraction data was collected for 14 protein crystals of Chk1-KD from conditions containing CKB phosphopeptide, at a temperature of 100K on beamline IO4 at DLS, Didcot. Diffraction data for all crystals was automatically processed at the beamline using Xia2 3dii (Winter, 2009). The structure of all Chk1-KD crystals was processed using programs from the CCP4 suite (Collaborative Computational Project, Number 4; Winn et al., 2011) and the phases were solved using molecular replacement with Chk1 KD (PDB: 1IA8). Initial structural refinement was carried out using REFMAC5 (Winn et al., 2011) using 95% of the diffraction data. Manual model re-building and refinement was carried out using Coot (Emsley and Cowtan, 2004) following each round of refinement. Protein structures were modelled using PyMol (Schrodinger, 2010).

Chapter 3

“Super” Combinatorial Domain Hunting

3.1 Introduction

The production of multi-milligram quantities of high-quality recombinant protein presents a significant bottleneck for downstream applications such as biochemical and biophysical studies, drug discovery and structural biology. As previously described in section **1.2.2**, sub-constructs of a target protein often have a high level of success in yielding high-quality protein suitable for further biochemical and biophysical analysis. The success of sub-construct design in recent years has come from high-throughput techniques, either through bioinformatics-driven or random mutation / fragmentation approaches for the generation of construct libraries and subsequent high-throughput protein expression screens. Whilst these screens readily identify highly-expressed, soluble protein fragments, if such constructs are to be used for drug discovery, they must actually contain the 'domain of interest' (DOI), and express functionally active protein.

A combinatorial library technique, termed 'Combinatorial Domain Hunting' or CDH, was developed by Dr Renos Savva and his colleagues at Birkbeck and UCL, London (Reich et al., 2006). CDH, explained in **Chapter 1** and undertaken in **Chapter 4**, has traditionally been used on previously well-defined proteins, where there is a particular DOI, and typically focuses on the 'druggable' domains of 'difficult to express' proteins; such as protein kinase domains. The CDH methodology has been of great use in identifying highly-expressed protein fragments, which are both proteolytically stable and soluble. A typical CDH screen will analyse over 20,000 colonies, looking for expression of His_6 affinity-tagged proteins. Small-scale expression testing will then be carried out, typically on 750 clones, which express tagged protein from the initial colony screen. Any soluble, expressed protein will then be purified in a one-step purification method (using IMAC) and eluates analysed by SDS-PAGE. An additional 'solubility' screen is carried out on those constructs expressing high levels of recombinant protein, and again analysed by SDS-PAGE and western blot (Reich et al., 2006). These subsequent steps determine the expression levels of the recombinant protein and assess protein stability. The DNA from 'positive' clones is then isolated, and sequenced. This protocol identifies all of the expression constructs that encode high-yield, proteolytically stable, soluble protein; but does not assess functionality.

Therefore, additional screening is generally required to ensure that each of these constructs encodes a protein capable of folding to a functional tertiary structure; thereby identifying those suitable for use in both drug screening and / or protein crystallography.

Currently, there are some modifications that can be made to the CDH protocol in order to enhance the chances of identifying constructs expressing a DOI. These include: the selective gel purification of fragments (from the total DNA fragmentation library) at around the size of DNA encoding the domain you wish to identify; ensuring that the fragmentation library is of sufficient size to enable sufficient sampling of the fragment construct pool; and focusing expression library screening, of the 750 selected clones, to only those encoding proteins that are of a sufficient size to fully encompass the DOI. Despite these precautions, the vast majority of 'hits' identified will not fully encompass a particular DOI, and of those that do, fewer still will express any functional protein. To overcome this fundamental disadvantage of random construct screening, we planned to enhance the current CDH protocol, and enable the selection of 'functional' constructs at a much earlier point in time in the screen, and thereby reduce the overall number of positive, but 'off target' constructs.

The experiments reported in this Chapter, aimed to develop a high-throughput functionality-screening methodology. If this was successful, this would firstly reduce the amount of parallel clone handling – and thereby reduce the considerable amount of time and money spent on the identification of 'functional' constructs – and secondly could potentially enable the rapid sampling of a much larger initial CDH library.

3.2 Method design

The proposed methodology involves the use of a 'selective binding partner', specific to the DOI of a target protein, in order to 'fish' for protein fragments that are capable of interaction, ideally directly from a cell lysate. To do this, it was proposed to immobilise of a specific ligand or binding partner, for example an inhibitor, co-factor or substrate, on an affinity matrix in order to create an Affinity

Probe (AP). For a protein fragment to bind the AP, it must be soluble, and more importantly, must contain the DOI in a correctly folded state. The AP will also be attached to a matrix that facilitates the detection of binding in a high-throughput manner compatible with the CDH methodology (**Figure 3.1A**).

There were a number of variables to consider in developing this AP methodology. In order to test the idea, we first required a ligand that bound specifically to a number of different protein targets. This would be of importance once the initial proof-of-concept tests and refinements to the methodology had been carried out, in order to determine the feasibility of this method when applied to other protein targets. In this respect, protein kinases were an ideal choice, as these have a specific DOI, i.e. the kinase domain, which are highly druggable target and are traditionally difficult to express in the heterologous host *E. coli*. The selection of protein kinases also enabled us to use the broad-spectrum (kinase-specific) inhibitor STU as the immobilised ligand; this compound typically binds protein kinases with nanomolar affinity. Attachment of STU to an affinity matrix was achieved by coupling the compound with a biotin moiety, allowing attachment of the compound to commercially available streptavidin-coated resins / surfaces. The biotin-streptavidin interaction is the strongest non-covalent biological interaction currently known, thereby ensuring the AP (STU) is stably attached to the matrix. A polyethylene glycol (PEG) linker was also introduced between STU and biotin (**Figure 3.1B**), as this is conformationally flexible, and thus would hopefully enable the interaction of both moieties with their respective binding partners, without any steric clashes or hindrance. This type of linker is also soluble in water, aiding the aqueous solubility of the STU moiety.

For detection of protein binding, we required a sensitive but high-throughput method. 'Pull-downs' using Streptavidin (SA) resin in a 96-well format would theoretically work – an adaptation to the current CDH protocol, which uses Ni-NTA resin for His_6 affinity tag binding, but this would require the same level of manual sample handling as the current CDH protocol. Additionally, this would require large quantities of AP to complete the assay. As an alternative, an enzyme-linked immunosorbent assay (ELISA) assay could be used, by attachment of an AP to the

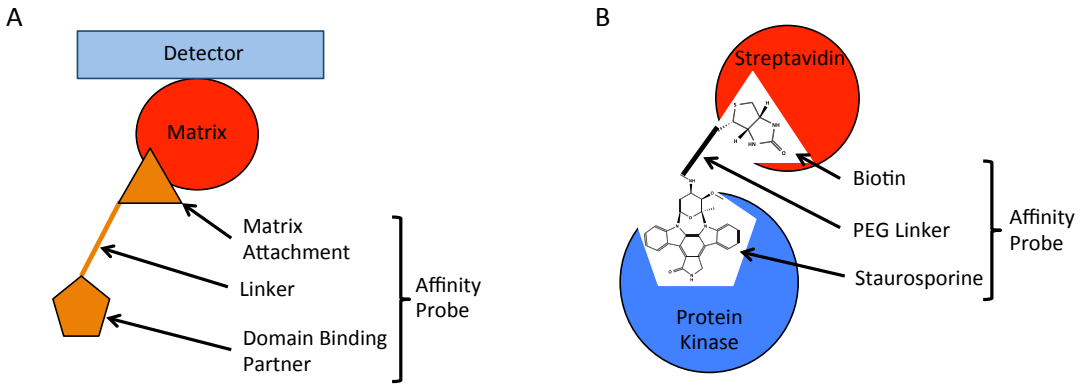


Figure 3.1 Design of the Affinity Probe.
(**A**) The proposed affinity probe comprised of a 'domain binding partner' linked to a moiety that is able to interact with a matrix attached to a detector / sensor; as indicated. (**B**) Proposed affinity probe for screening a protein kinase CDH library; composed of staurosporine joined by a PEG linker to biotin, for attachment to a streptavidin matrix as indicated.

bottom of a 96-well plate, enabling capture of individual proteins in each well (from cell lysates), and subsequent detection of binding using an anti-His antibody (or similar) and colourmetric detection. Whilst this would enable detection of both the functionality of the expressed protein and solubility, this would also require large quantities of AP to complete the assay, and this format of assay was not desired in the brief.

To detect protein interactions with an AP, we decided to use a ForteBio Octet Red96 (Octet), where eight samples can be run in parallel, and up to 96 samples (in 96-well plates) can be consecutively assayed. This instrument uses small biosensors that detect binding at the tip of the sensor, using Bio-Layer Interferometry (BLI) — an optical analytical technique. The instrument analyses the interference pattern of reflected white light as a result of a change in optical thickness at the probe tip, further described in **Figure 3.2** (Dayne, 2012). This system is extremely sensitive and enables the use of small and dilute volumes of sample. The Octet is also suitable for kinetic analysis for determining parameters for protein interactions with biomolecules and small molecules (Concepcion et al., 2009, Wartchow et al., 2011, Zuo et al., 2012). Unbound molecules, around the sensor tip are reported not to perturb the interference pattern, unless they are binding or dissociating, thereby enabling analysis in crude cell extracts (Concepcion et al., 2009). The Octet system has been successfully used to screen expression constructs, and has identified soluble His_6 affinity tagged recombinant proteins directly from cell lysates (Tian-Yu et al., 2010).

A screen using an AP and the Octet first involves hydration of the biosensor in an appropriate binding buffer (to be used throughout the assay) followed by loading of the AP to the biosensor, subsequent washes (to remove any unbound AP), and then probing of individual samples containing recombinant protein expressed in a CDH screen (**Figure 3.3**). This novel screening step would be most effectively used after the cellular lysis and clarification steps for the 750 selected clones (i.e. the positive hits from the colony-blot) and immediately prior to the 96-well purification step – therefore this methodology was to be developed in order to directly probe the soluble fraction of *E. coli* lysates. Whilst *E. coli* does express

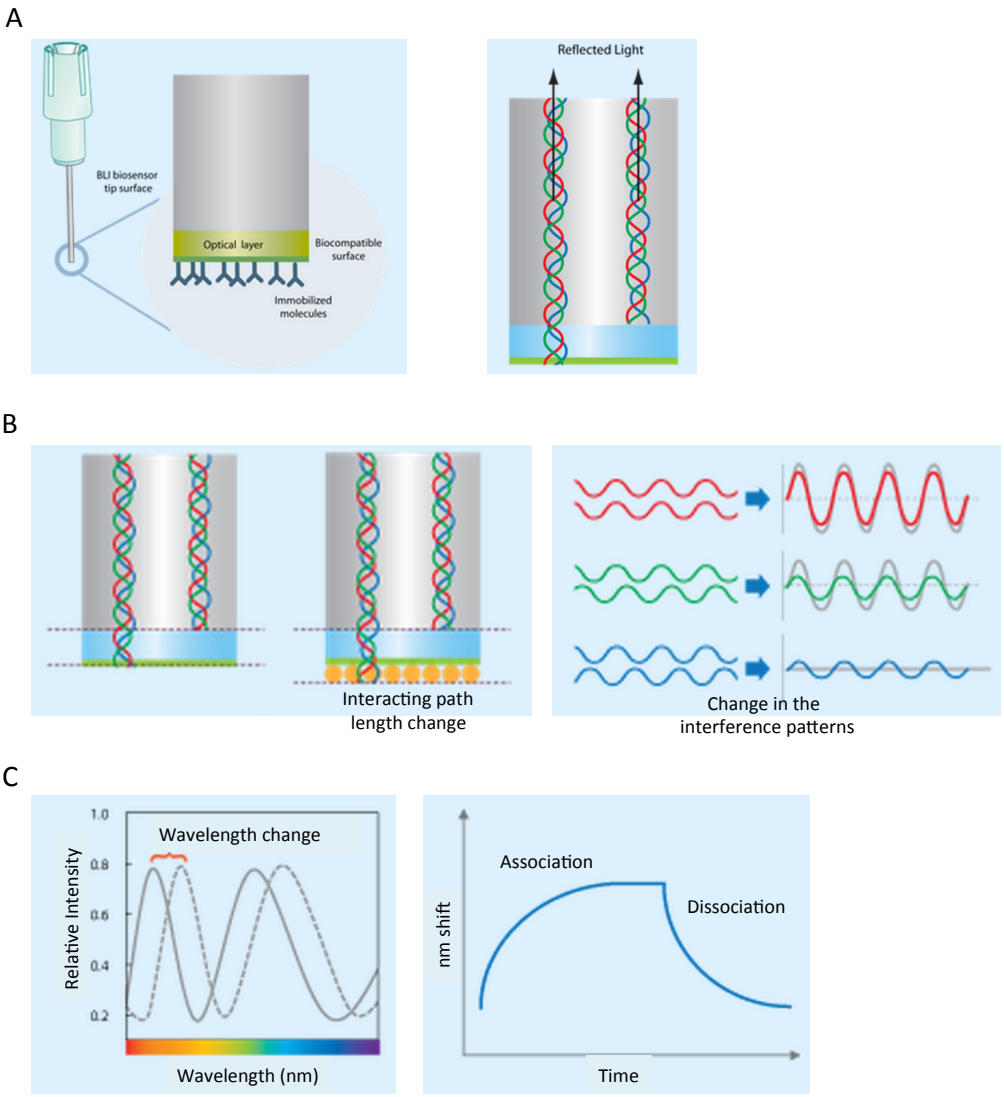


Figure 3.2 BioLayer Interferometry.
BioLayer Interferometry (BLI) is an optical technique for measuring biomolecular interactions, by detecting the change in a reflected white light interference pattern. (**A**) (Left) a ForteBio BLI biosensor tip surface, showing immobilised molecules attached to its surface. (Right) two reflection interfaces are contained within the biosensor, one at the interface of the glass fibre before the immobilised molecules on the tip and the second at the interface between in tip surface and solution. (**B**) (Left) the two reflections contain the same visable light wavelengths, and (right) the interference pattern for each individual wavelength reflection can be detemined. Binding of molecules to the surface of the tip results in a change in the reflected wavelengths, which results in a change in the interference pattern. (**C**) (Left) for the biosensor alone, the reflected wavelength intensity is plotted against the wavelength, the shift in the reflected wavelength upon binding at the surface of the tip is plotted and the nm shift in the wavelength is determined. (Right) the wavelength shift as a function of time is plotted, showing the association and dissociation of the molecule from the surface of the biosensor, from which the association and dissociation constants can be calculated. Adapted from Dayne, (2012).

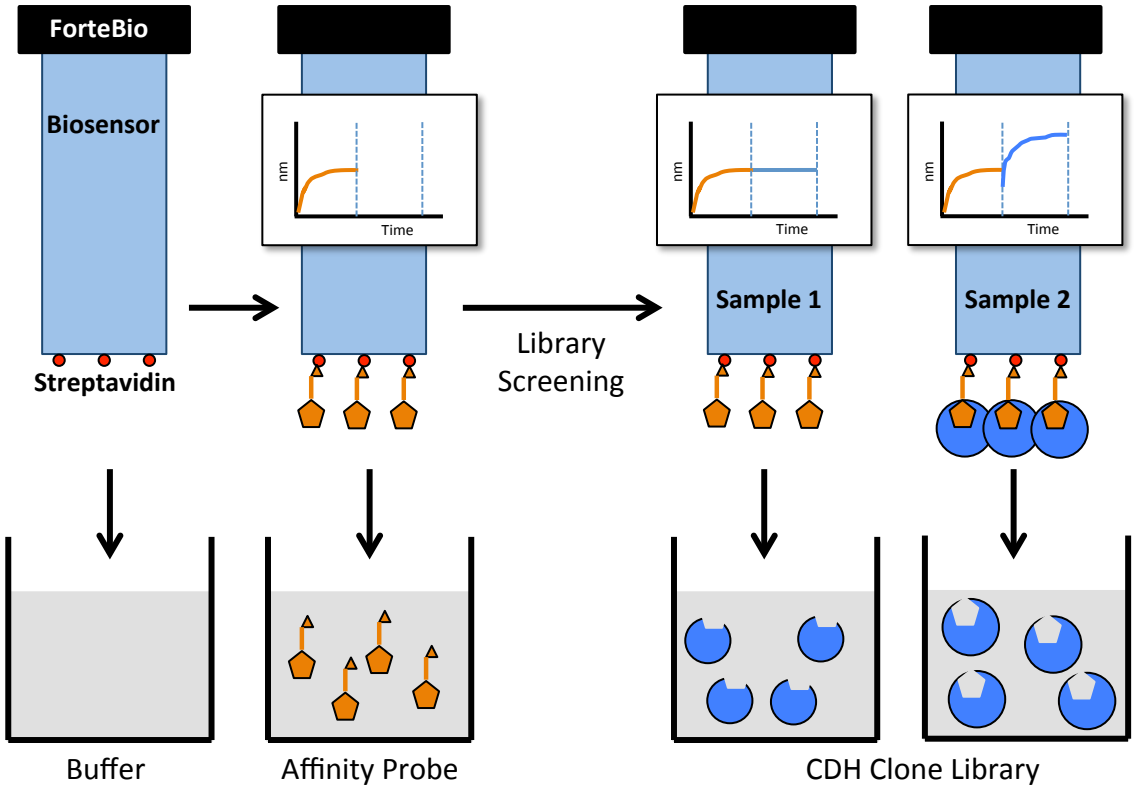


Figure 3.3 Schematic showing the Octet Affinity Probe screening methodology.
A schematic for the proposed ForteBio Octet methodology: where a streptavidin probe will first be hydrated in buffer and then coated in the AP from **Figure 3.1B**. This will then be used to probe individual protein samples expressed from a CDH library to look for correctly folded (functional) interactors. Insets: illustrative sensorgrams representative of (1) a sample where there is no increase in signal when a sensor is dipped into a cellular lysate, i.e. indicating no binding, and (2) an increase in signal due to correctly folded protein interacting with the Affinity Probe.

some histidine protein kinases, two tyrosine protein kinases and has some cellular serine/threonine autophosphorylation activity reported; there are no reports that any of these proteins interact with STU (Kennelly, 2002, Macek et al., 2008). The identification of any protein fragment/s binding to the AP on the Octet would then be channelled through the standard CDH protocol. This methodology could potentially be used to investigate a wide range of proposed druggable targets, for which the tertiary structures have yet to be determined.

In summary, in this Chapter, the intention was to develop a novel screening technique (Octet / AP) and show this could readily identify soluble, correctly folded protein kinases from cell lysates during a CDH screen.

3.2.1 Preparatory assay development by Domainex

Prior to embarking on this project, preparatory work had been carried out at Domainex (Cambridge, UK) using the Octet. However, this was carried out using an alternative biotinylated drug and protein target (Proprietary, unpublished data). This preliminary work showed promise, as specific binding of a purified protein at nanogram quantities to a biotinylated immobilised drug could be detected.

3.3 Modelling of the affinity probe

The desired characteristic of our AP, in order for it to be successful, was that it must be able to interact with both the SA matrix and the protein kinase at the same time; thus the optimal length of the PEG linker connecting the two moieties needed to be determined. If the PEG linker length were too long this may prevent the detection of kinase binding to the biosensor (by BLI). Conversely, if the PEG linker length were too short, this could hinder protein interaction with the AP when attached to the SA matrix. Ideally, whilst it would be preferable to make a number of APs with varying PEG linker lengths, due to cost constraints only one PEG linker length was selected and then tested (**Figure 3.4A**). This was examined by molecular modelling techniques using Coot (Emsley and Cowtan, 2004), using a structure of a protein kinase in complex with STU (PDB: 3RY2) and that of biotin bound to streptavidin (PDB: 1AQ1), we were able to ascertain that a PEG_4 linker would enable interaction of the AP with both the immobilizing affinity

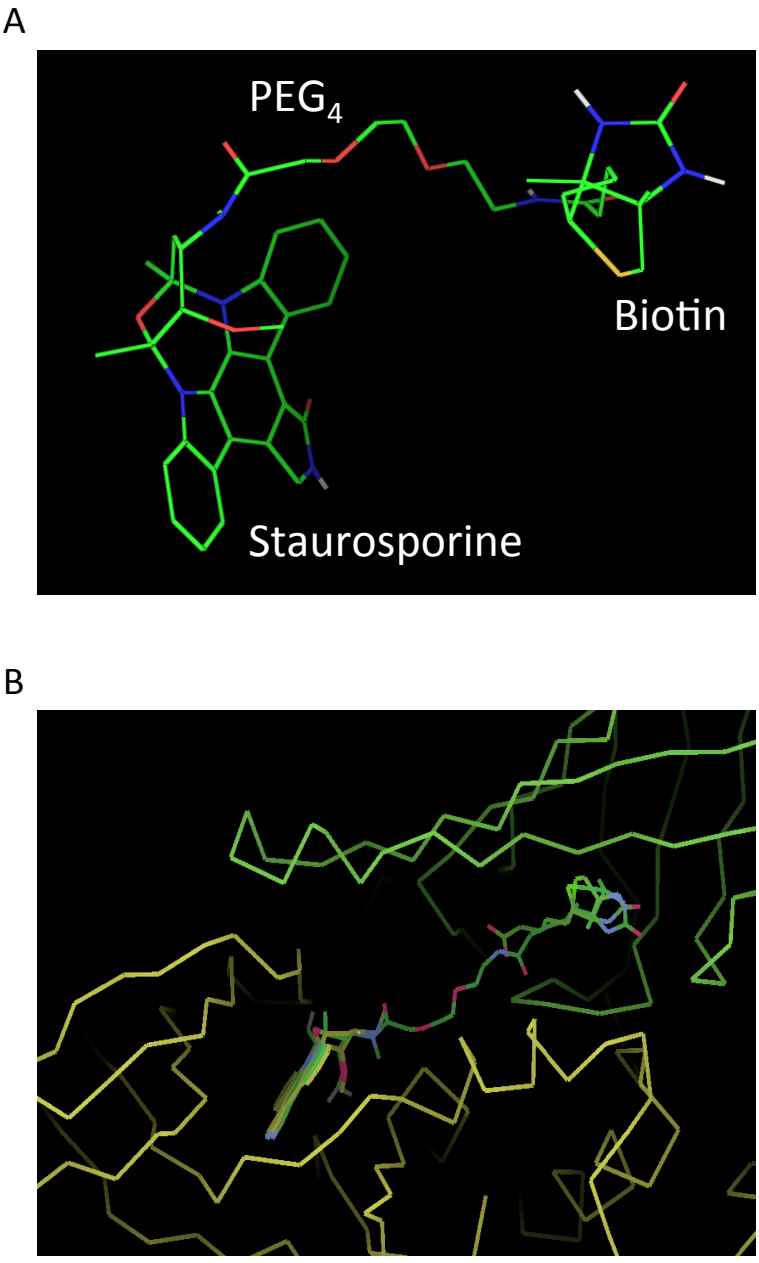


Figure 3.4 Structure and modelling of STU-PEG$_4$-BIOTIN.
(**A**) Structure of the affinity probe, a PEG$_4$ linker that links staurosporine to biotin. Image drawn using PRODRG (Schuttelkopf and van Aalten, 2004) and PyMol (Schrodinger, 2010). (**B**) Molecular modelling of the affinity probe bound to a protein kinase (PDB: 3RY2) and streptavidin (PDB: 1AQ1) using Coot (Emsley and Cowtan 2004).

matrix and the protein kinase, whilst keeping the interaction distance short enough to facilitate detection of kinase binding by BLI (**Figure 3.4B**).

3.4 Synthesis of the Affinity Probe

The STU-PEG_4-BIOTIN AP was synthesised by Michael Paradowski (University of Sussex) (**Figure 3.5A**). The synthesis was based on that previously described for a fluorescently-labelled STU compound, used in high-throughput, competitive screen of kinase inhibitors (Kawaguchi et al., 2008). The conditions described by Kawaguchi et al. (2008) to form the pegylated intermediate proved difficult to reproduce and led to difficulties in purification. Therefore, a more 'classical' and alternative approach to synthesis was considered. STU was first coupled to a Fmoc protected acid linker (Fmoc-NH-PEG_4-COOH) to give the required intermediate (1), and minor impurities were removed by flash chromatography (**Figure 3.5B**). Standard de-protection of intermediate (1) led to a de-protected amine intermediate (2), which was also purified by flash chromatography (**Figure 3.5C**). The primary amine of intermediate (2) was then coupled to NHS-BIOTIN by acetylation forming the STU-PEG_4-BIOTIN AP, which was again purified by flash chromatography (**Figure 3.5D**). Analysis of the compound by NMR and subsequent peak assignments confirmed synthesis of the expected compound with an experimentally determined mass of 925.40 Da, in good agreement with the calculated mass of 926.08 Da (**Figure 3.6**). The STU-PEG_4-BIOTIN AP was provided as a yellow, lyophilised powder.

3.4.1 Solubilisation of the Affinity Probe

STU is known to be highly insoluble in aqueous buffers and is instead typically dissolved in a water-miscible solvent such as DMSO. It was thought that coupling of STU, via the PEG_4 linker, to biotin (which themselves are both highly soluble molecules) would increase the overall solubility of the compound. However initial attempts to dissolve the AP at high concentration, in aqueous buffers, were unsuccessful. Instead, the AP was first dissolved in 100% (v/v) DMSO, to create a stock solution at high concentration, which could then be diluted into aqueous buffers, suitable for assays.

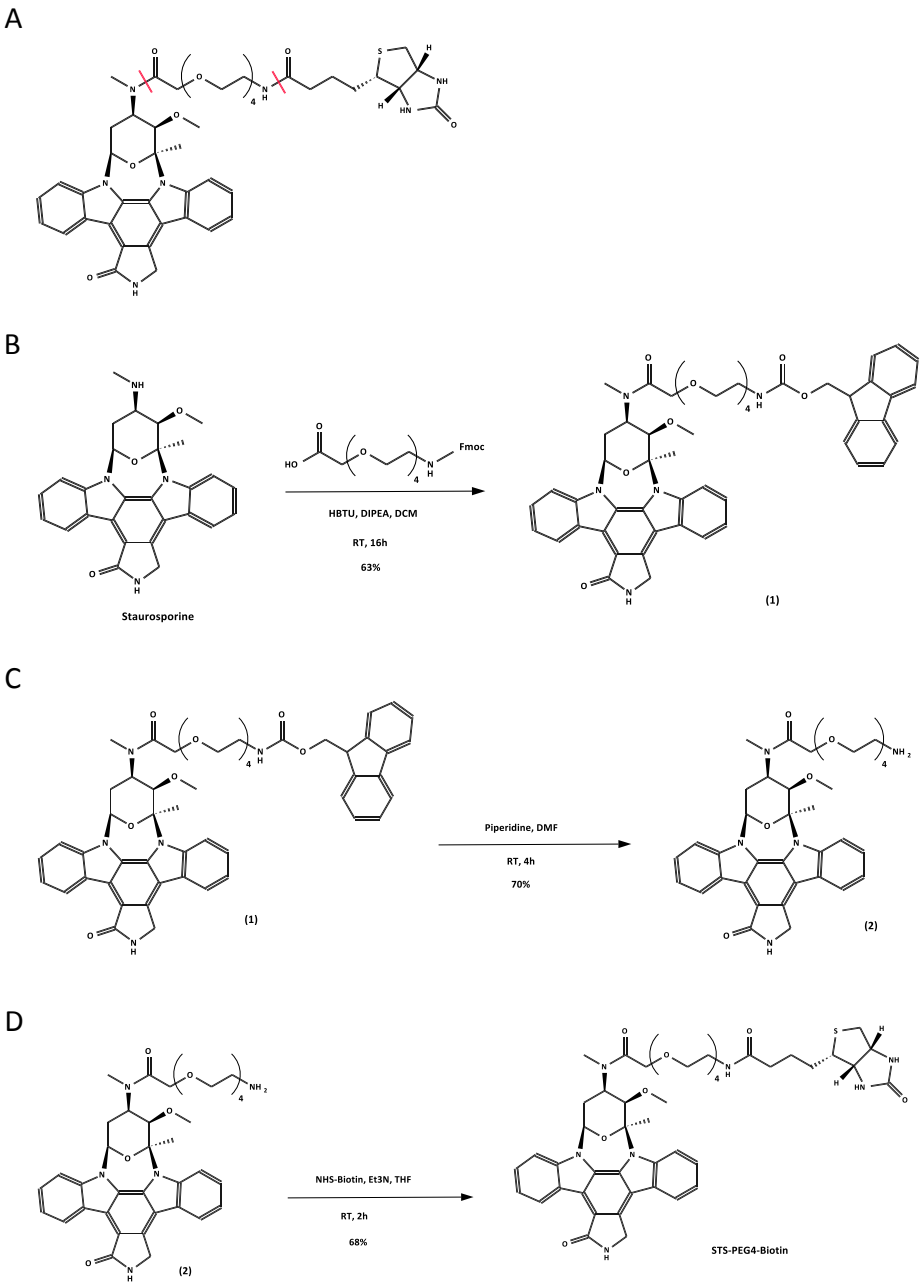


Figure 3.5 Synthesis of the Affinity Probe.
(**A**) Chemical drawing of STU-PEG$_4$-BIOTIN. (**B**) Coupling of the carboxylic acid group of the Fmoc-PEG$_4$ linker to the amine group of STU, in the presence of HBTU, DIFEA and DCM at room temperature for 16 hours, with a 63% yield of reaction intermediate (1). (**C**) De-protection of the second reactive group on the PEG$_4$ linker by the removal of the Fmoc head group (1), leaving an amino group, in the presence of piperidine and DMF at room temperature for 4 hours, with a 70% yield of reaction intermediate (2). (**D**) Coupling of intermediate (2) to biotin, in the presence of Et3N and THF at room temperature for 2 hours, with a 68% yield of the final product. Michael Paradowski (University of Sussex) developed and synthesized the compound, and developed this synthesis pathway and schematic.

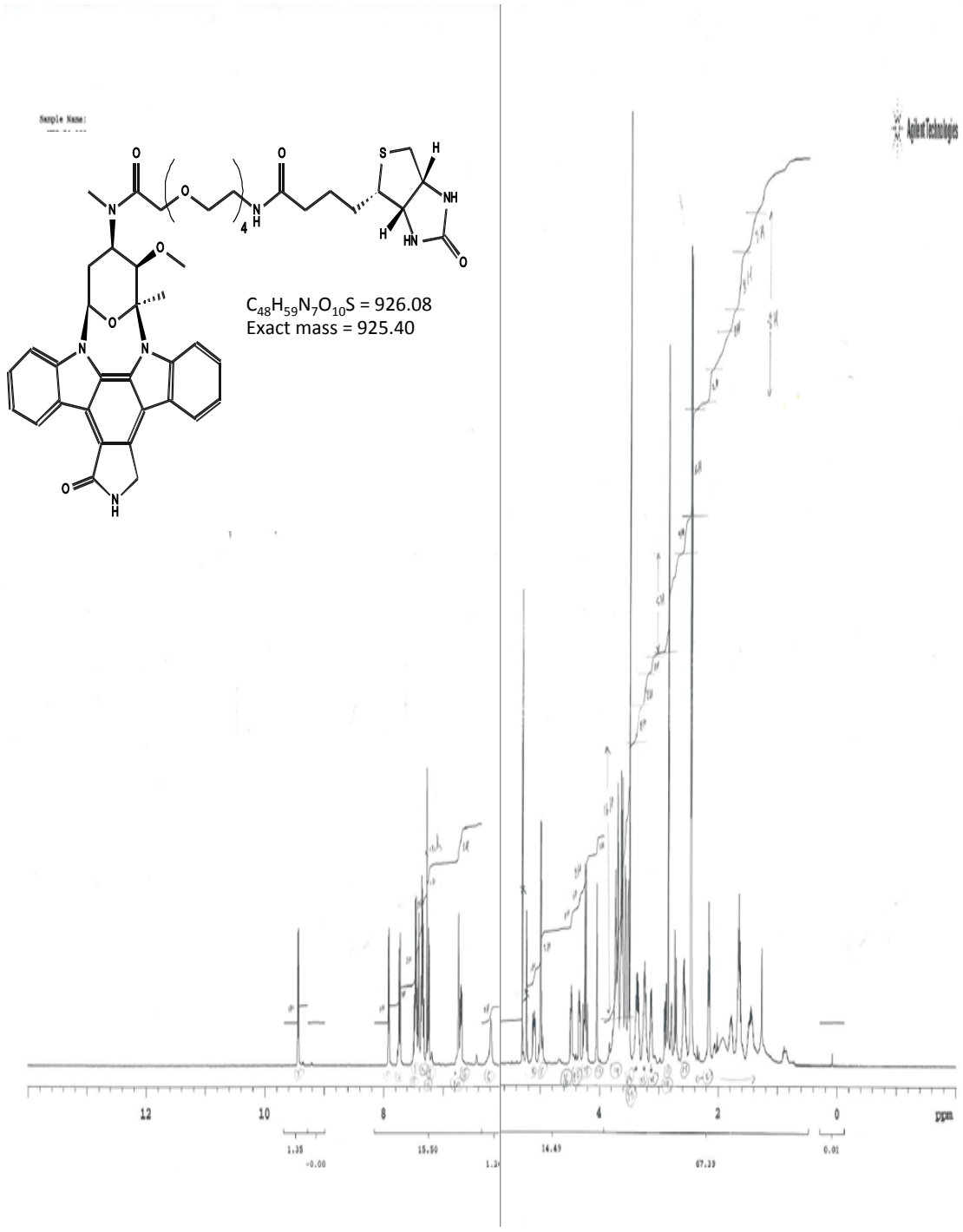


Figure 3.6 1D-NMR spectrum of STU-PEG$_4$-BIOTIN.
The data was collected on a VNMRS Varian 500 MHz instrument by Michael Paradowski (University of Sussex), who also carried out peak assignment, and determination of experimental compound mass.

3.5 A 'Proof-of-concept' protein kinase

Before using this method on a structurally undefined kinase target, it was important to first determine if this type of method could actually be used, and to investigate any factors that could affect the experimental outcome. Therefore a 'proof-of-concept' study was designed, using the previously characterised protein kinase MEK1 (MAPK/ERK kinase 1). MEK1 is a dual specificity kinase and forms part of the RAF / MEK / ERK signalling cascade, which is found to be constitutively activated in a number of different cancers. MEK-1 is a 393 amino acid protein comprised of a regulatory region (aa 1-60) and a catalytic domain (aa 61-393). The tertiary structure of the kinase domain of MEK1 (PDB: 1S9J) was determined in 2004 (Ohren et al., 2004), but in a crystal form that was inimical to further structure-based drug discovery efforts. Subsequently, MEK1 was put through a CDH screen, which identified an optimised expression construct that produced recombinant protein more suitable for fragment-based drug discovery methods (Meier et al., 2012). The construct 4F11, encodes the catalytic domain (aa 45-392) of MEK1, and the expression of the protein was further improved by replacing the unstructured activation loop (aa 264-307) with a 'Gly-Ser-Gly-Ser-Gly-Ser' linker.

3.5.1 Expression and purification of MEK1-KD

The kinase domain of MEK1 (MEK1-KD, encoded by construct 4F11) was purified as described in (Meier et al., 2012) by TALON-IMAC followed by SEC. After the first step, the captured MEK1-KD protein migrated at a position equivalent to a molecular mass of 36 kDa by SDS-PAGE, equivalent to the calculated molecular mass of 35.7 kDa (**Figure 3.7A**). Purification by SEC, MEK1-KD eluted as a single-peak from the SEC step, and selected fractions were examined by SDS-PAGE. (**Figure 3.7B**). The purity of MEK1-KD had been substantially improved, but some contaminants were still co-eluting from the column; whilst the purity of the sample would not be suitable for crystallography, it was sufficient for our analyses using the Octet instrument.

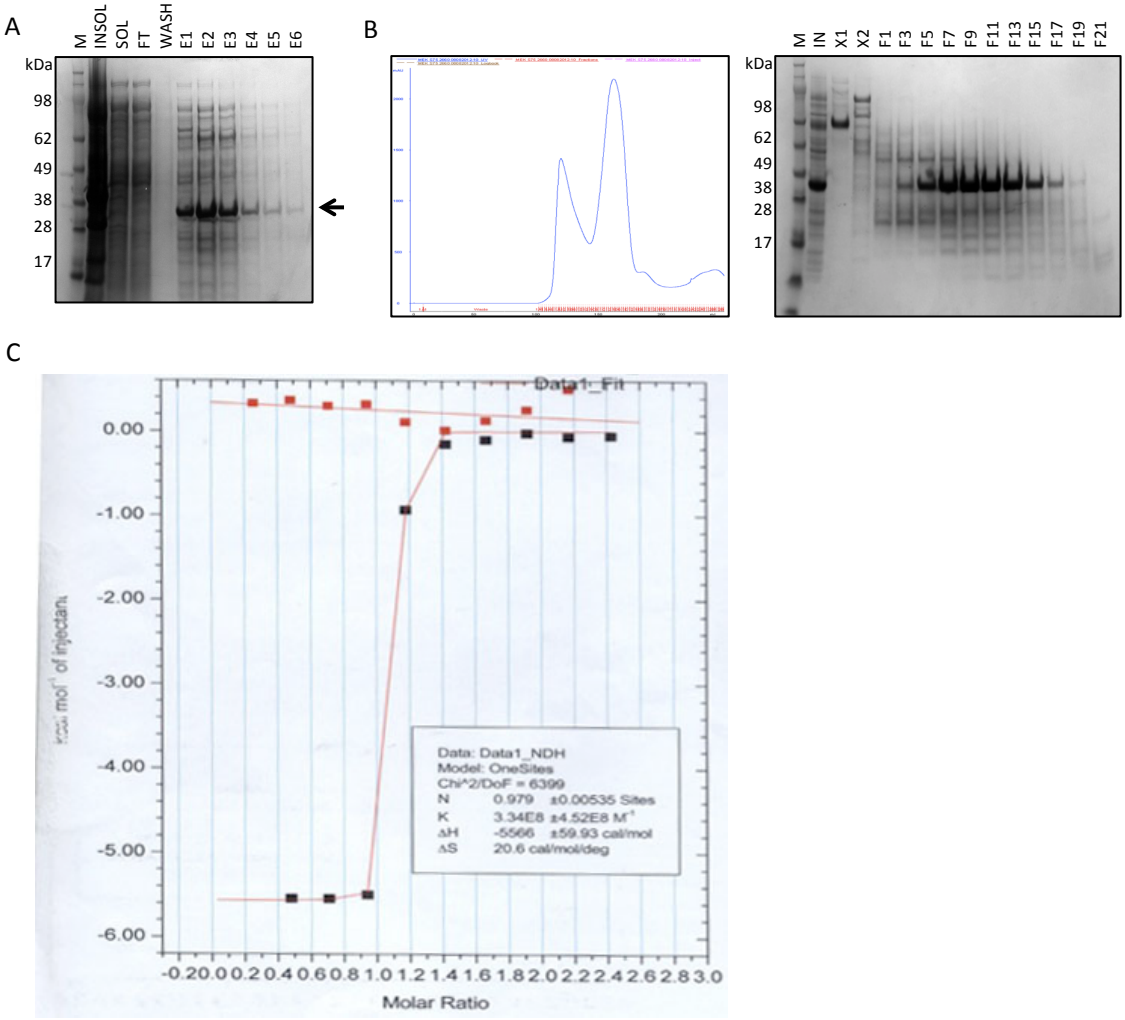


Figure 3.7 Purification of MEK1-KD and binding to STU.
(**A**) SDS-PAGE gel of *E. coli* cell lysate, and IMAC purification steps. The migration position of MEK1-KD is indicated by an arrow. M=molecular marker, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fractions, E1 to E6=successive elution fractions. (**B**) SEC using a 26/60 SD75 column. (Left) representative chromatograph, showing absorbance at 280 nm (blue line) and collected fractions. The elution peak for MEK1-KD is indicated with an asterisk. (Right) SDS-PAGE analysis of selected fractions. IN=input, X=fractions from the void volume, F=indicated fraction. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue. (**C**) Isotherm for titration of MEK1-KD with STU. For heats of interaction, nine 13.5 µl aliquots of 231 µM of MEK1 were injected into 20 µM of STU (black points). For heats of dilution, nine 13.5 µl aliquots of 231 µM of MEK1 were injected in buffer and subtracted from the heats of interaction. Corrected data were fitted using a non-linear least-squares curve-fitting algorithm (Microcal Origin), with three floating variables: stoichiometry, binding constant, and change of enthalpy of interaction (inset box).

3.5.2 Testing the functionality of MEK1-KD

Isothermal Titration Calorimetry (ITC) was used in order to confirm if the modified MEK1-KD retained its ability to bind to STU. ITC measures thermodynamic parameters of an interaction – allowing the determination of the dissociation constant (K_d) and the stoichiometry of interaction. An isotherm for binding of STU to MEK1-KD is shown in **Figure 3.7C.** Fitting of the measured experimental data indicated a 1:1 stoichiometry (MEK1-KD:STU) for the interaction and a K_d of 3.3 ± 4.5 nM. The K_d value is associated with a large standard deviation; this due to the high concentration of STU injected into the sample cell, leading to rapid saturation of MEK1-KD and few data points comprising the isotherm. However, the experiment confirmed that the 'engineered' MEK1-KD protein still bound to STU despite the shortening of the activation segment, and thus validated the use of this protein in our 'proof-of-concept' study.

3.6 Pull-down experiments

To confirm that BIOTIN-PEG_4-STU was suitable as an AP pull-down (co-precipitation) experiments were carried out using purified protein kinases and a Claspin protein fragment (A1H3) (section **2.10**). The AP (50 μM) was first immobilised on ~50 μl Dynabead MyOne Streptavidin T1 resin (SA Dynabeads) in MEK1 GF buffer supplemented with 50% (v/v) DMSO and 2% (v/v) Tween-20. After binding, the SA Dynabeads were washed in MEK1 GF buffer containing reduced concentrations of DMSO and Tween-20 [2% (v/v) DMSO and 0.05% (v/v) Tween-20]. The AP-beads were incubated with MEK1-KD and a fragment of Claspin, and with two additional purified protein kinase domains from Chk2-KD (Oliver et al., 2006) and GWL kinase from Dr Mohan Rajasekaran (University of Sussex), at a concentration of 25 μM. After incubation the beads were washed, and analysed by SDS-PAGE. MEK1-KD was clearly retained by the AP-Dynabeads whereas the Claspin protein fragment was not (**Figure 3.8A**). Similarly, both Chk2-KD and GWL protein kinases interacted with the AP-Dynabeads, albeit to somewhat different extents (**Figure 3.8B**). This experiment indicated that the immobilised AP could still interact with MEK1-KD; that the selected linker length was sufficient to allow both interactions.

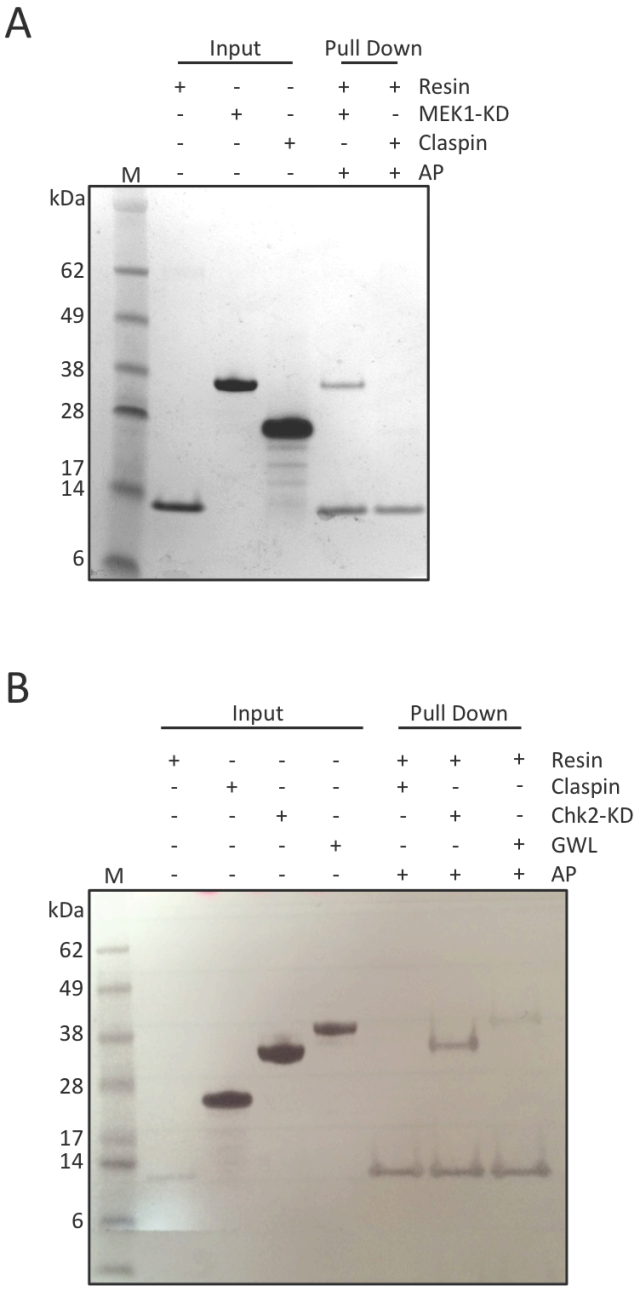


Figure 3.8 The AP specifically retains protein kinases.
The AP was immobilized on Dynabeads MyOne Streptavidin T1 resin prior to incubation with purified proteins. Bound samples were resolved on a 4-12% Bis-Tris SDS-PAGE gel and stained with Instant Blue. The input shown was 9% of the total sample, whilst the pull down sample was 28% of the total sample volume. Samples are as indicated. M=molecular mass marker. (**A**) AP interaction with MEK1-KD protein or a Claspin protein fragment. (**B**) AP interaction with Chk2-KD or GWL protein kinases or a Claspin protein fragment.

3.7 Experimental parameters for the ForteBio Octet RED96

The Octet in our laboratory is not equipped with a cooling system, therefore all experiments were carried out at ambient temperature (typically around 20 ˚C). Streptavidin biosensors were used for each of the following experiments. Following the manufacturer's recommended protocol, each biosensor was pre-hydrated for 30 minutes in the appropriate assay buffer (described individually) prior to use. The running of the assays is described in section **2.11.1**.

3.7.1 Testing the Affinity Probe

The first step was to determine if our AP could bind to the sensors, and what concentration would be optimal for its detection. Sensors were first hydrated, and moved into the assay buffer (MEK GF buffer) in order to determine a baseline. Sensors were moved into wells containing the same buffer supplemented with 25, 50 or 100 μM of the AP (23.15, 46.3 and 92.6 μg/ml respectively). They were incubated for 200 seconds to allow association, and transferred to another set of wells - again containing assay buffer - in order to measure dissociation (**Figure 3.9**). The loading step of the experiments showed an increase in signal for the AP binding to the SAB, indicative of interaction; this increased with the concentration of AP (Step 2; **Figure 3.9**). A rapid decrease in signal was observed during the dissociation phase (Step 3; **Figure 3.9**). However, once equilibrium had been achieved, the signal stabilised at a value higher than that observed for the initial baseline, indicating that the AP had bound successfully to the sensor. Some problems were observed with AP insolubility and non-specific binding of the AP to the surface of the sensors (i.e. not mediated by the biotin group). It was found that it was necessary to supplement the experimental buffer with 2% (v/v) DMSO and to include Tween-20 [initially 0.2% (v/v) Tween-20] in order to reduce this effect.

3.7.2 Octet assay development

Test experiments were first carried out with saturating concentrations of both the AP and MEK1-KD (**Figure 3.10**). As before, sensors were pre-hydrated and a baseline recorded. Sensors were then transferred into 10 μM (9.26 μg/ml) AP or into buffer (as a control). Any unbound AP was then removed by transferral of the

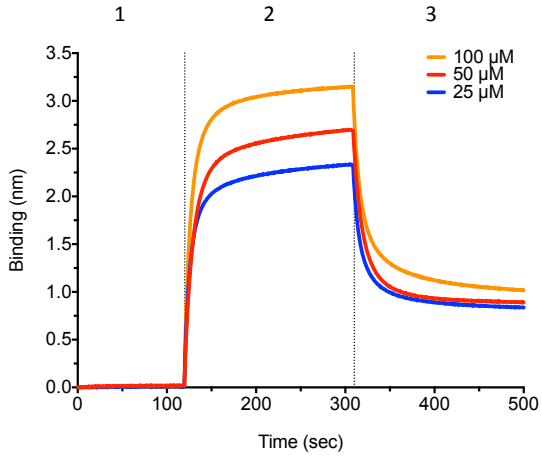


Figure 3.9 Binding of the AP to streptavidin-coated biosensors.
Sensors were pre-hydrated for 30 minutes in MEK1 GF buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for three independent sensors incubated with increasing concentration of AP (25, 50 and 100 µM, coloured blue, red and orange respectively. Step 1 represents baseline determination; Step 2 represents incubation in AP (loading); and Step 3 represents incubation in fresh-buffer (disassociation). The figure legend details Step 2 for each curve.

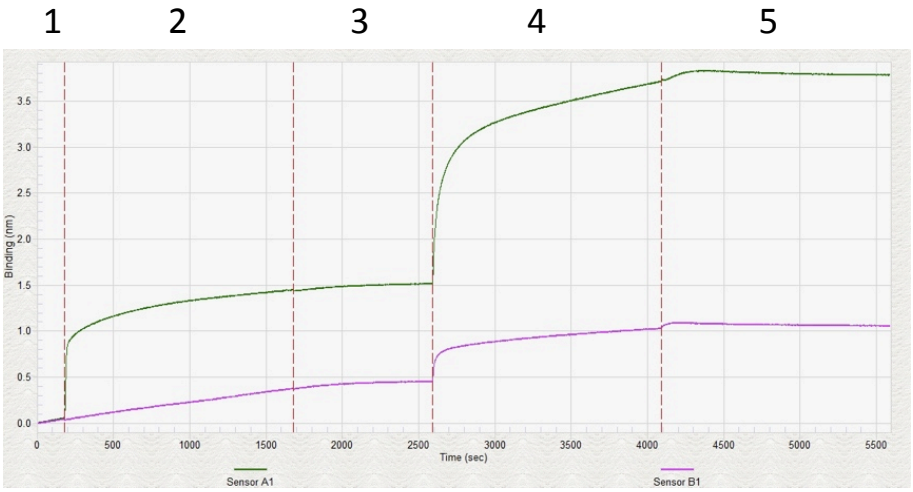


Figure 3.10 Initial 'proof-of-concept' test on the Octet.
Sensors were pre-hydrated for 30 minutes in MEK1 GF buffer [with 2% (v/v) DMSO and 0.2% (v/v) Tween-20]. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for two independent sensors incubated with AP (10 µM) and MEK1-KD protein (10 µM). Step 1 represents baseline determination (both lines); Step 2 represents incubation in AP (green line) or buffer (pink line) (loading); Step 3 represents incubation in fresh-buffer (both lines; disassociation); Step 4 represents incubation in MEK1-KD protein (both lines; association); and Step 5 represents incubation in fresh-buffer (both lines; disassociation).

sensors to fresh buffer. There was a small increase in signal seen for the buffer control (Step 2, **Figure 3.10**). However, this signal did not change when the sensor was moved into the 'washing' stage (Step 3, **Figure 3.10**). In contrast, a large change in signal was seen for the sensor dipped into the solution containing the AP, which was retained once the sensor was transferred into the 'washing' step (Step 3, **Figure 3.10**); again indicating successful immobilisation of the AP onto the sensor surface. For the binding experiment, the sensors were transferred into 10 μM MEK1-KD (357 μg/ml) for the association step, and then to wells containing fresh buffer for the dissociation step (green line, Steps 4 and 5, **Figure 3.10**). A significant increase in signal was seen for the AP-bound sensor, indicating an interaction with MEK1-KD. Moving the sensor into fresh buffer resulted in little change to the signal - indicating little dissociation of the protein. However, some level of non-specific binding of MEK1-KD to the control sensor was still observed despite the presence of both Tween-20 and DMSO (purple line, Step 4, **Figure 3.10**). However, this preliminary experiment, and corroborating data from pull-down experiments (section **3.6**), indicated that the immobilised AP could still interact with MEK1-KD, and that the selected linker length connecting the biotin and STU moieties was sufficient to allow an interaction, enabling detection of binding.

3.7.3 The effect of detergent on binding to Octet AP-Sensors

CDH buffers containing detergents are generally avoided, as these can result in false-positives through the artificial solubilisation of protein fragments. Whilst DMSO was found to be essential for the solubility of the AP, the effects of detergent were unclear. To test this, both the AP and MEK1-KD protein were prepared in varying concentrations of Tween-20 (or not containing Tween-20) and tested in binding experiments. In these experiments, a 10-fold lower concentration of both AP and MEK1-KD was used – i.e., 1 μM of AP (926 ng/ml) and 1 μM of MEK1-KD protein (35.7 μg/ml) in order to try and minimise the observed non-specific binding (**Figure 3.11A**).

The presence of Tween-20 in the assay buffer resulted in a greater amount of AP being immobilised on the sensors; the lowest concentration of 0.05% (v/v) Tween-

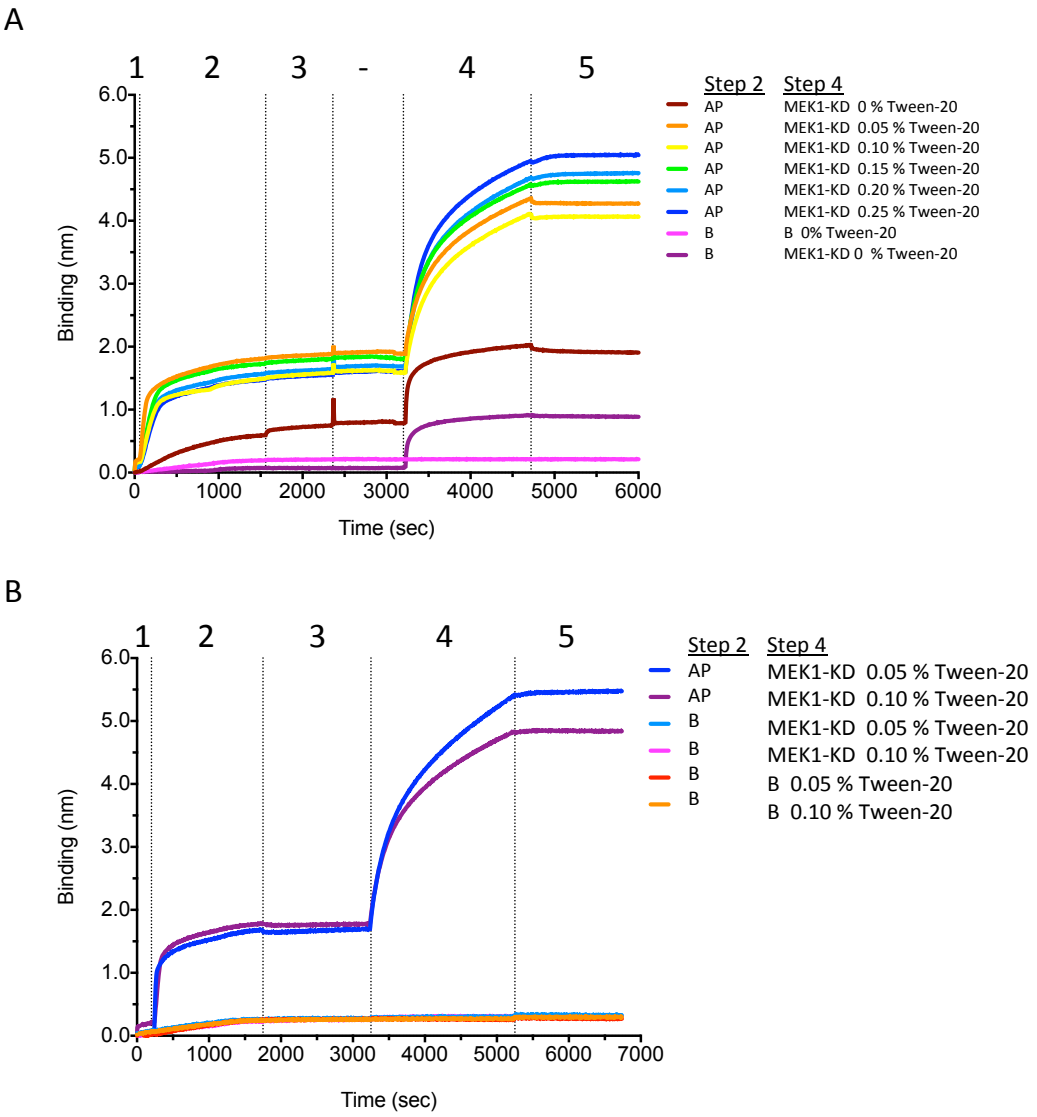


Figure 3.11 Determining and refining the optimal Tween-20 concentration for Octet assay development.

(**A**) Determining the optimal Tween-20 concentration for the assay. (**B**) Refinement of the optimal Tween-20 concentration for the assay. Sensors were pre-hydrated for 30 minutes in MEK1 GF buffer [with 2% (v/v) DMSO and Tween-20 concentration as specified]. The experiments were carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for '**A**' eight, and for '**B**' four independent sensors incubated with AP (1 µM) and MEK1-KD protein (1 µM), with changes in the Tween-20 concentration as indicated. Step 1 represents baseline determination; Step 2 represents incubation in AP or buffer (loading); Step 3 represents incubation in fresh-buffer (disassociation) (for '**A**' Step '–' was a buffer exchange step to the relevant Tween-20 concentration); Step 4 represents incubation in MEK1-KD protein (association); and Step 5 represents incubation in fresh-buffer (disassociation). The figure legends detail Steps 2 and 4 for each curve. AP=Affinity Probe, B=buffer.

20 was sufficient to observe this effect. Furthermore, the AP was stably associated with the sensors, as no dissociation could be seen when they were transferred into fresh buffer. The sensors were then moved into solutions containing MEK1-KD, in order to observe binding. The 10-fold reduction in protein concentration did not have an effect on the signal generated, although the association phase had slower kinetics than had been seen previously (10 μM MEK1-KD protein, refer to **Figure 3.10**). The concentration of Tween-20 did not greatly affect binding of MEK1-KD to the sensors, though the overall level of binding was slighter higher with the increased detergent concentrations. In the absence of Tween-20, binding of MEK1-KD to both the unmodified sensor is observed, thus indicating a high degree of non-specific binding. These results indicate that the presence of detergent is required for enhancing binding of the AP to the sensor, and also help to prevent non-specific interactions of MEK1-KD protein with the unmodified sensor surface.

To confirm the effect of detergent on the assay, the experiment described above was repeated as described, with a final concentration of either 0.05 or 0.10% (v/v) Tween-20 in all experimental buffers (**Figure 3.11B**). The binding curves for the association of the AP with the sensor, and for MEK1-KD binding to the modified sensor are essentially identical to those described above. The presence of Tween-20 eliminated any non-specific binding of the protein to the unmodified sensor. Therefore, in order to (a) prevent the artificial solubilisation of expression constructs arising from the CDH method (i.e. false positives) and (b) ensure robustness and reliability in the Octet-based assay, the lowest detergent concentration of 0.05% (v/v) was chosen [DTM buffer: MEK1 GF buffer with 2% (v/v) DMSO and 0.05% (v/v) Tween-20].

3.7.4 Specificity of the Affinity Probe

In screening of a CDH expression construct library, it is critical that the STU component of the AP is only able to bind correctly folded protein kinase domains. This could be demonstrated by two assays: the first by pre-incubation of a protein kinase with an inhibitor, and then testing the interaction with a AP-modified sensor, and second by screening additional protein kinase domains (positive controls) and other non-kinase proteins (negative controls).

For the first assay, MEK1-KD that had been pre-incubated with STU was used in a binding experiment (**Figure 3.12**). Only a small increase in signal was observed for the AP-modified sensor that was dipped into MEK1-KD pre-incubated with STU. Movement of this sensor into buffer (for dissociation) resulted in a drop of signal back to the baseline level. Pre-incubation of MEK1-KD with STU prevented the interaction of the protein with the AP-modified sensor, thus indicating that the kinase domain was making a functional interaction with the AP.

For the second assay, a second protein kinase domain termed 'Kinase X' and the protein NF-κB p65 (both provided by Dr Stefanie Reich, Domainex) were tested in the assay, to confirm that the AP could discriminate between constructs encoding protein kinases from those encoding other types of protein (**Figure 3.13**). Kinase X was able to interact with the AP, but the overall level of binding (signal) was less than that previously observed for MEK1-KD. For AP modified sensors dipped into NF-κB p65, there was an initial increase in signal, but this rapidly decreased over time. As this type of binding profile had not been observed before, the experiment was repeated, and an almost identical profile obtained. Despite this unusual behaviour, the assay still indicated that folded kinase-domains could ready interact with the AP-modified sensors.

3.8 Octet-based assay using cell lysates

As demonstrated above, this Octet-based methodology appeared to work for highly purified recombinant proteins, but this is not a feasible approach for a high-throughput CDH screen. The current CDH protocol involves on average 750 individual purifications in a 96-well format, and whilst this process has been streamlined, it is still incredibly time-consuming. The most effective position for this method, as part of CDH, would be after cellular lysis and removal of insoluble material, but prior to the IMAC purification step – i.e. to identify functional, folded protein kinase fragments from clarified lysates.

In order to test if the AP-modified sensors could be used directly in *E. coli* cell lysates, the following experiment was devised: AP-modified sensors were dipped

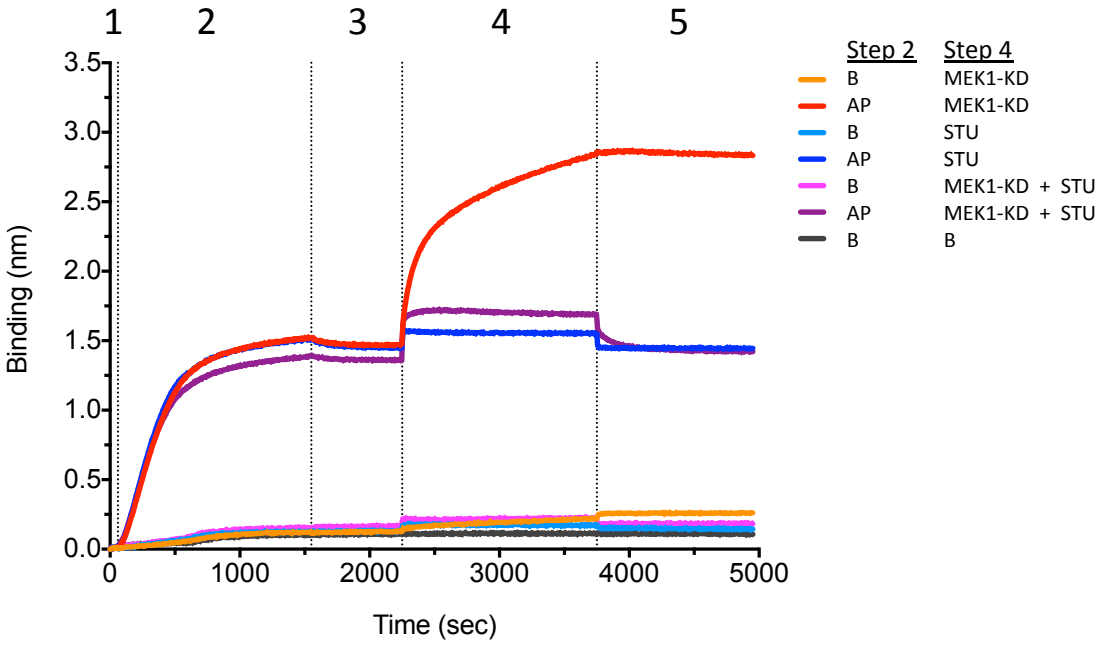


Figure 3.12 Determining the specificity of the AP using the Octet.
Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for eight independent sensors incubated with AP (1 µM) and MEK1-KD protein (1 µM), STU (1.2 µM) or inhibition of MEK1-KD with STU (1.2 µM). Step 1 represents baseline determination; Step 2 represents incubation in AP or buffer (loading); Step 3 represents incubation in fresh-buffer (disassociation); Step 4 represents incubation in STU, MEK1-KD or STU-inhibited MEK1-KD protein (association); and Step 5 represents incubation in fresh-buffer (disassociation). The figure legend details Steps 2 and 4 for each curve. AP=Affinity Probe, B=buffer.

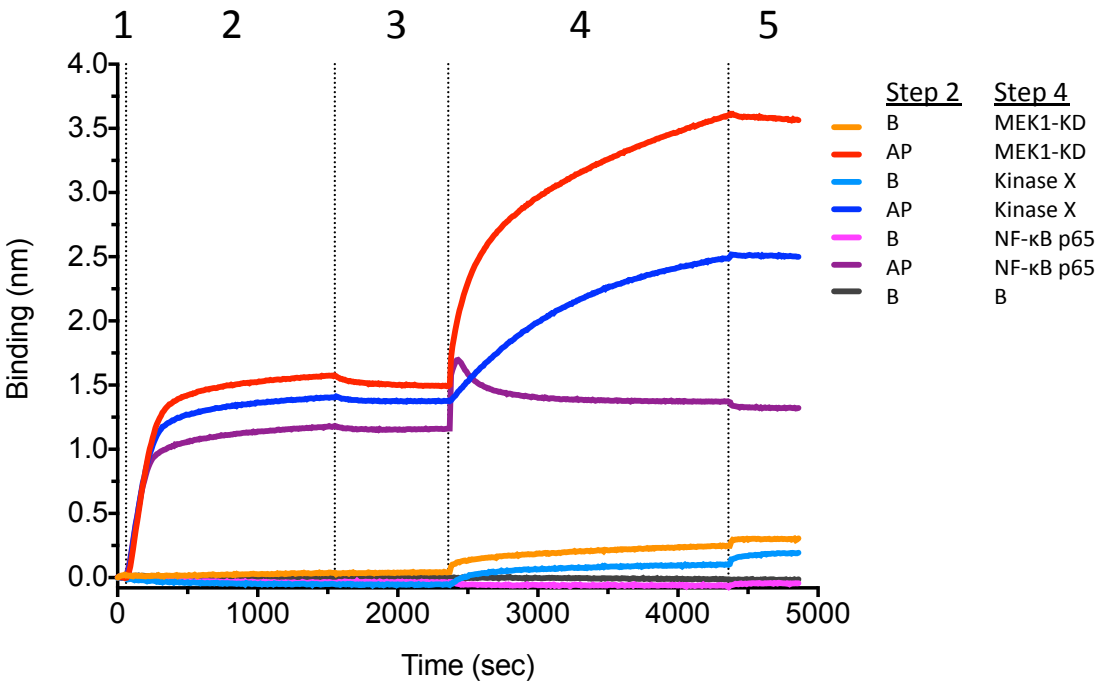


Figure 3.13 The AP binds specifically to protein kinases on the Octet.
Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for seven independent sensors incubated with AP (1 µM) and MEK1-KD, Kinase X or NF-κB p65 proteins (1 µM). Step 1 represents baseline determination; Step 2 represents incubation in AP or buffer (loading); Step 3 represents incubation in fresh-buffer (disassociation); Step 4 represents incubation in protein samples (association); and Step 5 represents incubation in fresh-buffer (disassociation). The figure legend details Steps 2 and 4 for each curve. AP=Affinity Probe, B=buffer.

into clarified cell lysates, where MEK1-KD had been expressed, or into lysates where no recombinant protein was present [both containing 2% (v/v) DMSO and 0.05% (v/v) Tween-20] (**Figure 3.14**). Unfortunately there was no discernable difference between the obtained binding signals. Movement of the sensors into fresh buffer also resulted in very similar dissociation profiles. Furthermore, spiking the cell lysate with previously purified MEK1-KD (1 μM), also did not affect either the observed association or dissociation profiles (data not shown). Together, these experiments indicated that there was a significant amount of non-specific or background binding to the modified-sensors, when cell lysates instead of purified proteins were used.

To determine if this high-level of background could be reduced by the inclusion of detergent, a number of different concentrations of Tween-20 were tested, using the same two lysates as described above. However, despite the inclusion of 1% (v/v) Tween-20 in experimental buffers, no improvement in the level of non-specific binding was seen (**Figure 3.15**).

3.9 Octet-based assay with auto-induction cultures

Although these initial results were discouraging, it was noted that the experimental protocol was somewhat different to that normally used in CDH; i.e. cell lysates were prepared from a 50 ml culture induced with IPTG, whereas 4 ml cultures grown in auto-induction medium are used during CDH. Auto-induction medium allows a much higher cell-density to be achieved and thus it is often possible to gain a substantially higher yield of recombinant protein. To see if this was an influencing factor, seven different protein kinase domains [BRaf V600E, MEK1-KD and Chk2-KD, and four different constructs of NIK (NF-κB-Inducing Kinase)] and five non-kinases proteins (p50, GST, NF-κB p65, NF-κB p65 RHR and C/EBPβ) (section **2.2.1**) were tested in the Octet assay. Each protein construct was expressed in 4 ml auto-induction culture, and was lysed following the standard CDH lysozyme lysis protocol, with lysate clarification by centrifugation (section **2.12**). Again, however, there were no notable differences between the binding profiles for lysates containing the protein kinase domains, and those for the other non-kinase proteins (**Figure 3.16A**; aligned **Figure 3.16B**). Samples were also

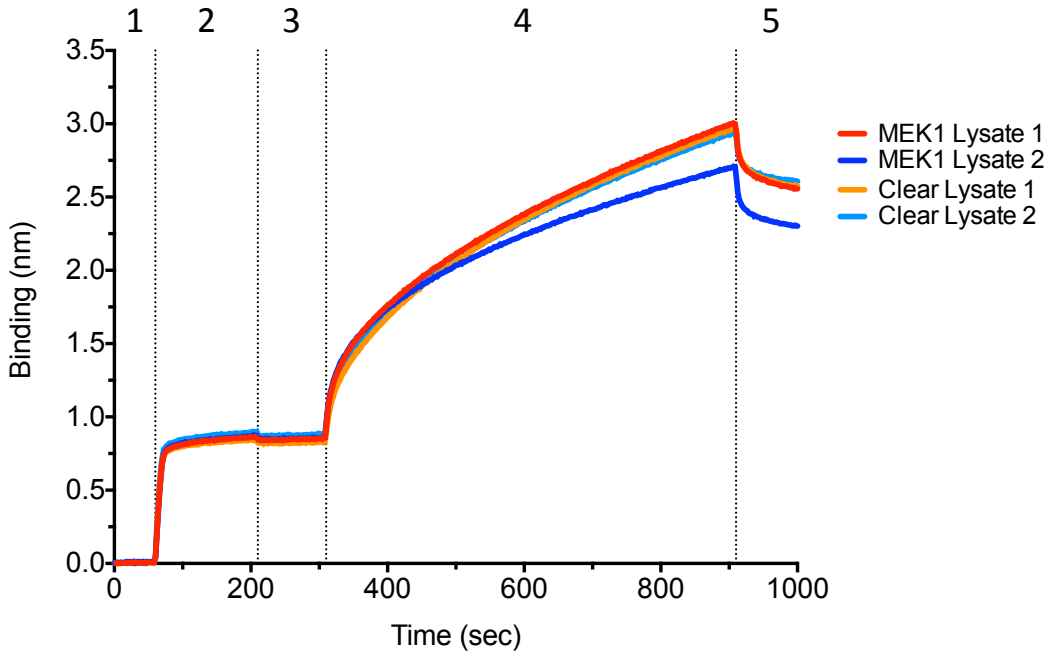


Figure 3.14 Octet assay development in a cell lysate background.
Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for four independent sensors incubated with AP (1 μM) and MEK1-KD containing or clear cell bacterial lysate clarified supernatants. Step 1 represents baseline determination; Step 2 represents incubation in AP (loading); Step 3 represents incubation in fresh-buffer (disassociation); Step 4 represents incubation in supernatant (association); and Step 5 represents incubation in fresh-buffer (disassociation). The figure legend details Step 4 for each curve.

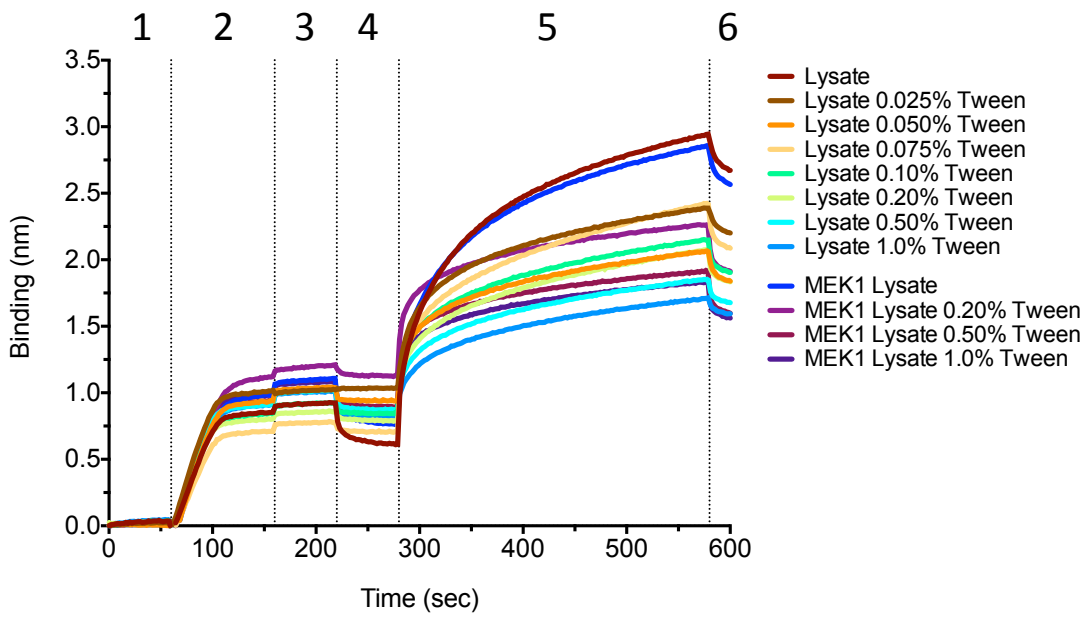


Figure 3.15 Octet assay development in a cell lysate background with increasing the detergent concentration.
Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for 12 independent sensors incubated with AP (1 μM) and MEK1-KD containing or clear cell bacterial lysate supernatants with Tween-20 concentration as specified. Step 1 represents baseline determination; Step 2 represents incubation in AP (loading); Step 3 represents incubation in fresh-buffer (disassociation); Step 4 was a buffer exchange step to the relevant Tween-20 concentration; Step 5 represents incubation in clarified supernatant (association); and Step 6 represents incubation in fresh-buffer (disassociation). The figure legend details Step 5 for each curve.

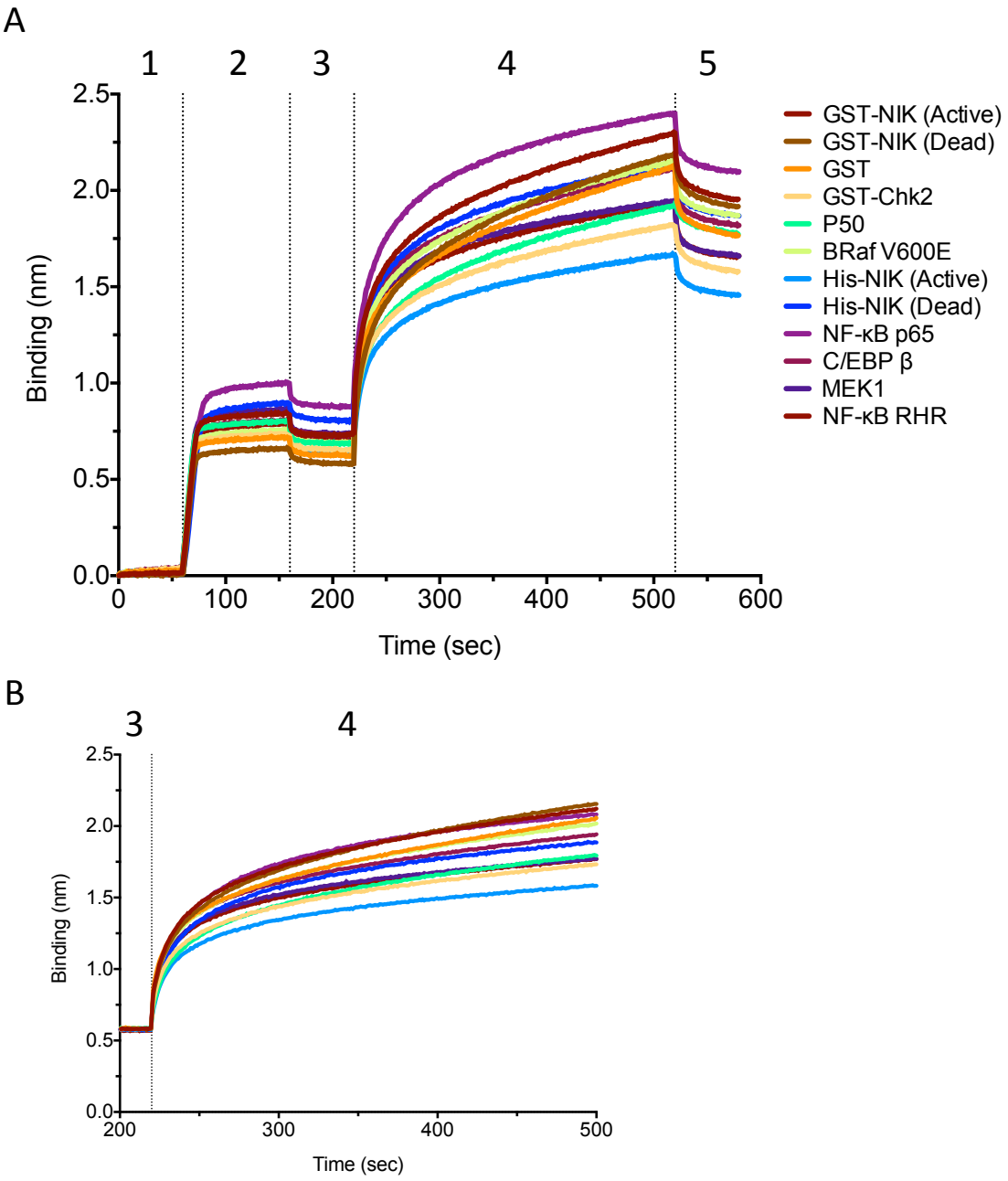


Figure 3.16 The Octet AP was unable to detect specific protein kinase binding in a cell lysate background by following a typical CDH protocol.
(**A**) Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for 12 independent sensors incubated with AP (1 µM) and recombinant protein containing bacterial lysate supernatants (as indicated). Step 1 represents baseline determination; Step 2 represents incubation in AP or buffer (loading); Step 3 represents incubation in fresh-buffer (disassociation); and Step 4 represents incubation in clarified supernatant (association). The figure legend details Step 4 for each curve. (**B**) Alignment of '**A**' at Step '3'; curves are as described in '**A**'.

prepared as above, but using sonication for cellular lysis, however as before, there were no notable differences between the binding profiles. The use of cell lysates and AP-modified sensors for the detection of constructs expressing functional, folded, kinase domains was deemed to be unsuccessful, due to the high levels of background (non-specific) signal observed; therefore it was investigated as to whether this Octet methodology could be applied to one-step purified proteins.

3.10 Octet-based assay using purified proteins

For modification of this assay, experiments using proteins pre-purified by a one-step affinity chromatography step were carried out. This modification of the method enables the target proteins to be available in a buffer suitable for binding to the modified sensors, and where non-specific background binding could hopefully be eliminated or drastically reduced. In this format, the assay detected binding for five of the different protein kinases tested (**Figure 3.17A**; aligned **Figure 3.17B**) with little, if any, non-specific background binding. Robust binding was observed for MEK1-KD, His_6-NIK (kinase dead), Chk2-KD and BRaf V600E, and weak-binding for His_6-NIK (kinase active). No binding was seen for the GST-tagged version of NIK (kinase dead) or for the other (non-kinase) proteins tested. To ensure that protein had been indeed been expressed in each of the samples tested, there were all analysed by SDS-PAGE. High levels of recombinant protein could be seen for five of the proteins tested – P50, GST, Chk2-KD, NF-κB p65, C/EBPβ and MEK1-KD (**Figure 3.17C**). This result indicated a very high sensitivity to the Octet assay by the detection of functional protein kinases that were not all identifiable by SDS-PAGE analysis. Furthermore, this was able to distinguish between highly-expressed protein kinases and non-protein kinases (only identifying protein kinases) that were detectable by SDS-PAGE analysis.

3.11 Summary

The work presented in this chapter aimed to develop a high-throughput technique to assess the functionality of proteins expressed by CDH clones. In particular, to develop a method that could be introduced at an early time-point in the CDH methodology, in order to minimise the down-stream handling of multiple clones, which may not in fact express a functional, folded protein. The current CDH

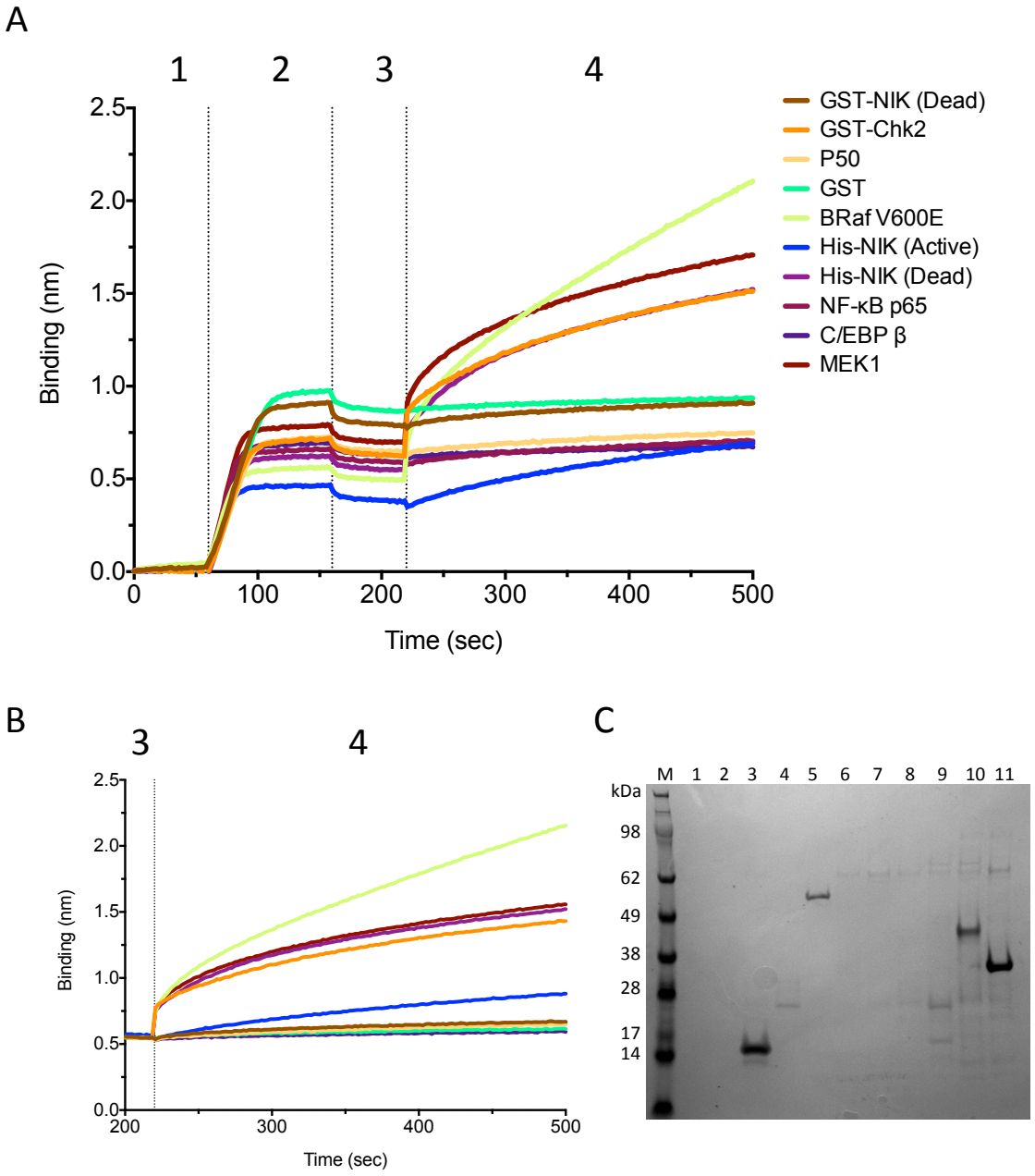


Figure 3.17 Octet AP binding in IMAC purified proteins.
(**A**) Sensors were pre-hydrated for 30 minutes in DTM buffer. The experiment was carried out at room temperature (~ 20 ˚C) with a mixing speed of 1000 rpm. Overlaid sensorgrams are shown for 12 independent sensors incubated with AP (1 µM) and one-step affinity-tag purified proteins (as indicated). Step 1 represents baseline determination; Step 2 represents incubation in AP or buffer (loading); Step 3 represents incubation in fresh-buffer (disassociation); and Step 4 represents incubation in supernatant (association). The figure legend details Step 4 for each curve. (**B**) Alignment of '**A**' at Step '3'; curves are as described in '**A**'. (**C**) One-step affinity-tag purified proteins used in '**A**'. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue. M=molecular mass marker, 1=GST-NIK (kinase active), 2=GST-NIK (kinase dead), 3=P50, 4=GST, 5=Chk2-KD, 6=BRaf V600E, 7=His-NIK (kinase active), 8=His-NIK (kinase dead), 9=NF-κB p65, 10=C/EBPβ, 11=MEK1-KD.

protocol has been highly optimised, but still typically takes four days to process 750 different samples – with the most time-consuming process being the analysis of samples by SDS-PAGE. A specific probe (AP) was designed, synthesised and tested, to be used as a screening molecule; composed of the broad-spectrum kinase inhibitor STU coupled to biotin for interaction with a streptavidin matrix via a PEG_4 linker (STU-PEG_4-BIOTIN). The ForteBio Octet Red96 was chosen as the platform for these experiments, using BLI to detect binding by changes in optical thickness at the sensor tip. Using this methodology, we were able to detect binding for several pre-purified protein kinases. The selectivity and the specificity of the AP (for protein kinases), was also confirmed using pull-down experiments. However, attempts to apply the same methodology to screening proteins in the context of cell lysates was unsuccessful, due to the high level of background signal that prevented the discrimination between lysates that contained protein kinases and those that did not. A number of different lysis conditions were also screened, as well as increasing the detergent concentration, but these were shown to have no effect on the overall binding signal and the ability to identify protein kinase binding.

The molecule biotin is synthesised in *E. coli* (Lin and Cronan, 2011), and this would bind to any unbound streptavidin on the sensor tip, resulting in a change in binding signal, however, the sensors were saturated using the AP prior to running the assay, therefore this would be unlikely. Furthermore, the binding of functionally folded protein (over-expressed in the cell lysate) to the biosensor should cause a more significant change in the binding signal than the binding of biotin, due to the size difference and the change in the reflected white light signal (refer to **Figure 3.2**). These results indicated there was either a non-specific interaction occurring at the sensor tip from the cell lysate, or the increased optical density of the cellular lysate was perturbing the reflected white light signal. As mentioned previously, the Octet system has been successfully used to screen expression constructs directly from cell lysates containing His_6 affinity tagged recombinant protein using His_6-biosensors (Tian-Yu et al., 2010); however, there is no material published on the use of the streptavidin coated biosensor, presumably due to the presence of biotin in cellular lysates.

Following this, experiments were carried out using proteins pre-purified by one-step IMAC. This modification to the method allowed the successful identification of recombinant protein kinase domains, and was able to distinguish these from controls, comprised of other non-kinase proteins. This particular methodology development does not reduce the number of expression constructs (identified by CDH) that need to be handled in parallel (the initial aim of the experiments presented in the chapter). However, it does allow the rapid identification of constructs that express functional, folded (kinase) domains. In this version of the Octet method, recombinant protein (from cultures grown in auto-induction media) must first be pre-purified by a one-step IMAC purification step, which typically takes around four hours to process two 96-well blocks (192 protein samples). For the Octet-based assay, the sensors would be pre-bound with AP prior to the assay to enable saturation of the sensors and to streamline to assay. For the assay itself, a full 96-well plate could be used where; column 1 would contain assay buffer (AP-coated sensor wash step), and columns 2 to 11 would contain CDH IMAC eluted samples (therefore 80 protein samples per plate), and column 12 would contain assay buffer in order to measure dissociation (**Figure 3.18**). The Octet would be pre-programmed to run through this assay continuously, with a new biosensors used for each sample, moving from buffer, to an IMAC eluted sample, and back to buffer – this would take a roughly 5 minutes per run. Analysis of 80 protein samples would take roughly 60 minutes, and if these were prepared back-to-back these would take roughly 2.5 hours to analyse all 192 samples. However as the Octet runs at room temperature it would be best to split the analysis down, to ensure sample stability. To fully analyse a CDH expression screen of 750 samples using the Octet, this would take 10 hours. Subsequently, 'hit' samples would require downstream processing to determine the sample quality and the identity of the constructs.

The cost of an individual modified probe is £3.76 (ForteBio, Pall Europe); therefore the approximate cost to screen 750 constructs is over £2800. It is technically feasible to regenerate the modified sensors, such that they can be used again,

	1	2	3	4	5	6	7	8	9	10	11	12
A	Buffer	1	9	17	25	33	41	49	57	65	73	Buffer
B		2	10	18	26	34	42	50	58	66	74	
C		3	11	19	27	35	43	51	59	67	75	
D		4	12	20	28	36	44	52	60	68	76	
E		5	13	21	29	37	45	53	61	69	77	
F		6	14	22	30	38	46	54	62	70	78	
G		7	15	23	31	39	47	55	63	71	79	
H		8	16	24	32	40	48	56	64	72	80	
		Protein samples										

Figure 3.18 A schematic of the developed methodology.
Sensors pre-loaded with AP would be moved through this plate, where the for the first run, eight AP-loaded sensors would move sequentially into buffer (wells A1-H1), then into protein samples 1-8 for association (wells A2-H2) and finally into buffer for dissociation (wells A12-H12). For each row of protein samples, a new set of sensors would sample the protein interaction of that row.

however, attempts to do this, resulted in very inconsistent binding signals and increased in background signal, even with highly purified protein samples.

Considering these factors, this particular screening method whilst useful, was not cost- or time-effective, and therefore was not developed or taken forward. Additional types of assays were considered, including the ELISA-like assay explained in **section 3.2**, but these were not desirable assay formats as described in the brief for the screen.

Chapter 4

Combinatorial Domain Hunting: Claspin

4.1 Introduction

The tertiary structure of a protein is linked to the function of the protein, and this is more highly conserved than the sequence of a protein; hence the requirement for the determination of the tertiary structure for the determination of the functionality of a protein. Structural genomics aims to solve the tertiary structure of all the proteins encoded within a selected genome, using both computational and experimental approaches. However, the accurate prediction of a domain or the structure, within a given protein sequence, requires some degree of amino acid sequence similarity / homology to other proteins where the structure is already known. Furthermore, as described in section **1.2.2**, sub-constructs of a target gene often are more tractable in yielding soluble, stable protein that is correctly folded and these can be identified through high-throughput bioinformatics or fragmentation / mutation library approaches.

A vast number of proteins (and ORFs) in the human genome have not yet been characterised, in terms of their component domains, due to their lack of sequence homology to other proteins (Prodromou et al., 2007). Prediction of a protein structure *ab initio* is possible, but currently only suited to smaller proteins, and can take many months of computational time to generate models. Lack of domain information, especially when working with very large proteins that potentially contain multiple domains, makes it very difficult to rationally design expression constructs in order to produce sufficient quantities of soluble protein for biochemical or biophysical analysis. This area of domain hunting is especially suited to fragmentation library approaches, where there isn't *a priori* information on the domain structure of the protein.

4.2 Target selection: Claspin

We wished to assess whether the CDH technique (described in section **1.5**) could be used for proteins with low amino acid sequence identity to other proteins, or that had no readily identifiable component domains – in order to produce soluble recombinant protein fragments suitable for downstream analyses. Previously, the fragmentation library technique ESPRIT (section **1.4.10**) was used to characterize

the PB2 (Polymerase basic protein 2) domain of the human influenza A polymerase. This was a novel protein with no homologues and where previous rational construct design had failed to design optimal expression constructs. The ESPRIT methodology resulted in the identification of novel protein folds within the sequence (Tarendeau et al., 2007, Guilligay et al., 2008, Tarendeau et al., 2008).

The objective of the lab of Prof Laurence Pearl and Dr Antony Oliver, is the structural and functional characterization of DNA damage and repair proteins. I therefore focused my attention on an annotated list of DNA damage repair proteins, kindly provided by Dr Frances Pearl (University of Sussex) [since published (Pearl et al., 2015)]. From the original list of 335 proteins (published as 450 proteins), all of the proteins for which there was already a deposited structure of greater than 50% sequence identity were eliminated, leaving 114 proteins. Proteins were also eliminated if they had small molecule inhibitors — as this also indicated that they contained a known domain / fold, which left 100 potential targets. By manual inspection of UniProt entries (Consortium, 2015), targets that either contained known domains, or were too short in length to be a practical target for the CDH methodology were also eliminated. The protein Claspin (UniProt ID: Q9HAW4) was eventually selected, as it had very few bioinformatic annotations. Further, Claspin aligned with the research being carried out in the Pearl / Oliver lab; a number of the proposed interactions of Claspin are also targets of the group.

The protein product of the CLSPN gene (Claspin) was originally identified through its interaction with Chk1 in *Xenopus* (Kumagai and Dunphy, 2000). Later experiments showed it to be necessary for DNA repair, as part of the replication checkpoint ATR-Chk1 signalling pathway, by mediating the phosphorylation of Chk1 by ATR (Kumagai and Dunphy, 2000, Kumagai and Dunphy, 2003, Chini and Chen, 2003, Lin et al., 2004, Clarke and Clarke, 2005, Lindsey-Boltz et al., 2009). Subsequently, Claspin has also been shown to be necessary for DNA replication facilitating high-rates of replication fork progression, by regulating fork stability during unperturbed replication, and stabilising replication forks during replication

arrest (Lee et al., 2003, Lin et al., 2004, Petermann et al., 2008, Scorah and McGowen, 2009, Yoshimura et al., 2011).

Claspin is described as a large, highly charged, protein 'adaptor' that has been shown to interact with a number of DNA-replication and -repair specific proteins (Kumagai and Dunphy, 2003, Jeong et al., 2003, Lee et al., 2005, Brondello et al., 2007, Kim et al., 2008, Gold and Dunphy, 2010, Nakaya et al., 2010, Uno and Masai, 2011, Serçin and Kemp, 2011, Rainey et al., 2013, Broderick et al., 2013, Yuan et al., 2014). Claspin is highly conserved in higher eukaryotes, however the degree of sequence homology varies between species. A multiple amino acid sequence alignment of Claspin from four different species; human, *Xenopus*, chicken and mouse, indicates roughly 50% sequence homology (**Appendix 1.1**) (Kumagai and Dunphy, 2000)**.** A functional homologue, Mrc1 has been identified in the yeasts, although the amino acid sequence conservation is relatively poor (**Appendix 1.2**) (Alcasabas et al., 2001, Tanaka and Russell, 2001). Claspin does not contain known functional domains or regions of structural homology to any previously solved proteins (Kumagai and Dunphy, 2000), but a HTH motif is predicted in the N-terminus of Claspin (aa 279-313) (Zhao and Russell, 2004) and DNA binding functionality has been described in the N-terminus of the protein (Sar et al., 2004, Zhao and Russell, 2004, Lee et al., 2005, Serçin and Kemp, 2011, Uno and Masai, 2011 and Yilmaz et al., 2011). The only structural information to date for human Claspin is a low-resolution EM image, using rotary shadowing, which indicates that Claspin has a ring-shaped structure when bound to branched DNA structures (Sar et al., 2004) (refer to **Figure 1.8**).

It has been reported that full-length recombinant human Claspin can be expressed and purified from mammalian cells, albeit at low yield and was highly prone to degradation – the recombinant protein was biochemically characterised and was found to run aberrantly by SDS-PAGE, and consistently eluted at a smaller retention volume than expected by SEC (Kumagai and Dunphy, 2000, Uno and Masai 2011). The full-length protein and smaller fragments have also been expressed in and purified from different heterologous hosts including *E. coli* and insect cells, in order to study protein-protein and -DNA interactions (Kumagai and

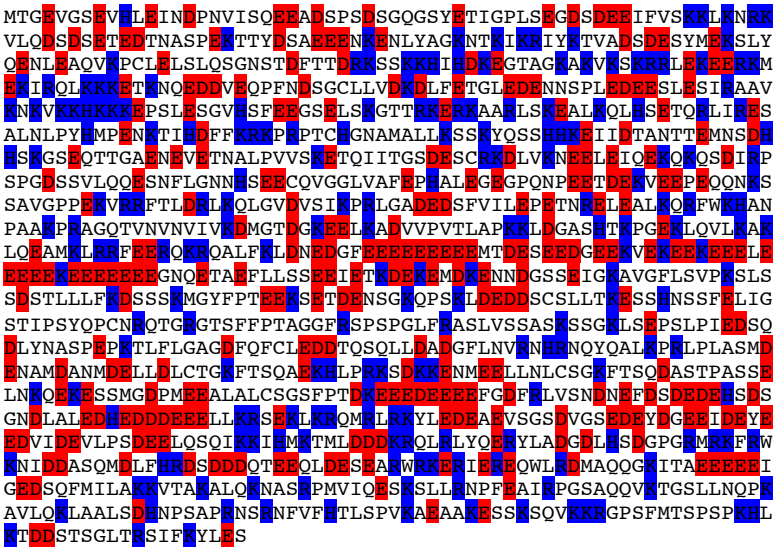
Dunphy, 2003, Jeong et al., 2004, Sar et al., 2004, Clarke and Clarke, 2005, Lee et al., 2005, Yoo et al., 2006, Chini et al., 2006, Lindsey-Boltz et al., 2009, Gold and Dunphy, 2010, Uno and Masai, 2011, Serçin and Kemp, 2011). However, the recombinant protein/s were found to be unstable and not expressed in sufficient quantities for structural biology approaches (Sar et al., 2004, Uno and Masai 2011, Serçin and Kemp, 2010).

4.2.1 Bioinformatic analyses

Initial bioinformatic analysis of the amino acid sequence of human Claspin was performed, to test what had already been published in the scientific literature. The amino acid sequence composition of Claspin indicates that it is a highly charged protein, with a high proportion of negatively (Asp and Glu; 22%) and positively (His, Arg and Lys; 16%) charged residues. In general, the charged residues are evenly dispersed throughout the protein, although there are a number of small sections that contain solely negative or positive charge (**Figure 4.1A**). Additionally, there are a large number of Ser and Thr residues, which could potentially be phosphorylation sites, whilst there are relatively few aromatic residues (**Figure 4.1B**). Amino acid compositional bias can affect the ability of a protein to fold in solution. (Hansen et al., 2006). Claspin is rich in charged residues and also has a high percentage of amino acids, which are commonly associated with disorder (Arg, Gln, Gle, Pro and Ser), whilst having a low percentage of amino acids associated with order (Cys, Iso, Leu, Phe, Trp, Tyr and Val). A web-based PSI-BLAST (Basic Local Alignment Search Tool) search for proteins with similar amino acid sequences to human Claspin did not produce 'hits' with high Z-scores, apart from Claspin itself and/or Mrc1 from other species.

Investigation of the UniProt entry for Claspin (Q9HAW4) identified three sequence repeats (between aa 910-985) termed the CKB motif 1, 2 and 3 (described in section **1.10.4** and **Chapter 7**); these are required for Chk1 interaction with Claspin. This entry also identifies two coiled-coils (CC): CC1 (aa 162-196) and CC2 (aa 592-681). CC2 is roughly in the centre of the human Claspin protein, and is mostly formed of a poly-glutamic acid (poly-Glu) repeat (Consortium, 2015).

A



B

Amino acid	Code	Total No.	Content (%)
Alanine	A	71	5.3
Arginine	R	60	4.5
Asparagine	N	57	4.3
Aspartic acid	D	104	7.8
Cystine	C	11	0.8
Glutamine	Q	62	4.6
Glutamic acid	E	194	14.5
Glycine	G	69	5.2
Histidine	H	28	2.1
Isoleucine	I	37	2.8
Leucine	L	118	8.8
Lysine	K	128	9.6
Methionine	M	26	1.9
Phenylalanine	F	42	3.1
Proline	P	58	4.3
Serine	S	139	10.4
Threonine	T	62	4.6
Tryptophan	W	4	0.3
Tyrosine	Y	18	1.3
Valine	V	51	3.8

Figure 4.1 Amino acid composition analysis of human Claspin.
(**A**) The amino acid sequence of human Claspin (Uniprot ID: Q9HAW4) coloured according to charge. Negatively charged residues are coloured red (D: Aspartic acid, E: Glutamic acid) and positively charged residues are coloured blue (R: Arginine, K: Lysine, H: Histidine). (**B**) Human Claspin sequence composition analysis (Gasteiger et al., 2005).

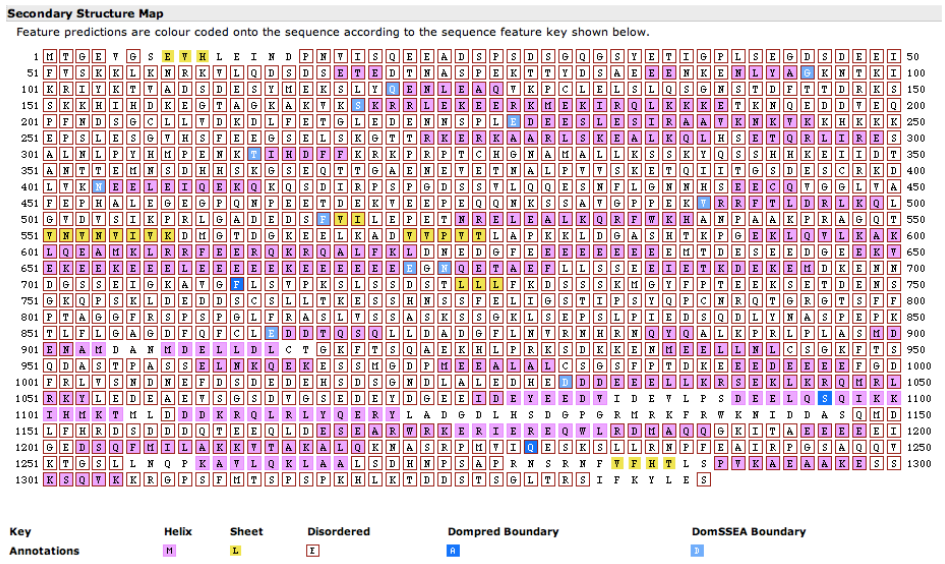
Using the web-based services Phyre^2 (Kelley and Sternberg, 2009) and PsiPred (Jones, 1999, Buchan et al., 2013) secondary structure predictions were generated for the human Claspin sequence (**Figure 4.2** and **Figure 4.3A**). Both programs use PSI-BLAST (Altschul et al., 1997) and then proprietary algorithms to predict the presence of secondary structure elements. Both programs indicated that Claspin is predominantly comprised of α-helices, but with some short regions of β-strand. However, there is some discrepancy between the two predictions, presumably due to the differing algorithms. There are also several regions with low confidence scores for secondary structure – as indicated by the coloured bar below the Phyre^2 output – likely due to the lack of sequence identity between Claspin and other proteins of known structure / fold. Both programs indicate that large segments of Claspin have no secondary structure, and are instead disordered in nature, as indicated by question marks below the sequence on the Phyre^2 output, and red boxes around residues on the PsiPred output. Using a separate web-based bioinformatics tool, DisoPred (Buchan et al., 2013) regions of predicted disorder were identified: a probability value of '0' indicates the likelihood of secondary structure, whilst a score of '1' indicates a region of disorder. The horizontal dashed line shown in **Figure 4.3B** indicates the order / disorder threshold. The Claspin sequence has a high disorder score (i.e. above the threshold) with only some potential order / fold around residues ~1100 to 1200 (**Figure 4.3B**).

Furthermore, tertiary structural predictions were identified for Claspin in existing databases from template-based models. Two template-based models, one for almost full-length protein (aa 19-1166) and a second for the C-terminus of Claspin (aa 678-1339) were found. The first model for almost full-length Claspin, produced by ModBase (Pieper et al., 2011; http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi) is based on the RecBCD:DNA complex (PDB: 1W36) a helicase and nuclease multi-enzyme in *E. coli* involved in DSB end processing (Singleton et al., 2004), but this only has 14% sequence homology to the Claspin template. This predicts there are distinct globular regions in the protein, but with large regions of low structural complexity connecting these together. The globular regions are mostly composed of α-helices with a number of small regions of β-sheet within these (**Figure 4.4A**). The second model, produced by DomSerf and annotated on

Figure 4.2 Secondary structure prediction for human Claspin generated by Phyre².

Regions predicted to form α-helical elements are represented by the green coloured helix, whilst a blue arrow represents those regions predicted to form β-strands. Regions of low amino acid sequence complexity are indicated by a question mark. The lower bar, coloured in a spectrum from red through to blue, represents confidence in the predictions, which indicates high to low scores respectively (refer to diagram key) (Kelley and Sternberg, 2009).

A



B

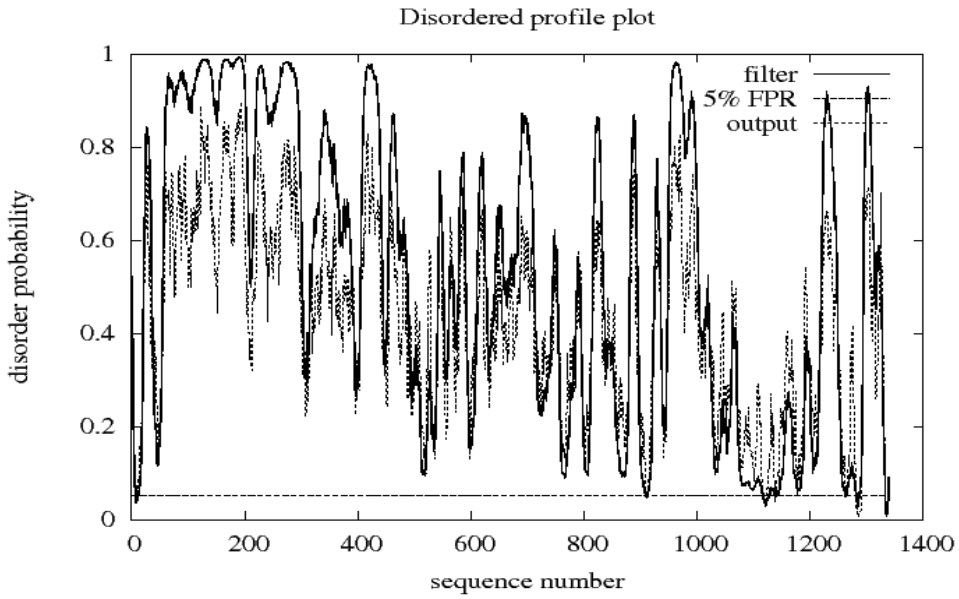


Figure 4.3 Secondary structure and disorder predictions for human Claspin generated by PsiPred.

(**A**) Regions predicted to form α-helical elements are represented by the pink coloured boxes, whilst yellow coloured boxes represent those regions predicted to form β-strands. Regions of low amino acid sequence complexity are indicated by a red outline (box) (refer to diagram key) (Jones, 1999, Buchan et al., 2013). (**B**) Sequence disorder prediction for human Claspin. The probability score ranges from '0' (folded) to '1' (unfolded, disordered) (Buchan et al., 2013).

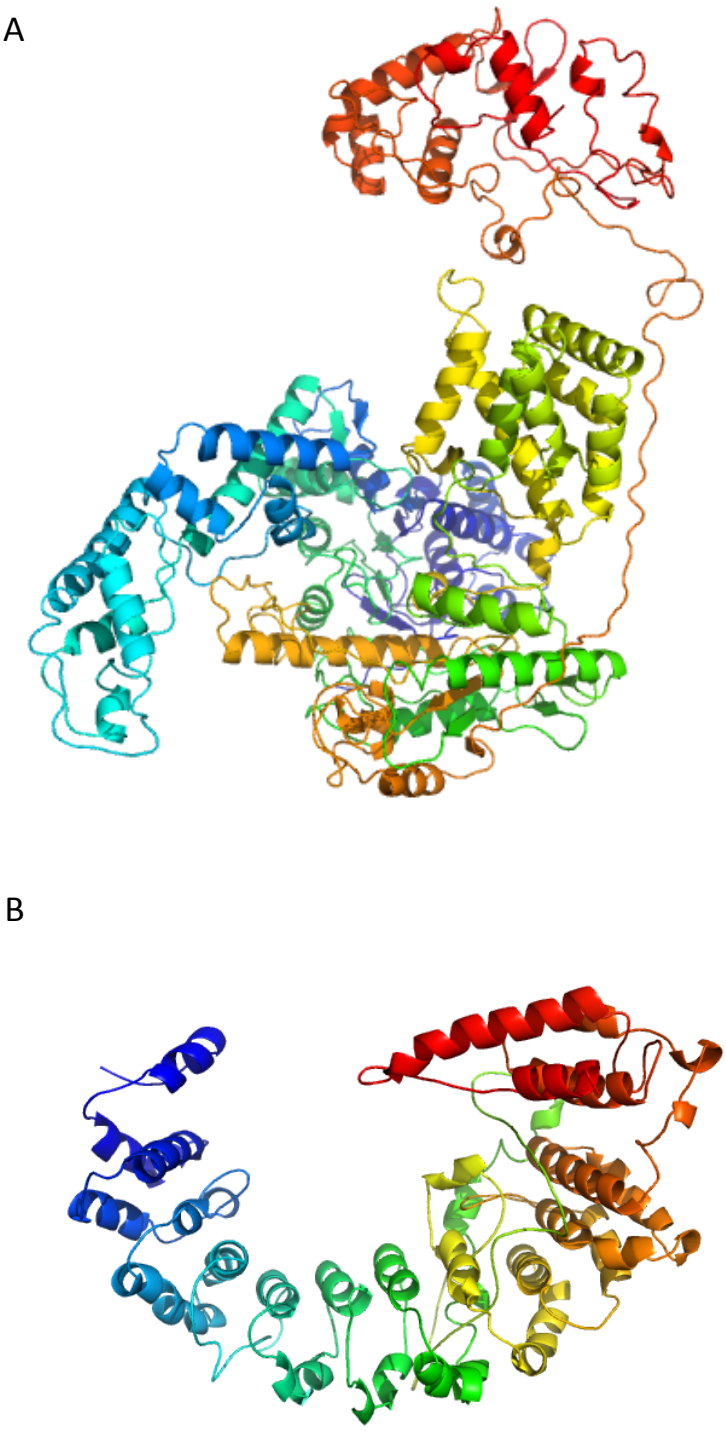


Figure 4.4 Template-based models of human Claspin.
(**A**) Structural prediction for Claspin amino acids 19-1166 from template-based modelling using ModBase, with the RecBCD:DNA complex as a template (PDB: 1W36) (Pieper et al., 2011). (**B**) Structural prediction for Claspin amino acids 678-1339, from template-based modelling using DomSerf with Karyopherin β2 as a template (PDB: 1QBK). (Lewis et al., 2013, Buchan et al., 2013). Models are shown in molecular cartoon representations, coloured from blue to red, from the N- to C-terminus respectively (Jones' Rainbow).

Genome3D (Lewis et al., 2013, Buchan et al., 2013; http://genome3d.eu/uniprot/id/Q9HAW4/annotations) represents the C-terminal region of Claspin, and is based on Karyopherin-β2 (PDB: 1QBK), but this only has 21% sequence homology to the Claspin template. Interestingly, this part of Karyopherin-β2 consists of an Armadillo (ARM) -repeat (**Figure 4.4B**). ARM-repeats are composed of a tandem-repeating three α-helix motif that is roughly 42 amino acids in length, which fold to form a super-helical structure. These have diverse but important cellular functions and whilst they are evolutionarily conserved in structure, the sequence conservation is typically low (Tewari et al., 2010). In both cases, the sequence homology between the template protein and Claspin is very low, and therefore the likelihood of this being a true representation of the tertiary structure of Claspin is slight. For a reliable assignment of secondary and tertiary structure, a threshold of > 30% sequence homology between template and target, is required; below this threshold the alignment of amino acid sequence and thus secondary structure elements becomes unreliable (Xiang et al., 2006). Comparison of the two Claspin models highlights this problem, as an overlapping region (aa 678-1166) is not predicted to have the same secondary or tertiary structure.

Additionally, the web-portal SwissModel was used with the intention of generating a homology model for Claspin. This program identifies any sequence homology in the target protein sequence with any other protein in its database, and then tries to identify any component fold / domain (Schwede, 2003). This identified a few potential fold / domain matches for the Claspin sequence, again the level of amino acid sequence homology was generally very poor, meaning that these were unlikely to be reliable models of the sequence. Furthermore, protein threading was performed for the N- and C-terminal regions (aa 1-625 and 625-1339 respectively) of Claspin using the Raptor X server (Kallberg et al., 2012). For the N-terminus of the protein (for aa 326-621), this predicted an α-helical bundle based on the fold template of a three helix bundle of a *de novo* protein (PDB: 4TQL); with 23% α-helix, 3% β-sheet and 72% disordered content (**Figure 4.5A**). For the C-terminus of the protein (for aa 890-1339), this predicted an elongated α-helical coil based on the fold template of tropomyosin, an elongated α-helical protein (PDB: 1C1G);

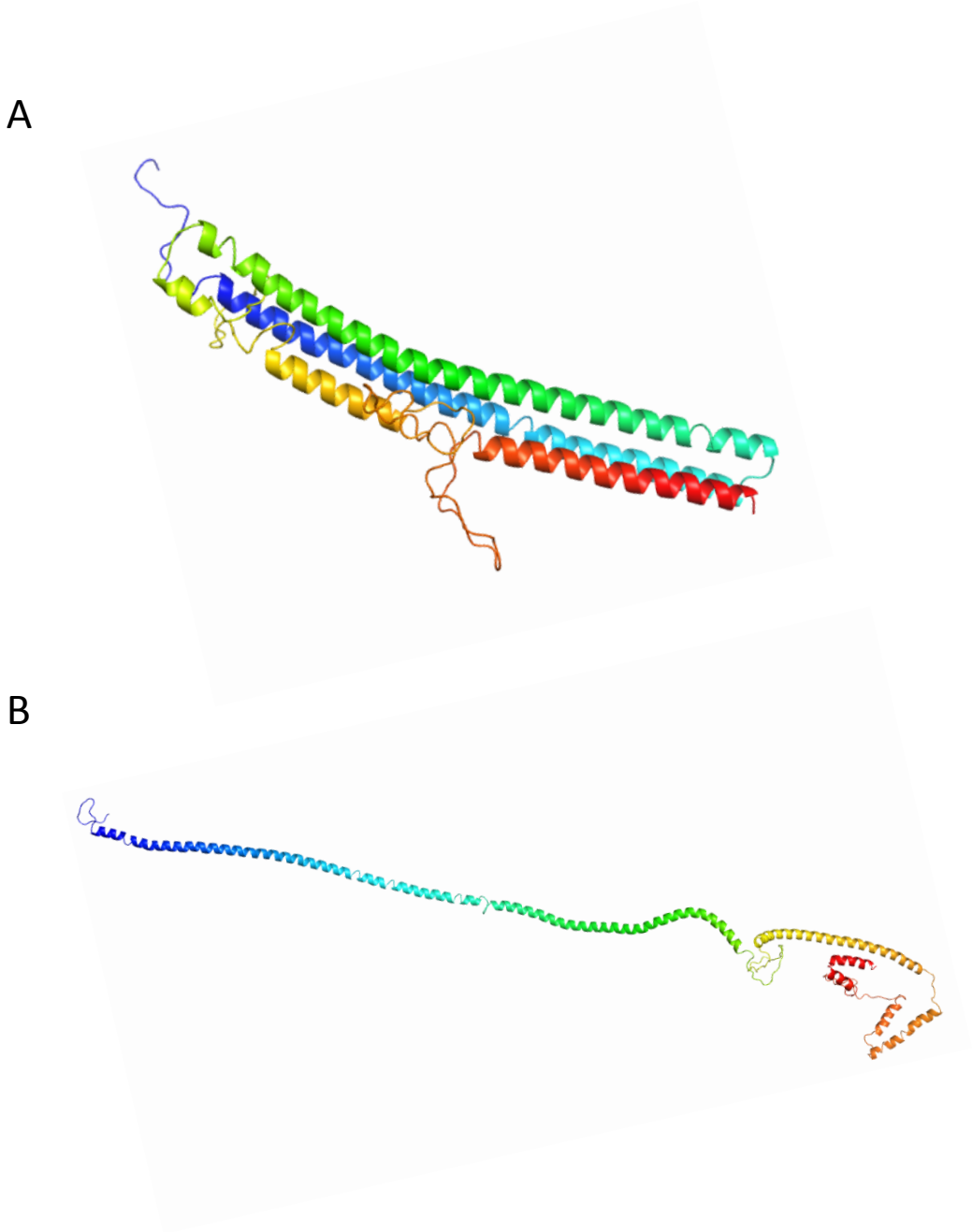


Figure 4.5 Protein threading models of human Claspin.
Structural prediction for Claspin by protein threading; (**A**) for amino acids 326-621 using the template for a de novo three helix bundle protein (PDB: 4TQL), and (**B**) for amino acids 890-1339 using the template of tropomyosin (PDB: 1C1G), predicted using the Raptor X server (Kallberg et al, 2012). Models are shown in molecular cartoon representations, coloured from blue to red, from the N- to C-terminus respectively (Jones' Rainbow).

with 12% α-helix, 0% β-sheet and 88% disordered content (**Figure 4.5B**). *Ab initio* modelling of Claspin was not attempted, as a protein of this size would require sizeable computing power and computational time.

4.3 Production of a random fragmentation library

This section is redacted for reasons of confidentiality.

Figure 4.6 S1 nuclease DNA fragmentation.

This section is redacted for reasons of confidentiality.

4.3.1 Quality assessment

Parts of this section are redacted for reasons of confidentiality.

A colony PCR was carried out on 190 transformed bacterial colonies, taking an average of 16 colonies from 12 plates. The resultant amplicons were then analysed by agarose gel electrophoresis. In total 152 reactions indicated the presence of a insert, of which 54 were greater than 1000 bp in size, with the remaining 98 containing smaller sized inserts (**Figure 4.7A**). 34 reactions also indicated the presence of 2 differently sized inserts – this was most likely due to the difficulty of picking a single colony from the L-agar plate, which were at a very high density, and therefore probably represents some level of cross-contamination. Only 4 reactions indicated the presence of empty vector, giving an amplicon of ~200 bp in size. Excluding the reactions that contained 2 inserts, there was a 97% insertion success rate, with 35% containing inserts > 1000 bp, with the remaining 62% containing inserts < 1000 bp, but still greater than the desired 200 bp cut-off.

35 representative clones (containing inserts) were then selected for sequencing, in order to confirm that the cloned DNA was in fact derived from the parental Claspin synthetic gene (and not from a potential contaminant), this also facilitated a gene coverage analysis. Of the 35 clones, sequencing reactions for 33 indicated an insert derived from Claspin, which could then be aligned against the full-length coding sequence, while no sequence was obtained for the two remaining constructs (**Figure 4.7B**). The coverage analysis showed a uniform distribution of fragment generation across the whole gene, with no apparent regional bias, with an average clone length of 994 bp. However, when visualised as a dot plot, the cloned fragment length does not follow a normal distribution (**Figure 4.7C**); this was due to an unintended bias in selecting longer constructs for sequencing. Subsequently a full sized CDH fragment library was generated.

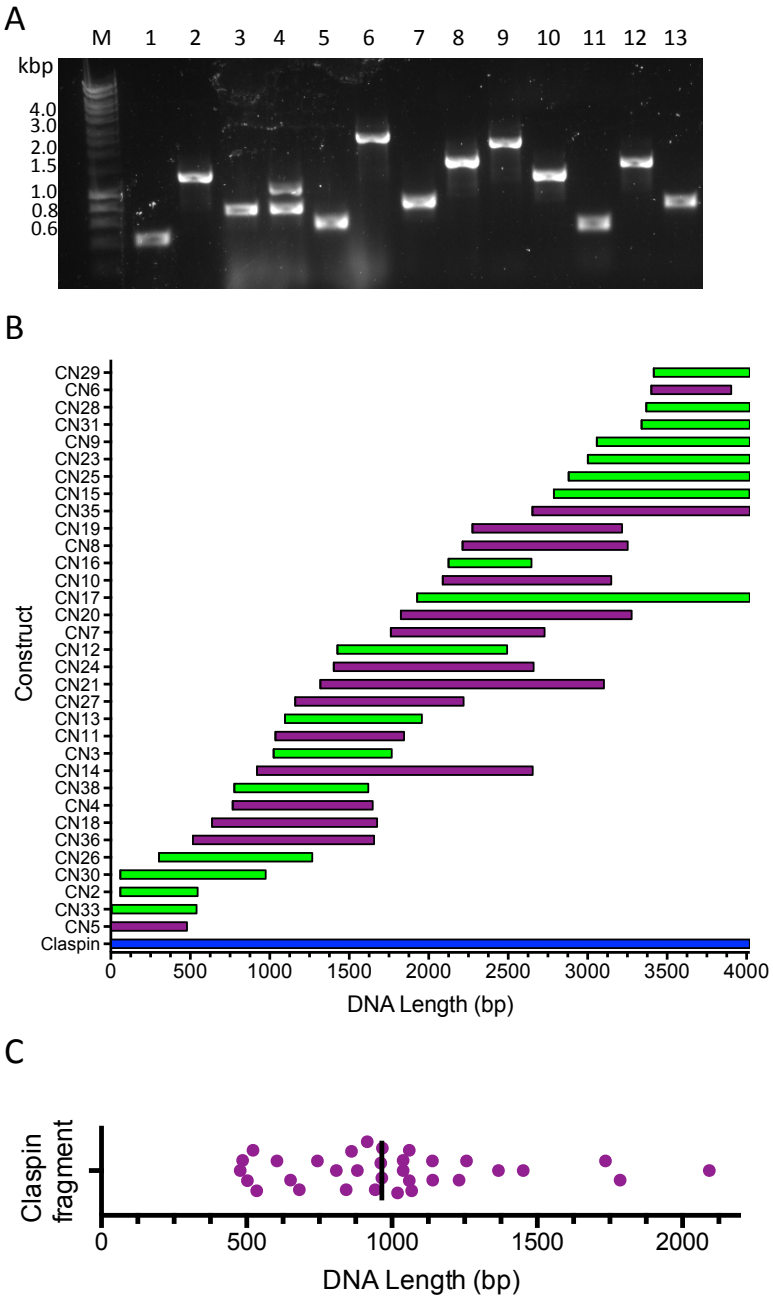


Figure 4.7 CDH fragmentation library quality assessment.
(**A**) Colony PCR analysis of the fragmentation library. A representative 1% (w/v) 1x TAE agarose gel showing amplicons for individual clones. M=Hyperladder I DNA marker, 1 to 13=colony PCR reactions. (**B**) Alignment of the fragment library, relative to the full-length synthetic Claspin gene (blue), as determined by individual sequencing of clones. Constructs in the 'correct' orientation with respect to the start codon of the pDXV4-TOPO vectors are coloured purple, and those in the reverse or 'incorrect' orientation are coloured green. (**C**) Dot plot of cloned fragment size, with a mean value of 994 bp indicated by the vertical black lines.

4.4 Creating and testing the expression library

Parts of this section are redacted for reasons of confidentiality.

To rapidly identify clones that expressed soluble Claspin fragments, colony expression tests were carried out by transformation of the *E. coli* BL21 (DE3) strain. In total, >23,000 transformed colonies were generated. To test protein expression, a colony lift was carried out (with a nitrocellulose membrane) enabling transfer of colonies (colony side up) to a replica L-agar plate containing IPTG to induce recombinant protein expression. After a period of 4 hours, cells were lysed *in situ* on the nitrocellulose membrane, and recombinant protein expression detected using a colourmetric anti-His western blot. The intensity of spot colour is roughly proportional to the level of recombinant His_6 affinity tagged protein expressed (**Figure 4.8A**). Colonies found to have expressed high levels of recombinant protein, were selected for further expression testing and purification. For this screen, 768 colonies were selected.

For micro-scale expression tests, single colonies were used to inoculate auto-induction medium. This medium enables high-throughput cell culture and spontaneous induction of recombinant protein expression once the preferential carbon source of glucose has been exhausted and is switched to lactose (Studier, 2005). Cultures typically have an OD_{600} of between 15 and 25 at harvest.

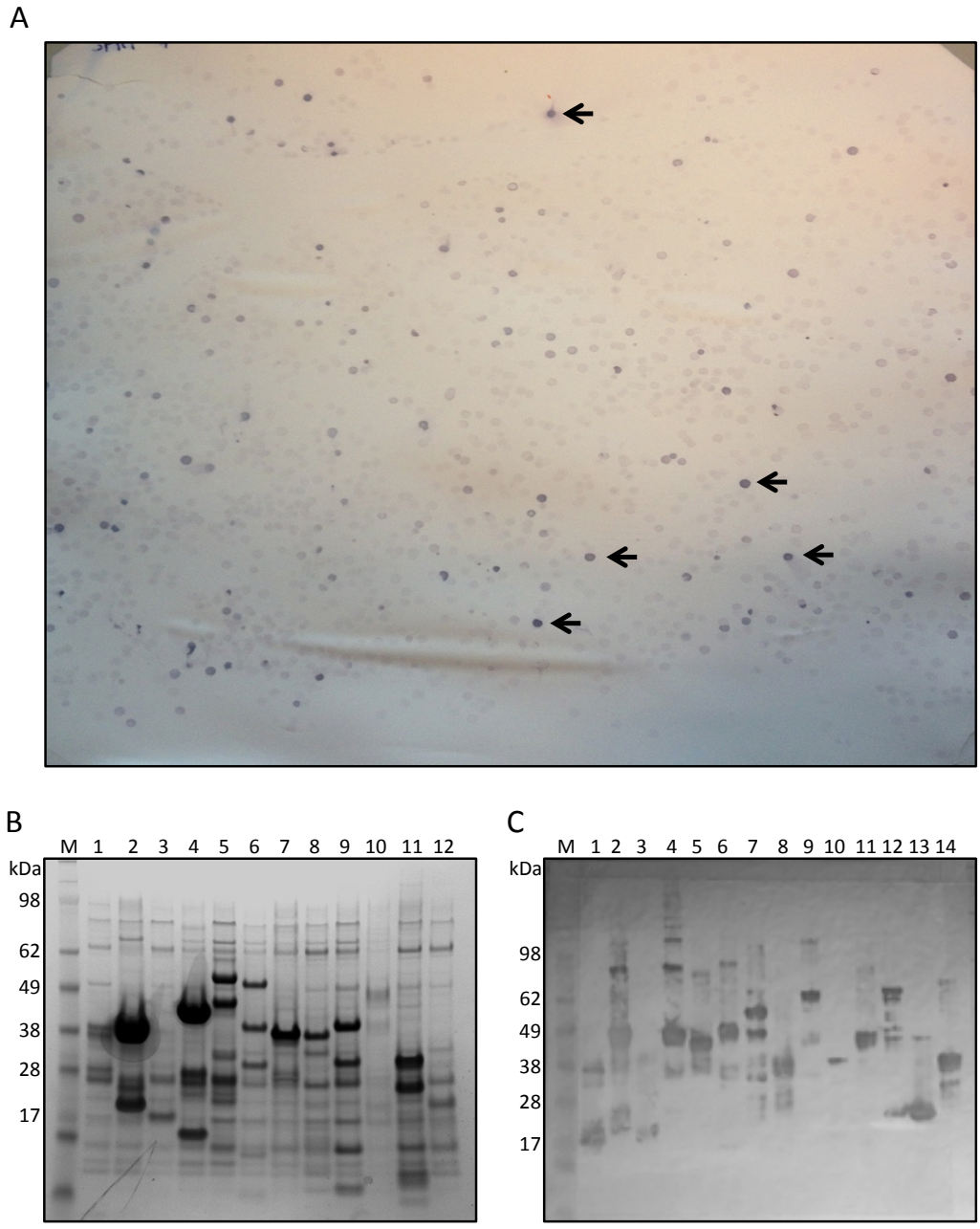


Figure 4.8 Expression library screening.
(**A**) A representative colourmetric colony blot; using an anti-His primary antibody, and an AP-conjugated secondary antibody. Arrows indicate colonies that would typically be selected for further analysis, i.e. those expressing high levels of recombinant protein. (**B**) A representative Instant Blue-stained 4-12% Bis-Tris SDS-PAGE gel analysing eluates from the one-step IMAC purification. M=molecular mass marker, Lanes 1 to 12=representative IMAC eluates (**C**) A representative colourmetric western blot, (using anti-His primary and AP-conjugated secondary antibodies), testing the solubility of recombinant proteins identified from IMAC purification step. Lanes 1 to 14=representative solubility test samples.

Purification of soluble C-terminally His_6 affinity tagged proteins was carried out in a high-throughput 96-well plate format, using a one-step IMAC purification procedure, in this case using Ni-NTA resin and a 96-well filter plate. The eluates from the IMAC plate purification were analysed by SDS-PAGE (**Figure 4.8B** and **Appendix 1.4**). A substantial band can be visualized in a large number of purified samples, which indicates over-expressed recombinant protein. Two further bands were consistently seen co-purifying with the desired recombinant protein in almost every sample, migrating at around 28 and 62 kDa, these are *E. coli* protein contaminants (SlyD and GroEL respectively) (Bolanos-Garcia and Davies, 2006), which are also binding to the Ni-NTA resin.

Expression constructs producing high yields of recombinant protein (as determined by SDS-PAGE) were then taken forward to the solubility screen, as determined by high-speed centrifugation of the eluates. 176 samples (indicated with an arrow in **Appendix 1.4**) were treated in this manner, in order to pellet insoluble protein, and the resulting soluble fraction reanalysed by SDS-PAGE and by anti-His western blotting (**Figure 4.8C** and **Appendix 1.5**). Solubility tests indicated 59 constructs, as being suitable for further study (indicated with an arrow in **Appendix 1.5**), and these were taken forward into the next step of the pipeline, and the DNA was sequenced (a data summary can be found in **Table 4.1**). In total, 42 of these constructs could be successfully mapped against the full-length sequence of the Claspin gene. Of these, 32 were found to be in both the correct orientation and reading frame. The other 10 constructs were found to be "out-of-frame, in-frame" constructs — due to the nature of the expression vector series, the ligation process can sometimes lead to fragments being out of frame at the 5'-end, whilst being in-frame at the 3' end. Of the remaining 17 constructs: 9 were found to contain multiple DNA sequences indicating cross-contamination, 5 were found to be out-of-frame, and 1 was in the inverse orientation. No sequence was obtained for the 2 remaining constructs. Due to the large number of initial 'hits', only 384 of the 768 cultures were eventually purified. The total number of constructs taken through each step during the CDH protocol can be found in **Table 4.2**.

Table 4.1 Summary of the sequenced Claspin expression constructs.

This Figure is redacted for reasons of confidentiality.

Table 4.2 Claspin CDH construct summary table.
This Table is redacted for reasons of confidentiality.

The DNA sequences of the 42 constructs were position-mapped against the full-length Claspin sequence (**Figure 4.9A**). This alignment identified two sub-regions within the human Claspin gene, which separated the protein into two distinct halves, the N-terminus (residues 1-625) and the C-terminus (residues 611-1339) (**Figure 4.9B**). The central region is predicted to contain CC2 (aa 592-681) with a poly-Glu repeat. Interestingly none of the N-terminal constructs contained this poly-Glu repeat, in contrast, 11 C-terminal constructs did. However, only 5 of these constructs were found to be 'in-frame'. This strongly indicates that this poly-Glu repeat region may be difficult to express / fold in the heterologous *E. coli* host. It is also worth noting that there is a clustering of 4 constructs that start around the N-terminus of CC1 (aa 162) and whilst CC1 (aa 162-196) is found in 4 further constructs (plus one "out-of-frame, in-frame" construct), no identified constructs start within the CC1 region.

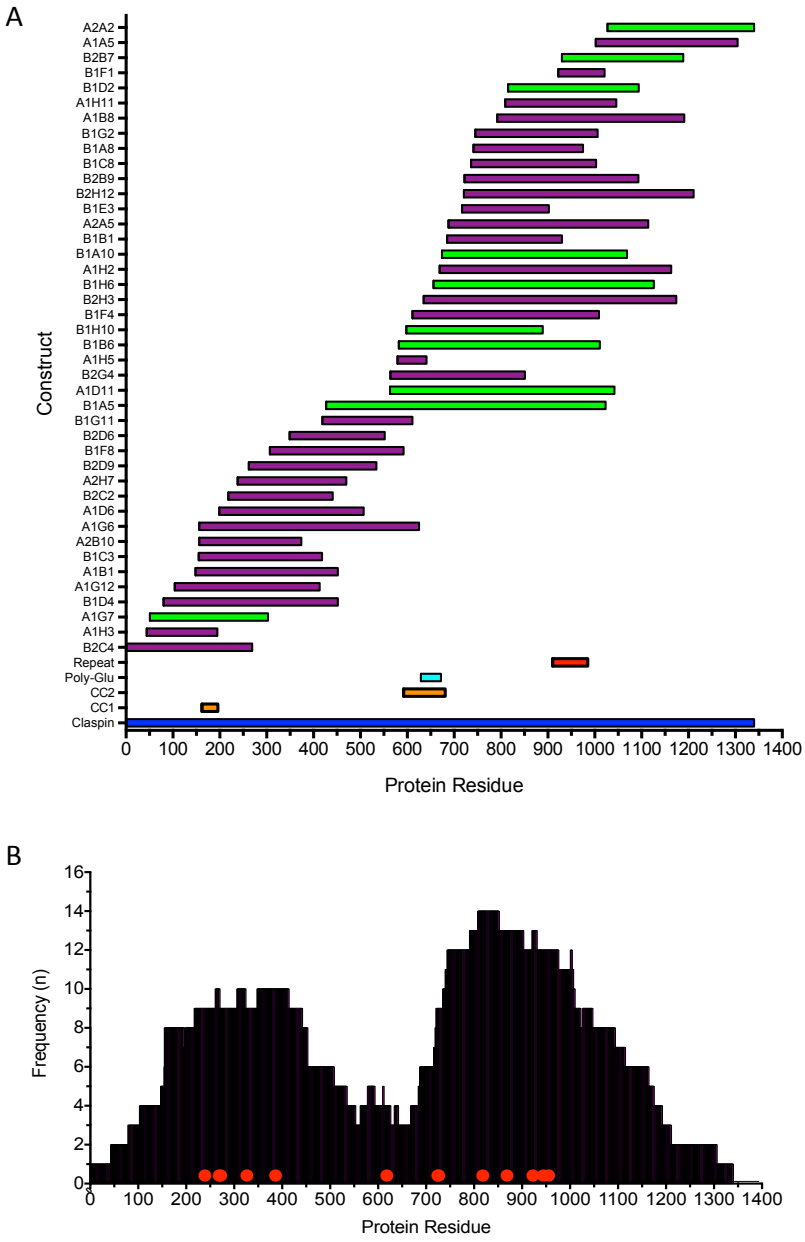


Figure 4.9 Sequence coverage and analysis of initial 'positive hits' from the CDH screen.

(**A**) Alignment of the fragment library, relative to the full-length synthetic Claspin gene (blue) as determined by sequencing of selected clones. The position of CC1 and CC2 (orange), the central poly-Glu repeat (cyan) and the CKBD C-terminal repeat motif (red) are highlighted. Constructs in the 'correct' orientation with respect to the start codon of the pDXV4-TOPO vectors are coloured purple, and those termed 'out-of-frame, in-frame' constructs are coloured green. (**B**) A frequency chart showing the occurrence of a particular amino acid residue in the soluble Claspin protein fragments. Observed mutations (amino acid substitution) are shown at their respective locations (red circles).

Many single nucleotide changes (mutations) were found the sequences of the Claspin fragment library. One third of the constructs sequenced contained a mutation resulting in a single amino acid substitution, whereas two constructs contained two amino acid substitutions (see **Figure 4.9B**). Taq DNA polymerase was used as it does not have any proofreading function, which is essential for the incorporation of dUTP, as other DNA polymerases would stall at uracil incorporation. The PFU Turbo Cx polymerase can also incorporate uracil without stalling, whilst retaining proofreading functionality. But single / multiple amino acid substitutions can in fact help to solubilise and/or stabilise a recombinant protein, so allowing these mutations to occur can be beneficial for the expression of recombinant proteins; some library techniques have been developed to exploit this, such as the technique of DNA shuffling (section **1.4.1**) (Stemmer, 1994a, Waldo et al., 1999). However, there is not a single region within the Claspin gene where mutations are more prevalent, and no one mutation is found in more than one construct, therefore these are almost undoubtedly as a result of using Taq DNA polymerase for the dUTP-PCR reaction.

Another observation, is that all of the soluble proteins migrate aberrantly on SDS-PAGE gels, running at a higher molecular mass than expected based on their amino acid sequence. This may be as a result of the overall acidic nature of the Claspin protein, leading to less SDS binding (via charge repulsion), and therefore retardation in the gel. This anomaly was previously reported in two published manuscripts; Kumagai and Dunphy, (2000) and Uno and Masai, (2011).

4.4.1 Small-scale expression and purification

Small-scale expression and purification tests were carried out with 24 constructs; nine taken from the N-terminal region and 15 from the C-terminal region of Claspin. Five of these contained part of the poly-Glu repeat and two C-terminal constructs were 'out-of-frame, in-frame' constructs (**Figure 4.10A**). The 24 constructs were selected by several criteria, including: over-expression level (SDS-PAGE), purity (SDS-PAGE), stability (SDS-PAGE and western blot, i.e. migrating predominantly as a single species), and their alignment position relative to the full-length Claspin gene. These were grown and expressed in small-culture volumes of

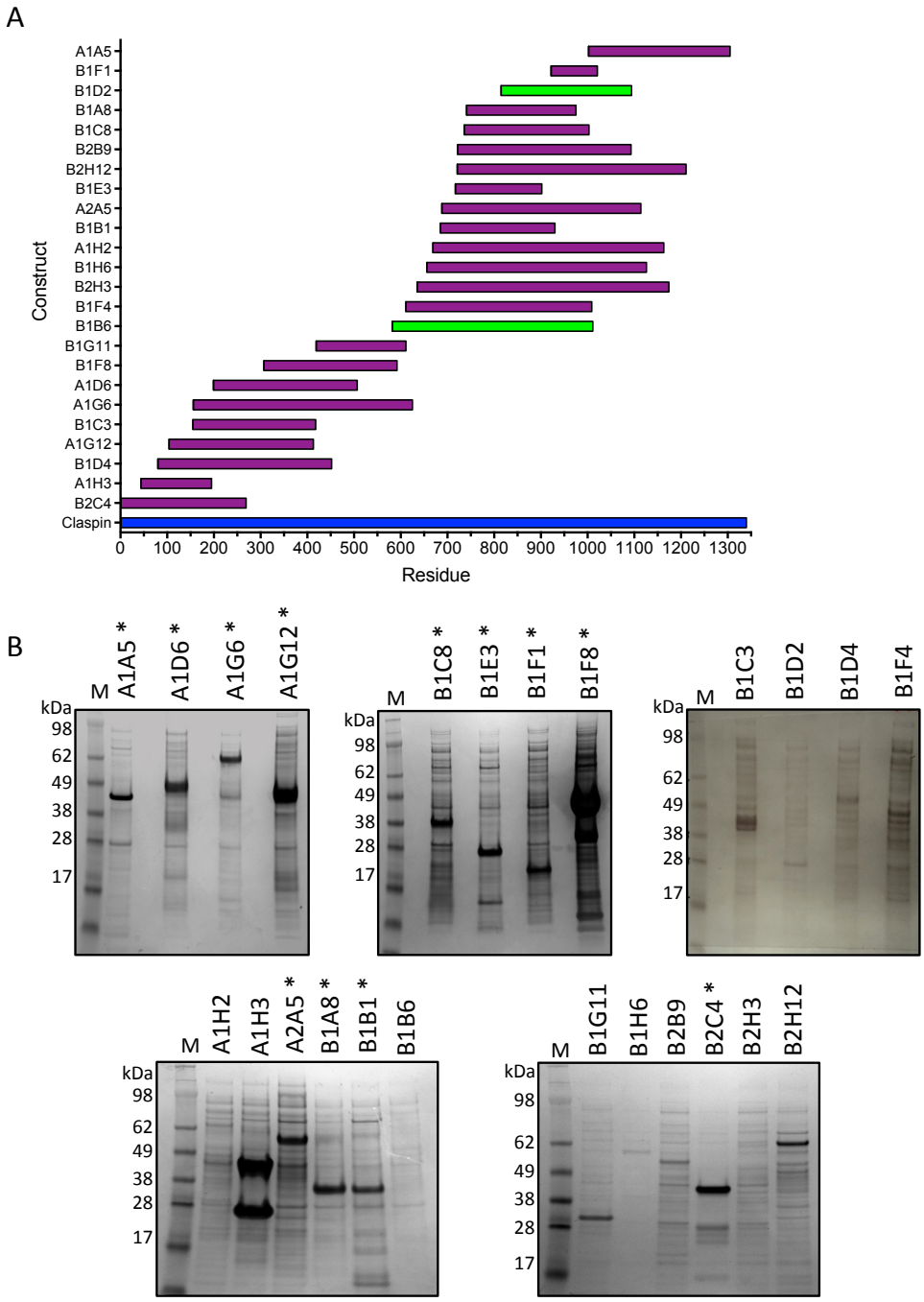


Figure 4.10 Small-scale expression screening for 24 Claspin expression constructs.

(**A**) Alignment of the 24 selected expression constructs relative to the full-length synthetic Claspin gene (coloured blue). Constructs in the 'correct' orientation with respect to the start codon of the pDXV4-TOPO vectors are coloured purple, and those termed 'out-of-frame, in-frame' constructs are coloured green. (**B**) Concentrated eluates from a one-step IMAC purification step for each of the 24 selected fragments, examined using Instant Blue-stained 4-12% Bis-Tris SDS-PAGE gels. Samples with high levels of recombinant protein expression are indicated with an asterisk. The unique expression construct IDs are indicated above each lane of the gel. M=molecular mass marker.

auto-induction medium, and the recombinant protein was then purified. Talon resin as opposed to Ni-NTA resin was preferentially used for the first IMAC purification step, due to its higher performance with concentrated cellular lysates (non-blocking) and reduced non-specific binding. The IMAC-bound proteins were eluted, concentrated, and resolved by SDS-PAGE. High-level expression of recombinant protein was seen for most of the constructs tested including 5 N-terminal and 7 C-terminal constructs (marked by asterisks in **Figure 4.10B**), but the constructs A1G12 and B1F8 both expressed significantly higher levels of recombinant protein; both of these constructs correspond to the N-terminal region of Claspin. The construct A1H3 when concentrated consistently resolved as two migrating species on SDS-PAGE gels. These could be made to form one species by increasing the concentration of β-ME in the SDS-PAGE sample loading buffer; thus indicating disulphide bond formation. Analysis of this sample by CD showed an anomalous CD spectra (strong negative ellipticity at 230 nm), and subsequently this construct was omitted from any further experiments. The five constructs that contained all or part of the described poly-Glu repeat had low to undetectable levels of recombinant protein expression, again indicating that the heterologous *E. coli* host had particular difficulties in making / folding this part of the human Claspin protein. Further, the two 'out-of-frame, in-frame' constructs showed no recombinant protein expression

4.5 Large scale expression and purification studies

One of the primary goals for this project, was to produce soluble protein, to enable biochemical and biophysical analysis for different parts / regions of human Claspin. For such analyses, milligram quantities of stable and homogenous recombinant protein are required. To determine if any of the Claspin expression constructs, produced from the CDH screen, had the required characteristics to warrant such analysis, 6 initial constructs were taken forward into large-scale expression / purification tests: C-terminal fragments: A1A5 (aa 1102-1305), A2A5 (aa 688-1114) and B1A8 (aa 2224-2926), and N-terminal fragments: A1D6 (aa 199-507), A1G12 (aa 104-413) and B1F8 (aa 307-592). Expression and purification of the C-terminal fragments is described below, whilst the N-terminal fragments purifications are expanded upon in **Chapter 5**.

4.5.1 Expression and purification of A1A5, A2A5 and B1A8

Large-scale expression studies were carried out for each of the 3 C-terminal constructs (section **2.15.2**) The soluble fraction resulting from 1 litre of *E. coli* cell culture, induced with IPTG, was purified using Talon-IMAC, Ion Exchange Chromatography and SEC; the eluted fractions were analysed by SDS-PAGE.

After Talon-IMAC, all three recombinant proteins were found to migrate at a greater than expected molecular mass, but consistent with those previously observed in the initial CDH screen and small-scale expression tests. The A2A5 protein seemed particularly prone to degradation and was therefore not taken any further (**Appendix 1.6**). The A1A5 and B1A8 proteins were both highly expressed, but several contaminants were visible by SDS-PAGE (**Appendices 1.7A** and **1.8A**). A second chromatography step was introduced, and the pooled elution fractions from Talon-IMAC were loaded onto a Q-sepharose ion exchange column and a linear salt gradient was then applied to the column. The A1A5 protein eluted from the column as a single peak (at ~500 mM NaCl, ~25% IEX B buffer), but there did not appear to be any significant improvement in the purity of the protein (**Appendix 1.7B**). For B1A8 protein, a proportion of the protein was retained and eluted from the column identified in the first peak on the elution trace ('bound' sample) (at ~300 mM NaCl, 15% IEX B buffer), but some protein had flowed straight through the column ('unbound' sample). Interestingly, the 'unbound' sample appeared to be much cleaner than that had been retained (**Appendix 1.8B**). As a final 'polishing step', each recombinant protein sample was purified using SEC. The A1A5 protein was still found to contain a large number of apparent contaminants, with little improvement in purity over the eluate from the initial IMAC step, therefore this construct was not taken any further (**Appendix 1.7C**). The B1A8 protein that was retained on the Q-sepharose column ('bound' sample) eluted as a complex mixture from SEC; some eluted at the void volume, as well as in 2 additional peaks. As this protein has a predicted molecular mass of 25.7 kDa, elution at the void volumes indicates a highly aggregated sample (**Appendix 1.8C**). The protein that was not retained on the Q-sepharose column ('unbound' sample) was also applied to the SEC column. This sample eluted within the resolving range of the column, but whilst the purity of the sample had been improved, the overall

concentration was significantly reduced compared to the 'bound' sample (**Appendix 1.8D**). Therefore the B1A8 construct was not taken any further. Each of these purifications strategies failed to produce recombinant protein at sufficient quantity or quality for downstream studies for these C-terminal Claspin constructs.

4.6 Summary

Parts of this section are redacted for reasons of confidentiality.

Domain-hunting technologies and bioinformatics-based predictions, which enable sub-construct production, have resulted in a major leap forward in the amount of structural information available for many proteins from many different species (Savitsky et al., 2010, Littler, 2010). However, these techniques have mostly been applied to situations where the target protein either contains a known domain/s or has a known function/s. Here the CDH methodology was applied to a largely uncharacterized human protein Claspin, which has little sequence homology to other known proteins and contains no predicted motifs or domains, as previously described by Kumagai and Dunphy, (2000). There is a predicted HTH motif (Zhao and Russell, 2004), in addition to four basic sequence patches (Lee et al., 2005). Independent bioinformatic analyses indicate that Claspin is comprised predominantly of α-helices, but with a high degree of predicted structural disorder, with two regions of sequence that predict as coiled-coils and a C-terminal region that is defined as a repeat motif (Consortium 2015). Two existing models based on low homology template sequences, were found in publically available repositories. Protein threading identified helical elements to the protein in addition to large regions of structural disorder. However, due to the low sequence identity / homology between target and templates, and the lack of agreement for regions of common amino acid sequence, these models are highly unlikely to be accurate in their prediction of the tertiary structure of human Claspin.

The CDH methodology was used to create a fragmentation library, which covered the entirety of the human Claspin gene. From a library of clones, 42 constructs were identified that expressed soluble recombinant protein to a high level. DNA

sequence analysis of these 42 clones indicated the presence of two distinct regions producing soluble protein, one at the N-terminus and the other at the C-terminus of Claspin. However this may simply be due to unintended bias, due to the presence of a central coiled-coil (with poly-Glu repeat), which seems to present the heterologous *E. coli* host with a particular expression or 'folding challenge'. Small-scale expression tests of 24 (selected) constructs indicated that approximately 50% of these were amenable to larger scale expression and purification. Six expression constructs were selected as being representative, three for the N-terminal region, and three for the C-terminal region of Claspin. However, all three C-terminal fragments proved refractory to purification: either remaining at insufficient purity, being prone to degradation, or to aggregation. As constructs expressing N-terminal fragments of Claspin proved to be far more tractable (see **Chapter 5**), a decision was made not to continue working with the C-terminal constructs.

Interestingly, all recombinant protein fragments (whether N- or C-terminal) consistently migrated at a higher than expected molecular mass when analysed by SDS-PAGE. This may simply be a reflection of the general acidic nature of the soluble recombinant proteins, and thus affecting the amount of SDS-bound, and thus the migration of a protein though the gel matrix as previously was described by Kumagai and Dunphy, (2000), and Uno and Masai, (2011). This observation is examined in more detail for the more tractable N-terminal fragments in **Chapter 5**.

One of the goals of the CDH screen is to only identify 'positive' expression constructs. However, 'false-positive' constructs are often identified during a screen as was found from this Claspin CDH. A number of factors could have caused these results including, but not limited to: some expression constructs are only produced when auto-induction medium is used rather than TB medium with IPTG induction; some expression constructs form soluble aggregates that are not removed by centrifugation in the solubility screen; there may have also been cross-contamination or human errors during the high-throughput protein purification step or the DNA sequencing step, resulting in the mixing of samples. Additionally the 'false-positive' initial hits seen on the SDS-PAGE gel may be have been a

commonly overexpressed *E. coli* protein as a number of these that are commonly captured on Ni-NTA (Bolanos-Garcia and Davies, 2006). The individual cause for each 'false-positive' construct was not determined.

To identify soluble and stable protein fragments from the C-terminal region of Claspin, a more restrictive CDH screen could be carried out focusing specifically on the C-terminus of the protein.

Chapter 5

Biochemical and biophysical analyses of the N-terminal region of human Claspin

5.1 Introduction

Claspin was originally identified through its interaction with Chk1, and has since been identified as an essential component of the intertwined processes of DNA replication and replication-coupled repair (Kumagai and Dunphy, 2000, Kumagai and Dunphy, 2003, Chini and Chen, 2003, Lin et al., 2004, Petermann et al., 2008). Claspin is described as a large, highly charged, protein 'adaptor', which mediates a number of protein-protein interactions that may be important for its cellular function (Kumagai and Dunphy, 2000, Lee et al., 2005, Kim et al., Gotter et al., 2007, Gold and Dunphy, 2010, Uno and Masai, 2011, Serçin and Kemp, 2011). Both full-length protein and fragments of Claspin have been purified from heterologous hosts, but to date little biochemical or biophysical analysis has been carried out (section **4.2.1**). Bioinformatic-based structural predictions for Claspin indicate that it has a high α-helical content, in addition to large regions of disorder (section **4.2.1**). In **Chapter 4,** CDH was used to identify several N-terminal fragments of Claspin suitable for biochemical analysis (those corresponding to the C-terminal region were unsuitable). This chapter therefore focuses on the expression and purification of N-terminal fragments of Claspin, and their subsequent characterisation by ASEC, CD, AUC, thermal denaturation, chemical crosslinking and finally structural characterisation by bioSAXS and NMR experiments.

5.2 N-terminal Claspin fragments

The N-terminal region of Claspin (amino acids 1-625) is located upstream of a poly-Glu repeat which starts at amino acid 629 (part of CC2; aa 592-681). There were 13 different expression constructs, identified by CDH, which together spanned the N-terminal region of Claspin. Four of these constructs, A1B1, B1C3, A2B10 and A1G6, had similar N-termini, which mapped between amino acids 148 and 156 and were N-terminal with respect to CC1 (aa 162-196) (**Figure 5.1A**). The high incidence of constructs staring at this position therefore indicates a preferred start site for soluble protein expression. Similarly, there was a clustering of C-termini, between amino acids 413 and 452, and therefore indicating a preferred site for the ends of soluble protein expression (**Figure 5.1B**).

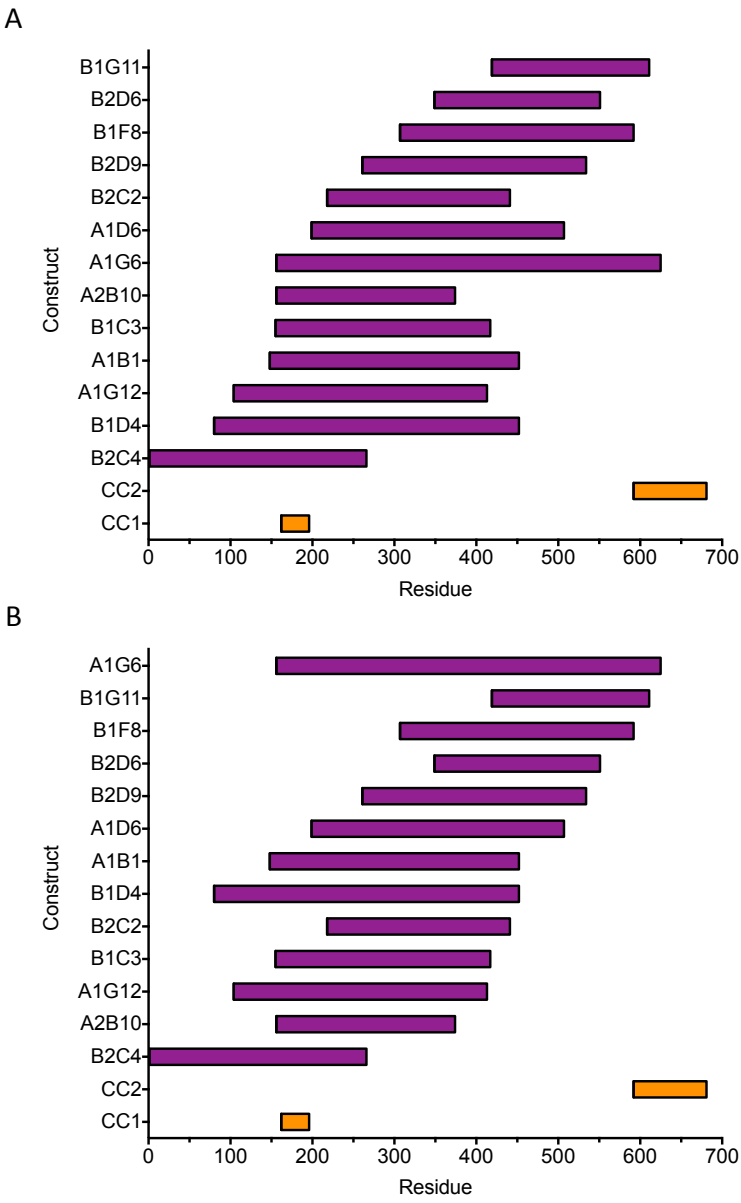


Figure 5.1 Sequence alignment for the N-terminal CDH fragment library. Alignment of the 13 N-terminal Claspin expression constructs (purple) relative to the predicted coiled-coil motifs (CC1 and CC2; orange). (**A**) Alignment with respect to the position of the N-terminus of the Claspin fragments (**B**) Alignment with respect to the position of the C-terminus of the Claspin fragments.

5.3 Purification of N-terminal Claspin fragments

Large-scale expression studies were carried out for each of the 13 N-terminal constructs (section **2.15.2**). Typically, recombinant protein in the soluble fraction, resulting from three litres of *E. coli* cell culture, was purified in a three-step protocol consisting of Talon-IMAC, Heparin Chromatography and SEC. Selected fractions from each of these purification steps were analysed by SDS-PAGE (see **Appendices 2.1** to **2.13**).

Seven of the 13 protein fragments: B2C4, A1G12, A1D6, B2C2, B2D9, B1F8 and B2D6, produced soluble recombinant protein at high levels, and these were deemed 'positive-hits'. The initial IMAC step achieved substantial purification for each of the fragments, but some contaminating protein species remained in all cases. A Heparin chromatography step enhanced the purity of each fragment. Protein fragments B2D9, A1G6D, B1F8 and B2D6 all required pre-dilution (in order to reduce the salt concentration) before application to the Heparin column, and were all found to elute in buffer containing low concentrations of sodium chloride. The remaining protein fragments eluted in relatively high concentrations of sodium chloride, indicating a strong and potentially specific interaction with the resin (discussed in section **6.7**). Finally, SEC recovered each of the recombinant proteins as a single symmetrical peak - albeit eluting at a lower retention volume than would be expected for their molecular mass and thereby indicating that a significantly greater hydrodynamic radius and non-spherical behaviour (i.e. non-globular). It is worth noting that similar characteristics had previously been observed for full-length Claspin (Uno and Masai, 2011). Furthermore, each of the protein fragments was found to migrate with a larger relative molecular mass than expected by SDS-PAGE, as previously described in section **4.6** and described in Kumagai and Dunphy, (2000), and Uno and Masai, (2011). Each fragment migrates approximately 10 kDa above its predicted mass. This is likely to be to compositional bias, and the acidic charge of these fragments; where pI is predicted to be between 5.1 and 6.7 in each case.

The remaining six protein fragments B1D4, A1B1, B1C3, A2B10, A1G6 and B1G11, were intractable, as they did not produce highly-expressed and/or proteolytically

stable recombinant protein at sufficient levels for study, and were therefore not taken any further. However, for three of these constructs, A1B1, A2B10 and A1G6, a stable degradation product was visible during purification. The exact amino acid termini of the degradation products were determined by western blot analysis and/or Edman degradation (N-terminal sequencing) (see **Appendices 2.3E**, **2.5E** and **2.6D**). Both A1B1 and A1G6 were found to have a common N-terminal sequence, 'YQSSHH', starting at amino acid 339. The DNA encoding the proteolytic product of A1G6 was cloned to create the expression construct A1G6D, which was then expressed and purified as before. This protein fragment was found to be both highly-expressed and proteolytically stable (**Appendix 2.14**).

Together, these eight constructs spanned the N-terminal region of human Claspin. An alignment of the constructs to functionally annotated regions of Claspin is shown in **Figure 5.2A**, and a summary figure showing the purified protein samples is shown in **Figure 5.2B**.

5.4 Bioinformatic analysis of the N-terminal region of Claspin

Secondary structure predictions using PsiPred and Phyre^2 were carried out for the full-length Claspin amino acid sequence (section **4.2.1**). Similar predictions were repeated for each of the individual protein fragments, to determine if predictions were different when only a fragment of the protein was used as an input. Secondary structure predictions for the eight fragments (using PsiPred) showed some differences in both α-helical and β-strand containing regions, and a global reduction in the amount of predicted disorder (**Appendix 2.15**). Phyre^2 predictions were consistent with this observation (data not shown). However, these analyses do not provide unequivocal evidence of folded regions or domains within each of the constructs. Consequently, homology modelling using SwissModel (Schwede, 2003) was used to investigate the potential tertiary structure in each of the fragments (section **4.2.1**). SwissModel successfully identified templates for a number of the fragments, which were all bromodomain-containing proteins. The models were built from part of the sequence containing the bromodomain fold; which is a highly structurally conserved four-helix bundle that interacts with specific acetylated lysine residues

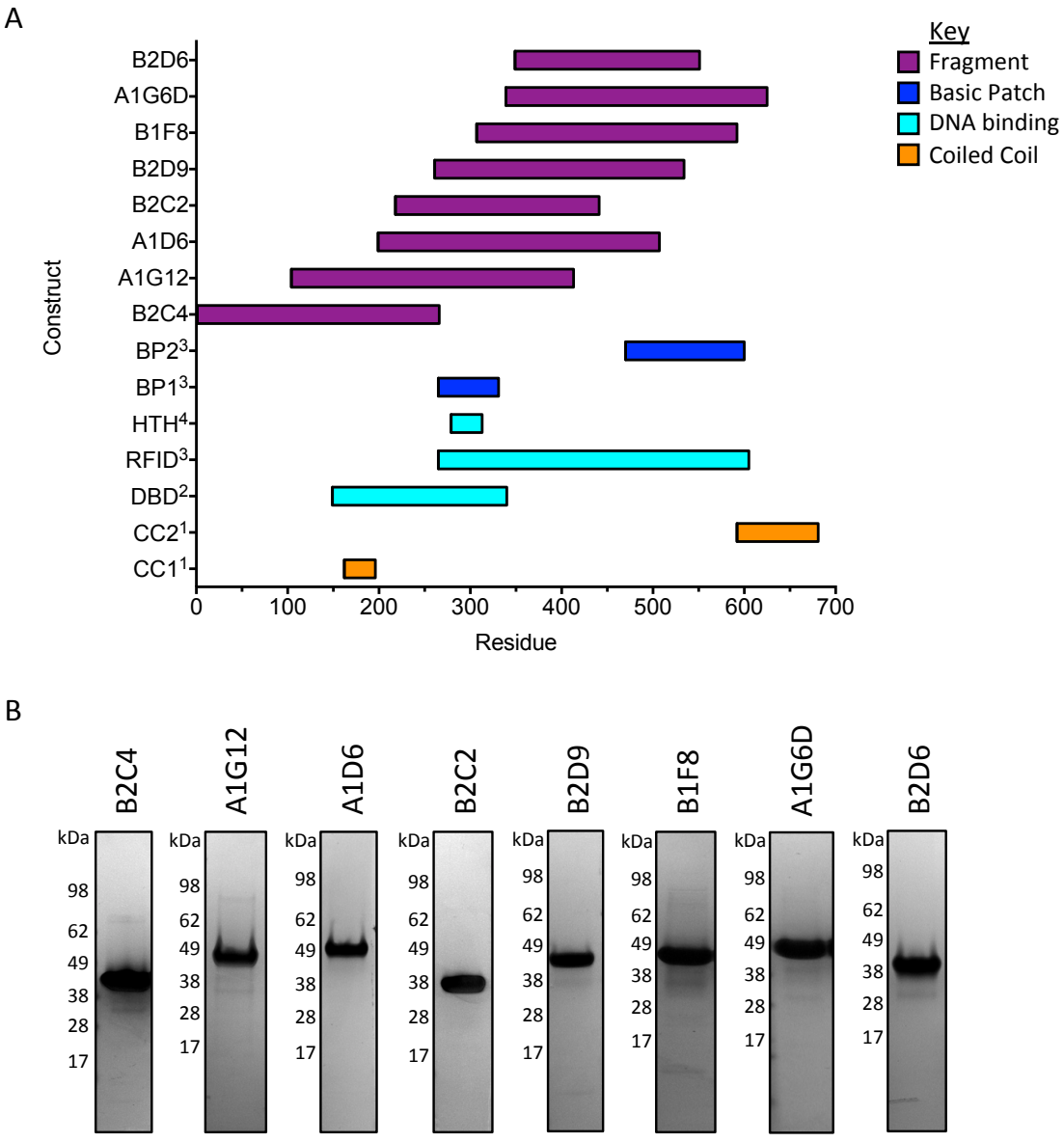


Figure 5.2 Successful N-terminal Claspin protein fragments.
(**A**) Alignment of the eight selected N-terminal Claspin fragments (purple), relative to functionally annotated regions of the Claspin N-terminus sequence. CC1 and CC2=Coiled Coil 1 and 2 (orange), RFID=Replication Fork Interaction Domain, DBD=DNA Binding Domain, HTH=Helix-Turn-Helix (cyan), BP1 and BP2=Basic Patch 1 and 2 (blue) ([1]Consortium, 2015, [2]Sar et al., 2004, [3]Lee et al., 2005, [4]Zhao and Russell, 2004). (**B**) SDS-PAGE analysis of concentrated purified Claspin protein fragments. Samples are as labelled above individual lanes and relative molecular mass markers sizes are on the left for each sample. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

in chromatin (Sanchez and Zhou, 2009). These analyses did indicate the presence of a HTH motif in four of the fragments, A1G12, A1D6, B2C2 and B2D9, which spanned amino acids 276-342 (**Figure 5.3**). However, the overall amino acid sequence homology was very low, being ~12-14% homology between the template and Claspin sequence. This region has been previously predicted as a HTH motif (aa 279-313) and is thought to be required for DNA binding (Zhao and Russell, 2004). Additionally, protein threading was performed for A1G12 using the Raptor X server (Kallberg et al., 2012), which based on the fold-template for RAP (receptor-associated protein; PDB: 2PO1), a Low Density Lipoprotein receptor molecular chaperone, predicted three HTH-like motifs in the sequence with large regions of disorder connecting these regions; with 36% α-helix, 0% β-strand and 63% disordered content (**Figure 5.4**).

5.5 Analytical Size Exclusion Chromatography

Here the, Claspin protein fragments were shown to have a lower retention volume by SEC and consistently a larger relative molecular mass by SDS-PAGE. ASEC, enables protein sizes to be more accurately examined with an improved estimation of molecular mass and sample dispersity than preparative SEC. Concentrated purified protein was diluted to 5 and/or 10 mg/ml and then applied to an ASEC column (section **2.18**). A relatively high protein concentration was necessary for ASEC, because each of the protein fragments tested had a relatively low molar extinction coefficient and therefore would not be readily detected at 280nm (at low concentrations). Two different protein concentrations were tested in order to examine any changes in the retention volume of the sample, which might be caused by concentration-induced protein association (dimerisation, tetramerisation, etc.) or by aggregation.

Each of the protein fragments tested, eluted as a single peak, within the resolving volume of the column, thereby indicating that the proteins are not forming mixed species in solution (**Figure 5.5**). Comparison of the 5 and 10 mg/ml chromatographs for each protein sample found these to be consistent with each other, with little change in the elution profile or retention volumes observed. Further analysis of the peak profile indicated that five of the fragments eluted with

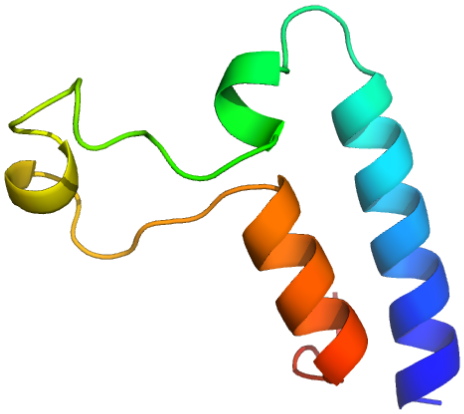


Figure 5.3 Homology-based model of human Claspin.
Structural prediction for Claspin amino acids 280-342, from homology-based modelling using the template TRIM24 (PDB: 2YYN), predicted using SwissModel (Schwede, 2003). Model is shown in molecular cartoon representations, coloured from blue to red, from the N- to C-terminus respectively (Jones' Rainbow).

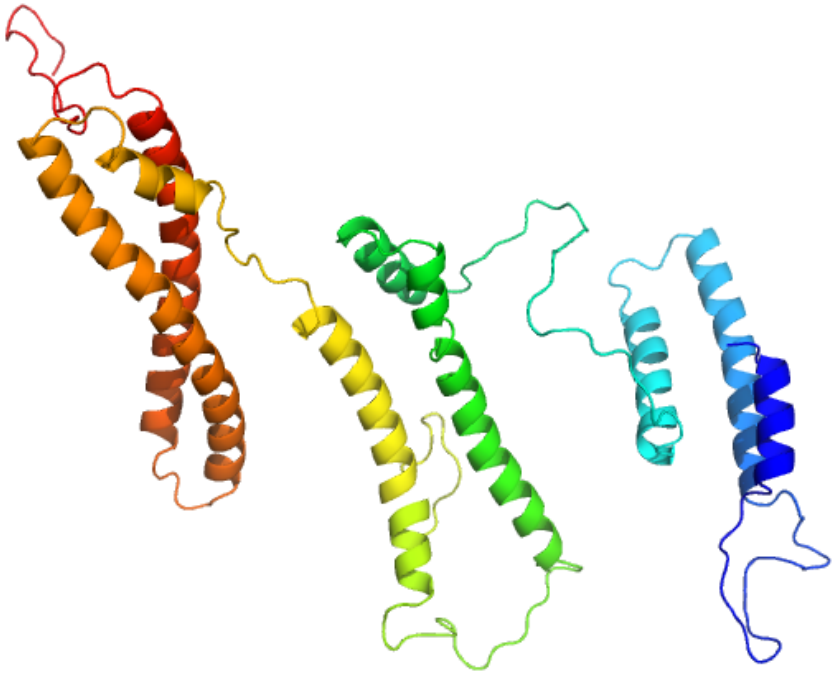


Figure 5.4 Protein threading model of human Claspin.
Structural prediction for Claspin amino acids 80-428, from protein threading using the template RAP (PDB: 2PO1), predicted using the Raptor X server (Kallberg et al, 2012). Model is shown in molecular cartoon representations, coloured from blue to red, from the N- to C-terminus respectively (Jones' Rainbow).

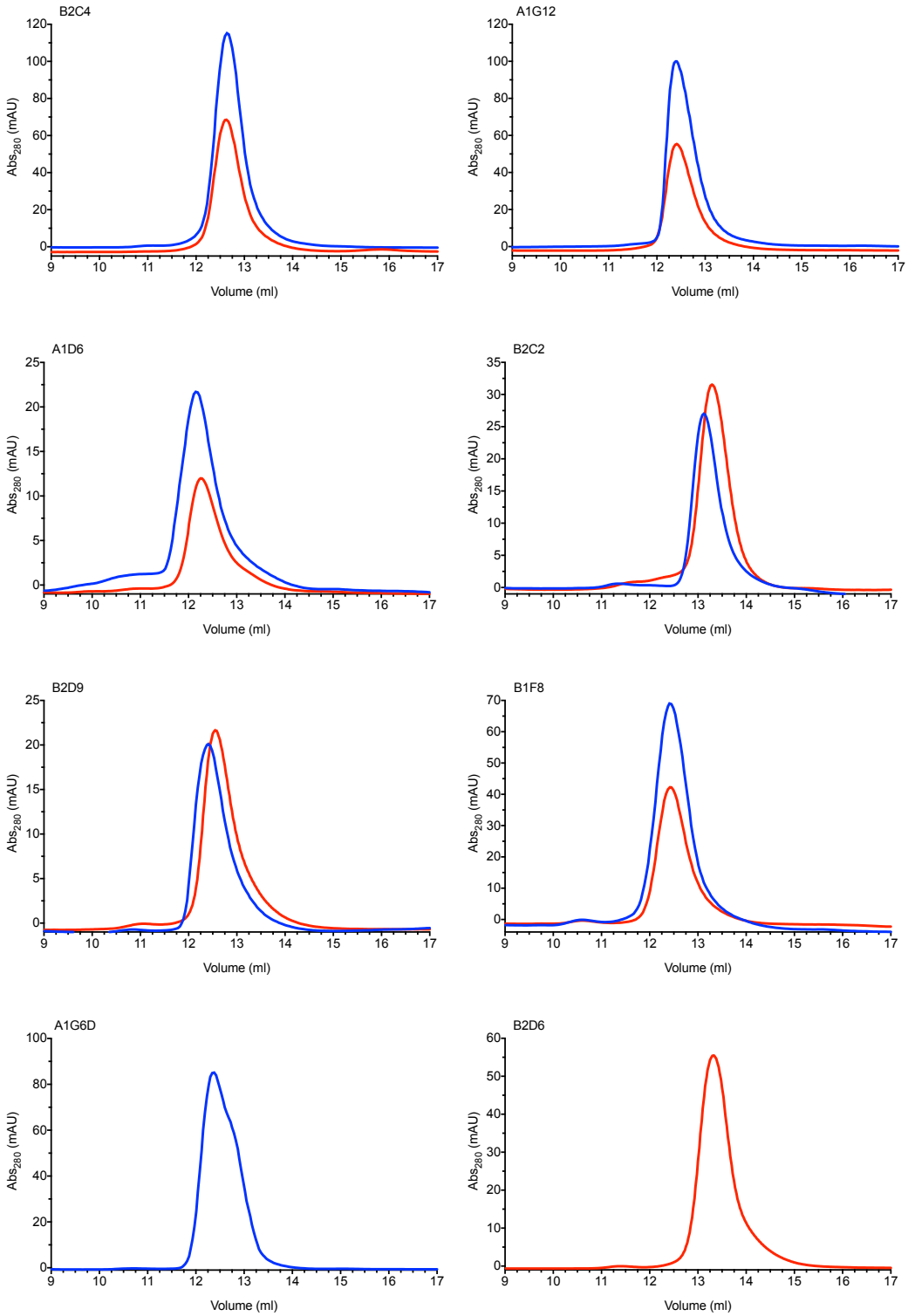


Figure 5.5 ASEC of the Claspin protein fragments.
Represenative ASEC chromatograms for purified Claspin protein fragments. Protein fragments at a final concentration of 5 mg/ml (red line) or 10 mg/ml (blue line) were applied directly to the column (10/300 SD200 Increase). The identity of each of the protein fragment is indicated in the top left hand corner of the trace.

a symmetrical peak profile, indicating monodispersity. For the other three samples: A1D6 and B2D9 showed that the trailing edge of the elution peak was slightly prolonged, and for A1G6D there was a distinct shoulder on the peak, suggesting that these samples were not monodisperse, or were making non-specific interactions with the resin of the column.

Using a calibration curve generated from molecular mass standards on the ASEC column (refer to **Figure 2.1**), the expected relative molecular mass for each of the fragments could be calculated. This indicated a four to five-fold increase between the determined molecular mass and the theoretical relative molecular mass (**Table 5.1**). Such a concomitant increase in the hydrodynamic radius of these apparently monodisperse protein fragments, indicates that they are likely to be either elongated in structure or contain unstructured regions. A second calibration curve was generated for the ASEC column, using the theoretical relative molecular masses and the elution volume of each fragment. This showed an almost linear relationship between the relative molecular mass of the fragments and the elution volume. A1G12 and B2C2 did not follow this relationship; eluting as a smaller mass compared to the remaining Claspin fragments; indicating these may contain increased structural order in comparison to the other fragments (**Figure 5.6**).

5.6 Circular Dichroism Spectroscopy

CD spectroscopy was used to investigate the secondary structure composition of the Claspin protein fragments. Measurement of CD spectra can reveal particular 'signatures' for different types of secondary structure element, such as α-helix or β-sheet, with the percentages of each element estimated through deconvolution-type analyses (Whitmore and Wallace 2004, Whitmore and Wallace, 2008, Kelly et al., 2005) (section **1.11.2**). To investigate if there were any concentration-dependent changes in protein structure, the CD spectra for each fragment was measured at concentrations of 30, 60, 90, 120 and 150 μM (section **2.19**).

The CD spectra units were converted to molar ellipticity to enable spectra comparison (**Figure 5.7A** and **Appendix 2.16A**). The CD spectra for each protein fragment, at low concentrations (30 and 60 μM) showed strong negative ellipticity

Construct	Predicted MW (kDa)	SDS-PAGE MW (kDa)	MW fold increase from SDS-PAGE	ASEC estimated MW (kDa)	MW fold increase from ASEC
B2C4	31.1	40.0	1.3	144.7	4.7
A1G12	36.5	47.0	1.3	162.1	4.4
A1D6	35.7	49.0	1.4	174.0	4.9
B2C2	26.4	38.0	1.4	106.9	4.0
B2D9	31.8	45.0	1.4	162.1	5.1
B1F8	32.8	45.0	1.4	160.6	4.9
A1G6D	33.0	45.0	1.4	166.0	5.1
B2D6	23.7	38.0	1.6	107.4	4.5

Table 5.1 Claspin protein fragment size comparison analysis.
Comparison of the Claspin protein fragments relative molecular mass as determined from bioinformatics predictions (ExPAXy ProtParam; Gastgeiger et al., 2005), SDS-PAGE analysis, and ASEC calibration curve extrapolation for the 5 mg/ml (or 10 mg/ml for A1G6D) elution volume from **Figure 5.5**. The fold-increase in the relative molecular mass with respect to the bioinformatics calculated molecular mass is shown.

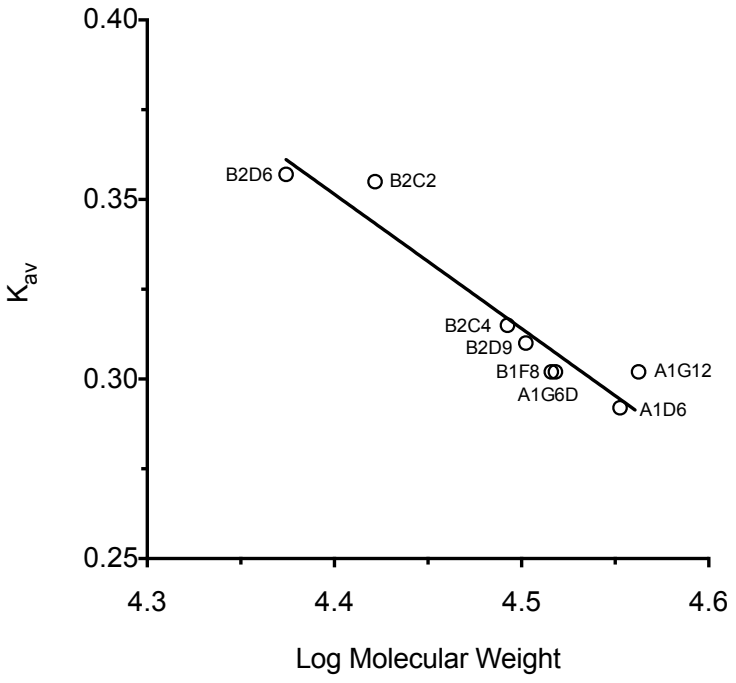


Figure 5.6 Calibration curve for the Claspin protein fragments.
Calibration curve for the 10/300 SD200 Increase column, using the relative molecular mass and elution volumes for the eight Claspin protein fragments. The curve was generated using calculated values for the log molecular mass and K_{av}, for the elution of the 5 mg/ml protein samples (or 10 mg/ml for A1G6D) from **Figure 5.5**.

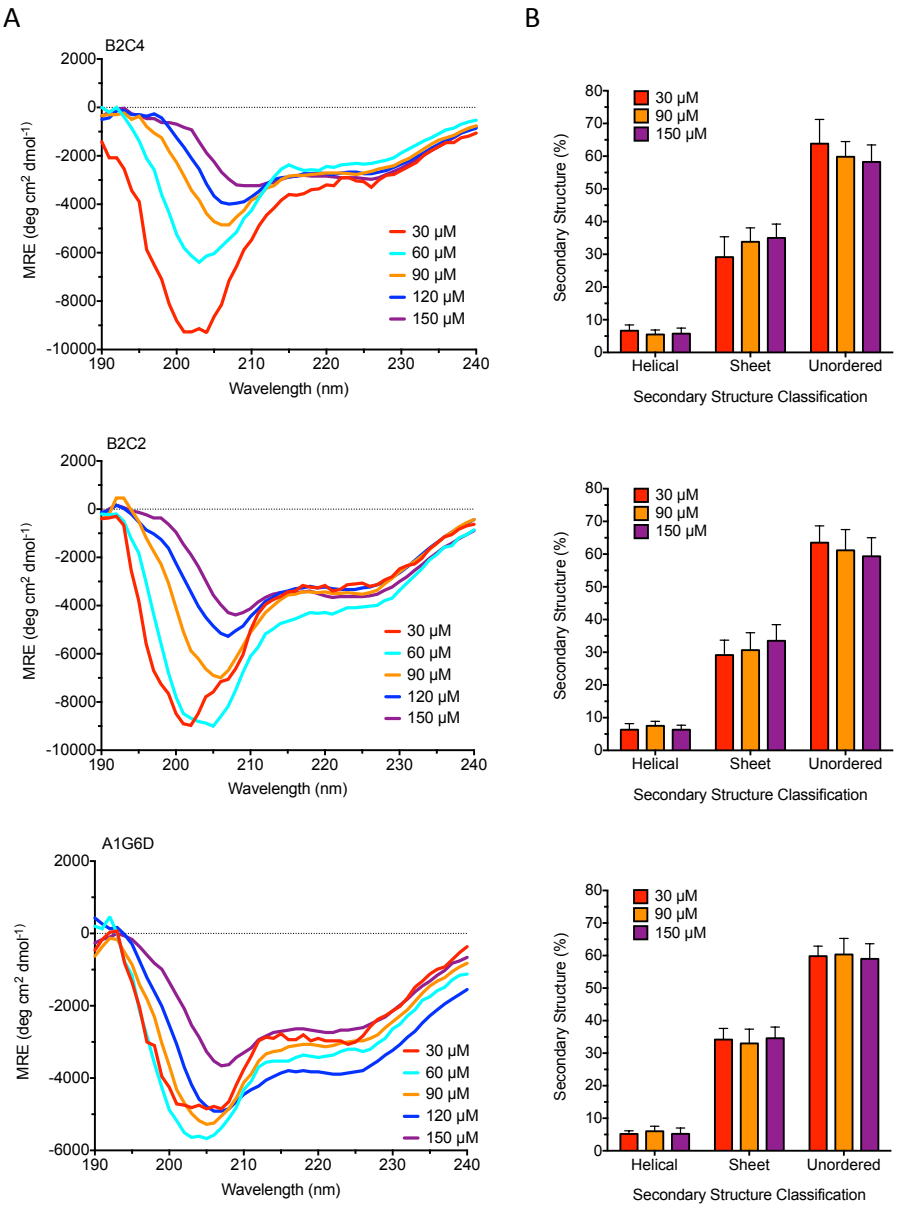


Figure 5.7 CD spectra and Dichroweb deconvolution for selected N-terminal Claspin protein fragments.

(**A**) CD spectra of recombinant Claspin protein fragments from top to bottom: B2C4, B2C2 and A1G6D (labelled top left). Protein samples were placed in a 0.1 mm quartz cuvette and the spectra measured between a wavelength of 190 and 240 nm using a JASCO J-715 spectropolarimeter, at 20 °C controlled by a JASCO PTC-384W peltier temperature control system. Data represents the average of three scans, from which the spectrum of buffer alone has been subtracted. The protein concentration is as indicated in the associated key. (**B**) Percentage of predicted secondary structure. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm, using six individual data sets, and then averaged. Error bars show one standard deviation. Protein concentration is indicated in the associated key (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).

at ~ 202 nm and slight negative ellipticity at 222 nm. Increasing the protein concentration resulted in a shift of the negative 202 nm peak to 208 nm. The peak at 222 nm, however, remained un-shifted. CD spectra for proteins with high α-helical content typically have ellipticity minima at 222 nm and 208 nm, and a maxima at 193 nm. A protein consisting of mostly random coil has a broad maxima peaking at 215 nm and a minima at 198 nm (see **Figure 1.10**). The negative ellipticity at 222 nm recorded for each of the protein fragments indicates some α-helical content, whilst the negative ellipticity at 202 nm is indicative of random coil (unfolded chain), therefore it is likely that the fragments contain a mixture of helical and random coil structures. The DichroWeb server, using CDSSTR program algorithm, was used to deconvolute the CD spectra for each protein fragment (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008, Sreerama et al., 2000a, Sreerama et al., 2000b) (section **2.19.1**). In each case, the calculated percentages of secondary structure elements were roughly 6% α-helix, 32% β-sheet and 62% unstructured (random coil), with no apparent trends / changes due to protein concentration (**Figure 5.7B** and **Appendix 2.16B**).

5.7 Thermal denaturation

Experimental data from preparative SEC, ASEC and CD spectroscopy consistently indicated there was a significant degree of disorder within each of the eight Claspin protein fragments tested. In an 'inverse' experiment, in order to determine if there was a folded region within any of the fragments, thermal denaturation experiments were carried out. This type of analysis can be used to investigate the stability of a protein in a number of different experimental conditions (or additives), each of which aim to increase the observed midpoint of transition from a folded to unfolded state with increasing temperature (*T*m). It can also be used to probe for the presence of a co-operatively unfolding (sub-)domain within a given protein.

5.7.1 ThermoFluor

ThermoFluor (or Differential Scanning Fluorimetry) makes use of the proprietary fluorescent dye SYPRO orange, which interacts with an exposed hydrophobic core of a protein during thermal denaturation, resulting in a measurable increase in its

fluorescent signal. To measure the thermal stability of the eight Claspin fragments, a series of protein concentrations were tested; 0.75, 1.9 and 3.75 μM, each of which were incubated with 5x SYPRO orange. The assay indicated a single, cooperative transition in all of the fragments, which occurred between temperature of 52 and 55 °C and corresponded to a single unfolding event. However, the change in fluorescent signal was very slight (**Figure 5.8**). The melting of a globular protein with a large hydrophobic core would typically have shown a greater change in fluorescent signal (as shown for Chk1-KD$^{1-270}$-His in **Figure 7.8A**). It is therefore likely, and consistent, that the observed change in fluorescence for each Claspin fragment is due to melting of a small region of secondary structure, rather than a large hydrophobic core.

5.7.2 Protein unfolding monitored by CD

CD melt spectra were collected in comparison to the native CD spectra, as described in section **5.6**. The CD spectra for each protein fragment was collected using 90 μM protein, measured at 20 °C, and then the temperature was raised to 80 °C (using a water-bath and peltier pump attached to the spectropolarimeter) before measuring a second (unfolded) CD spectrum.

Here, there were visible differences to the CD spectra as a result of denaturation (**Figure 5.9A** and **Appendix 2.17A**). For B2C4 and A1G12 there were differences to the ellipticity minimum at 222 nm, whereas A1D6 and B1F8 showed changes at wavelengths of 202 and 208 nm. Interesting, for A1G6D, B2C2, B2D9 and B2D6 the denatured CD spectra were entirely different, indicating that there was a change in secondary structure for these fragments as a result of denaturation. Each CD spectra was also submitted to the DichroWeb server as described in section **5.6** (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008, Sreerama et al., 2000a, Sreerama et al., 2000b). Whilst there were some notable changes to each spectrum upon heating to 80 °C, the deconvolution analysis indicated little change in the predicted percentage of disordered regions for each of the eight protein fragments (**Figure 5.9B** and **Appendix 2.17B**). This gives support to the notion that the Claspin protein has a high degree of intrinsic disorder.

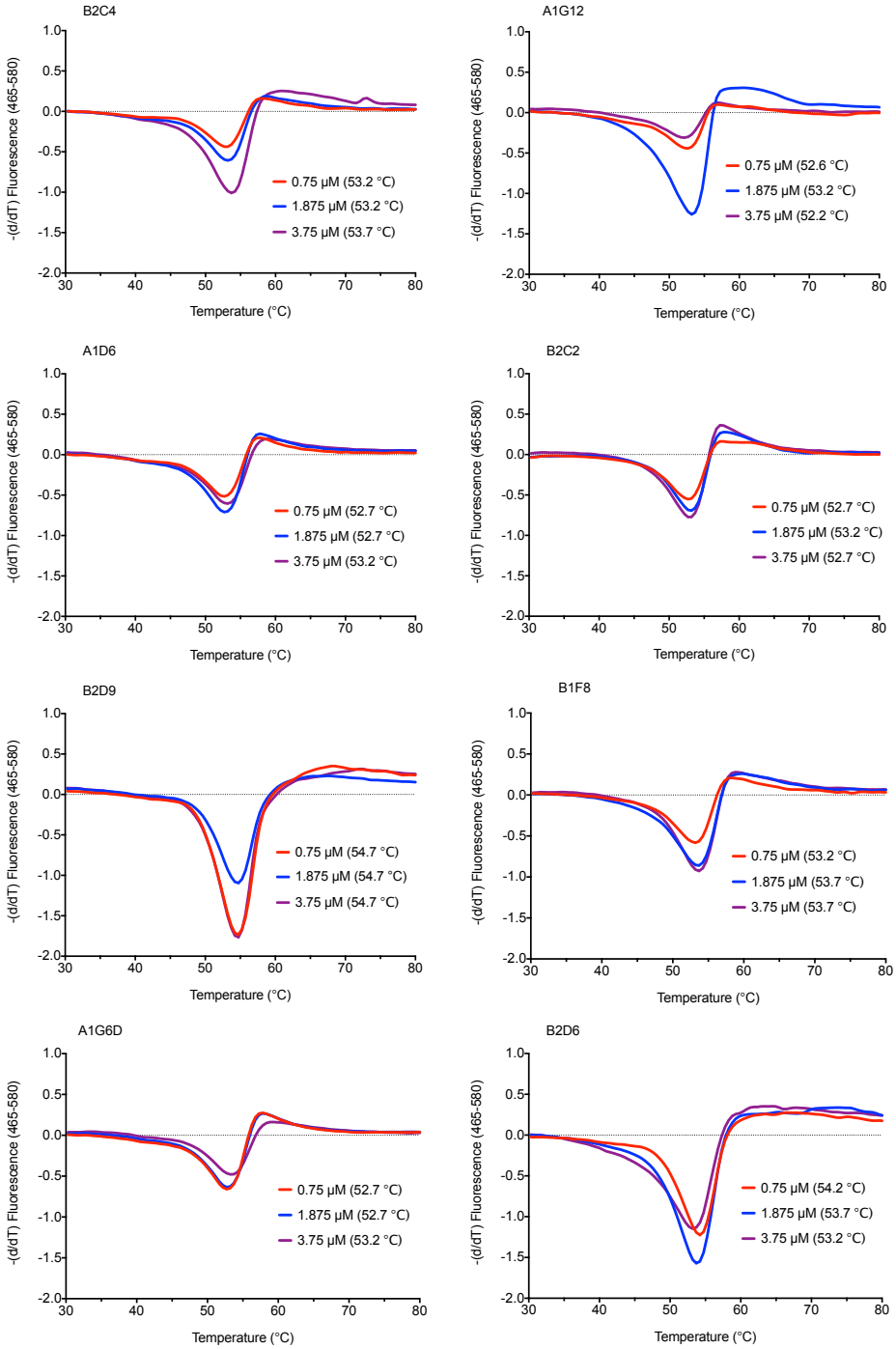


Figure 5.8 Thermal denaturation of the Claspin protein fragments.
For thermal denaturation, protein samples were mixed with 5x SYPRO orange, and melting profiles were obtained from 20 to 80 °C, with a temperature ramp of 0.02°C s^{-1} using the Roche LightCycler 480. The protein sample is shown top left on each graph. The protein concentration as well as the melting temperature (*T*m), determined by the point of the minima of the negative peak of the denaturation curve, is shown in the associated key. Each plot is the mean of three individual experiments.

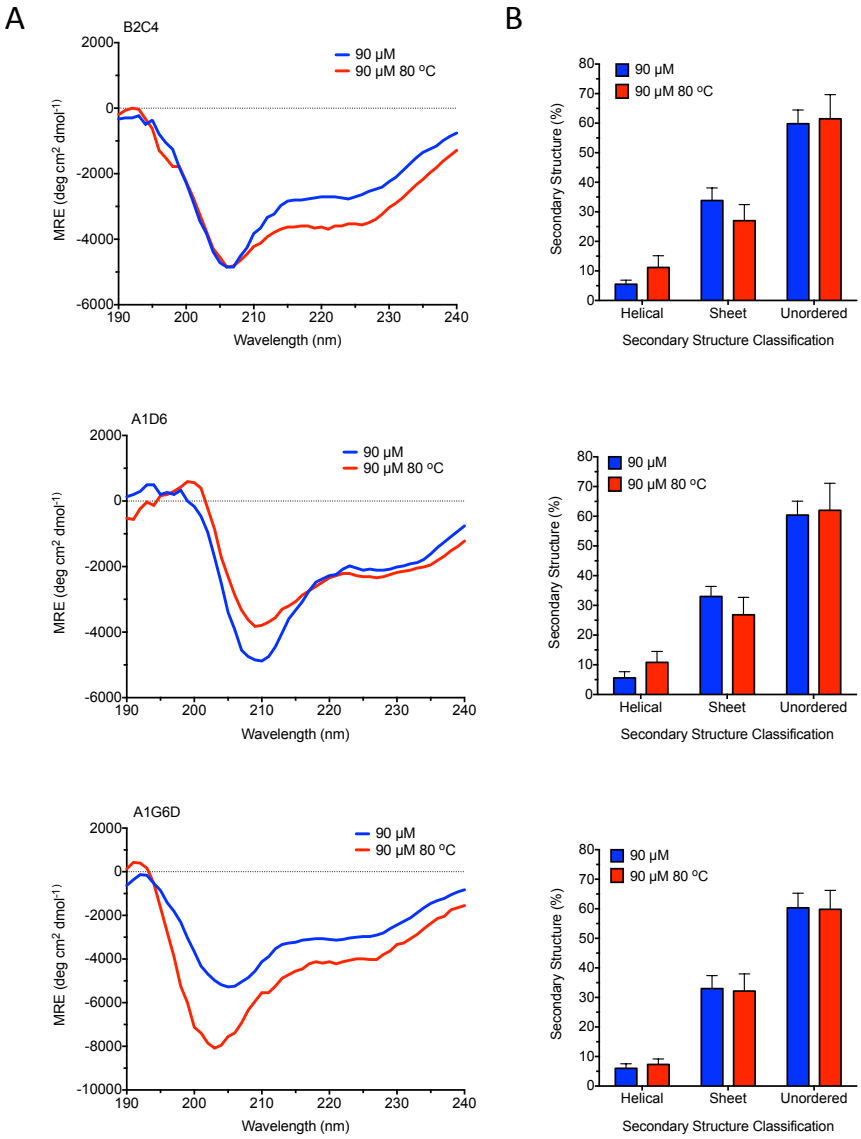


Figure 5.9 Two temperature point CD spectra and Dichroweb deconvolution for selected N-terminal Claspin protein fragments.

(**A**) CD spectra, measured at two temperature points, for selected recombinant Claspin protein fragments. From top to bottom: B2C4, A1D6 and A1G6D. The protein concentration as well as the temperature (80 °C) is shown in the associated key. Samples were placed in a 0.1 mm quartz cuvette and CD spectra were measured at wavelengths between 190 and 240 using a JASCO J-715 spectropolarimeter. Samples in cuvettes were heated at 80 °C for 3 minutes controlled by a JASCO PTC-384W peltier temperature control system and the CD spectra were re-measured. Data represents an average of three scans, from which the spectrum of the buffer alone has been subtracted. (**B**) Percentage of predicted secondary structure for recombinant Claspin protein fragments. The protein concentration as well as the temperature (80 °C) is shown in the associated key. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm using six individual data sets, and then averaged. Error bars show one standard deviation (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).

5.8 Chemical crosslinking

The data presented thus far indicate that the Claspin protein fragments are elongated molecules with little globular structure propensity. However, this does not mean that they are unable to form oligomeric structures. To determine if the protein fragments were multimeric (or otherwise), chemical crosslinking experiments with BS3, which irreversibly crosslinks primary amines that are in close proximity, were undertaken. Crosslinking was performed at two protein concentrations, 30 and 60 μM, and incubated with a dilution series of BS3 (50 mM to 0.5 μM) before being analysed by SDS-PAGE (**Figure 5.10**). At high concentrations of BS3, it was possible to visualise higher order oligomers of each protein fragment. However, at lower concentrations of BS3, only monomeric and dimeric species were observed. It was also noted that at high BS3 concentrations, the monomeric species of the protein fragment migrated somewhat faster on SDS-PAGE, which is due to intramolecular protein-crosslinks resulting in compaction of the protein. The results from the two different protein concentrations were very similar. The crosslinking studies therefore indicate that the protein fragments are most likely monomeric, but with some potential for dimerisation.

5.9 Analytical Ultra-Centrifugation

In order to better understand any potential monomeric to dimer transition, the protein fragments were also analysed by sedimentation velocity experiments in an analytical ultra-centrifuge (SV-AUC; Biophysics Laboratories, University of Portsmouth). For the protein fragments B2D9, A1G12 and B2C2, sedimentation was detected by absorbance optics — these can be set to measure protein concentration at different wavelengths, as long as the total absorbance of the sample at that particular wavelength is below 1.5. Data for fragment B2D9 (0.6 mg/ml) was collected at a wavelength of 280 nm (aromatic residues), whilst data for fragments A1G12 (0.3 mg/ml) and B2C2 (0.3 mg/ml) were collected at 230 nm (peptide backbone).

The raw absorbance sedimentation profile for the proteins are shown in **Figure 5.11A**. AUC data were collected and then fitted to a continuous size distribution

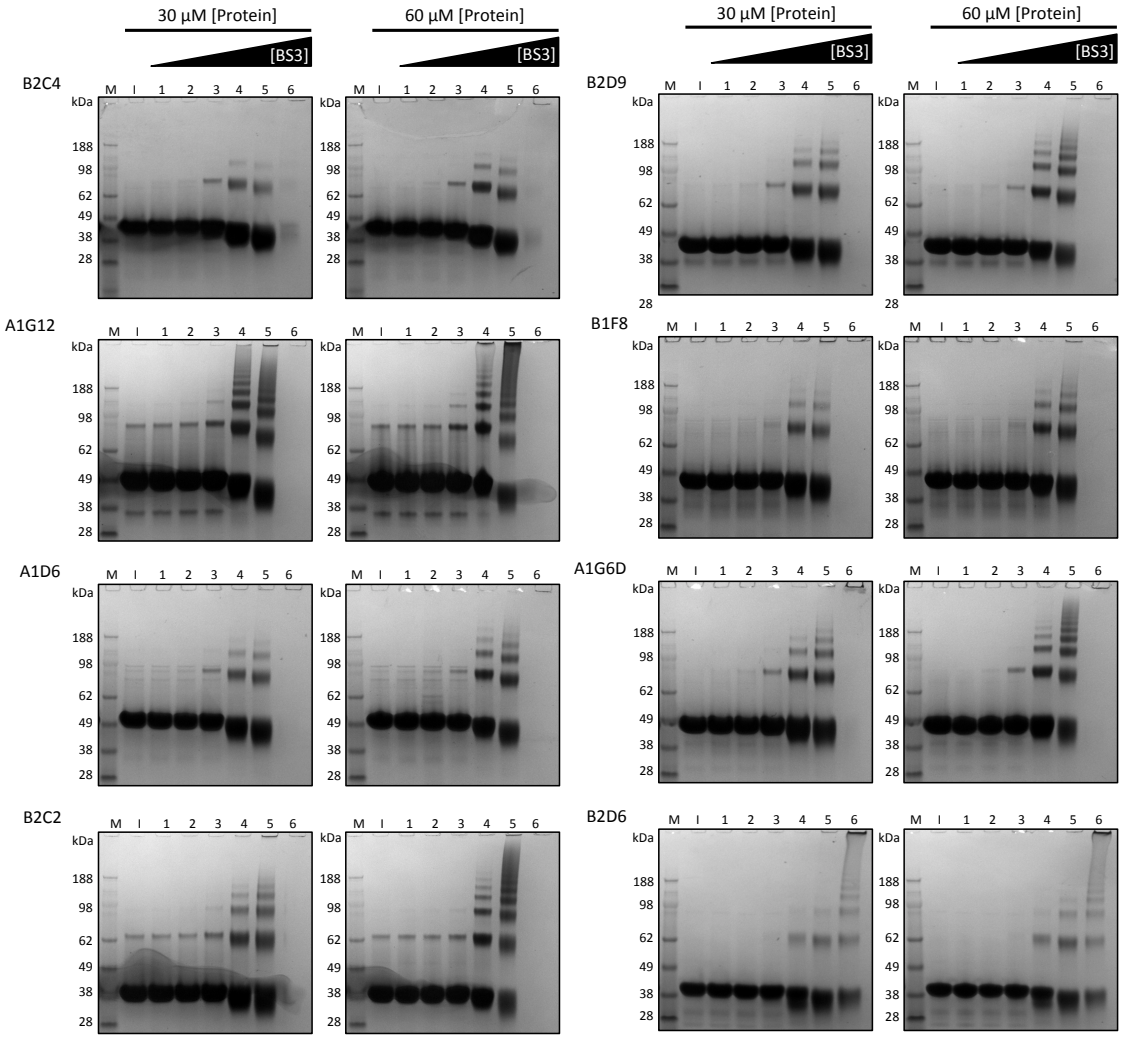


Figure 5.10 Chemical crosslinking of the Claspin protein fragments.
SDS-PAGE analysis of BS3 chemical crosslinking of purified Claspin protein fragments at a final concentration of either 30 or 60 µM. Proteins were incubated with 10-fold serial dilutions of BS3 crosslinker for one hour before the reaction was quenched with Tris pH 8.0. The protein fragments are labelled on the left of the SDS-PAGE gels. The protein concentration is shown above each column, as is the increase in BS3 concentration. M=molecular mass marker, I=input sample, 1 to 6=BS3 concentration where: 1=0.5 µM, 2=5 µM, 3=50 µM, 4=0.5 mM, 5=5 mM, 6=50 mM. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

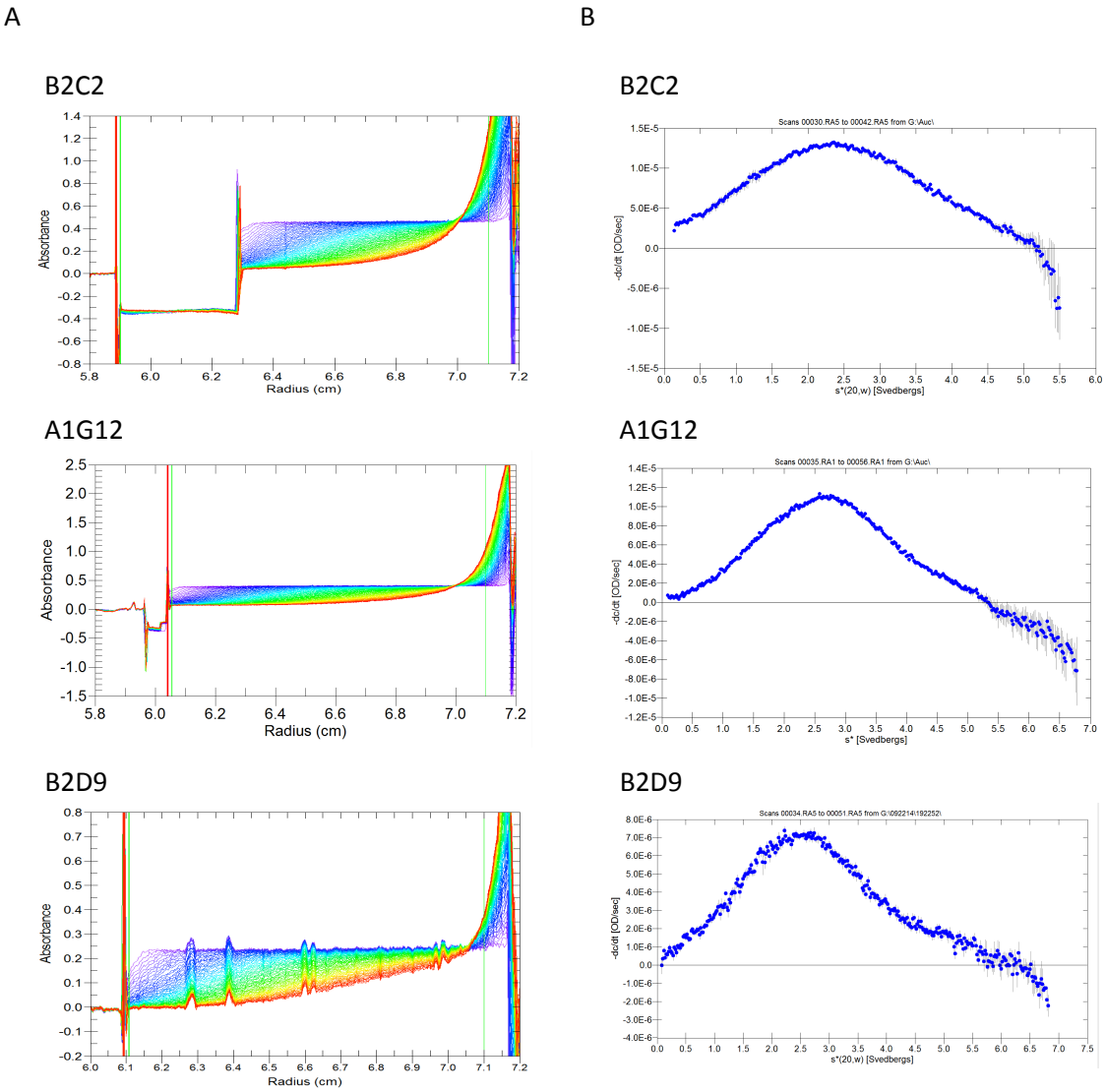


Figure 5.11 Analytical ultracentrifugation for the Claspin fragments.
(**A**) The raw absorbance sedimentation profile scan acquired every 10 minutes, recorded at 230 nm at 23 °C at 35,000 rpm for B2C2 and A1G12 and 15,000 rpm for B2D9. (**B**) The sedimentation profile for each protein sample was converted to a dc/dt plot using DCDT^{+2} (Philo, 2006). Plots are labelled in the top left corner; where from top to bottom these are B2C2, A1G12 and B2D9.

model using DCDT+ software (Philo, 2006) (**Figure 5.11B**). B2C2 and A1G12 both sedimented as a monomeric and monodisperse proteins; B2C2 (molecular mass = 26.51 kDa) had a sedimentation coefficient of S(20,w) 2.001 ± 0.005 and a molecular mass of 27.57 ± 0.30 kDa, whilst A1G12 (calculated molecular mass = 36.52 kDa) had a sedimentation coefficient of S(20,w) 2.361 ± 0.004 and a molecular mass of 36.69 ± 0.40 kDa; both of these molecular masses are just slightly larger than the calculated molecular mass of the proteins. For B2D9 (calculated molecular mass = 31.78 kDa), dc/dt analysis indicated a protein monomeric in molecular mass with a sedimentation coefficient of S(20,w) 2.226 ± 0.007 and a molecular mass of 31.86 ± 0.47 kDa, once again just slight larger than the calculated molecular mass of the protein. However, the dc/dt analysis indicated a potentially second large species to be sedimenting in the sample, but the reduced data quality in this region prevented accurate study of this region (large grey error bars in **Figure 5.11B** bottom). This data showed, that at the protein concentrations tested the Claspin protein fragments existed as monomeric proteins in solution, and that for B2C2 and A1G12 these were monodisperse samples.

5.10 Circular Dichroism with 2,2,2-trifluoroethanol

2,2,2-trifluoroethanol (TFE) is a solvent, which can be used to stabilise secondary structures in solution, in particular for α-helices and unfolded sequences with some intrinsic helical propensity (Sonnichsen et al., 1992, Reiersen and Rees, 2000). TFE is thought to act by the dehydration of specific residues within a random coil conformation or α-helix, resulting in an increase in stability of the α-helix (Starzyk et al., 2005). TFE has been used to study peptides and proteins with intrinsic disorder by CD and NMR by inducing a native-like structure to the chain, enabling characterisation of the protein (Buck, 1998, Kaczka et al., 2014, Maestro et al., 2013). However, the addition of TFE to a protein can also result in a conformational change divergent from the correct native tertiary structure (Gast et al., 1999). The effect of TFE on the Claspin protein fragments, was measured by incubating each of the eight protein fragments (at a concentration of 30 μM) with TFE. The addition of 10% (v/v) TFE caused some protein samples to precipitate, whilst the addition of 20 or 30% (v/v) TFE caused all of the protein samples to

precipitate. The precipitation of the protein fragments with lower concentrations of TFE could be due to partial structural changes, which either destabilized or cause non-specific associations of the protein molecules. However, all of the protein samples were stable in 40% (v/v) TFE when rapidly mixed, and the CD spectra were measured as described in section **5.6**.

Each representative CD spectra for the protein fragments in 40% (v/v) TFE shows a significant change when compared to the previously measured spectrum measured in the absence of any TFE - with a large positive maximum at 193 nm and large negative minima at both 208 nm and 222 nm - a characteristic signature for an α-helix containing protein (**Figure 5.12A** and **Appendix 2.18A**). Again, spectra were submitted to the DichroWeb server for spectral deconvolution (**Figure 5.12B** and **Appendix 2.18B**) (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008, Sreerama et al., 2000a, Sreerama et al., 2000b). A large increase in the predicted percentage of α-helical structure from ~6% to 20-45%, and decrease in the predicted percentage of β-sheet structure by ~20% was observed. In addition, there was an overall reduction in the predicted unstructured (random coil) content for each protein fragment, but this still represented a significant proportion (~50%) of the predicted structure for each protein fragment. These analyses showed that secondary structure (α-helix) was readily induced by the addition of TFE, but these still contained a significant amount of disorder.

In the light of these set of experiments, it is apparent that Claspin is a largely intrinsically disordered protein and is therefore unlikely to be amenable to crystallisation. Consequently, in order to structurally characterise these protein fragments, we adopted a three-method approach – involving SAXS, NMR and protein crystallography (Sibille and Bernado, 2012) although the latter was unlikely to yield any structure.

5.11 Crystallographic trials

Protein fragments B2C4, A1G12, B2C2 and B1F8 at concentrations between 5 and 60 mg/ml were set up in sitting drop format in plates as a 0.2 µl protein with 0.2 µl mother liquor drops, against a well volume of 40 µl using an ARI Crystal Phoenix

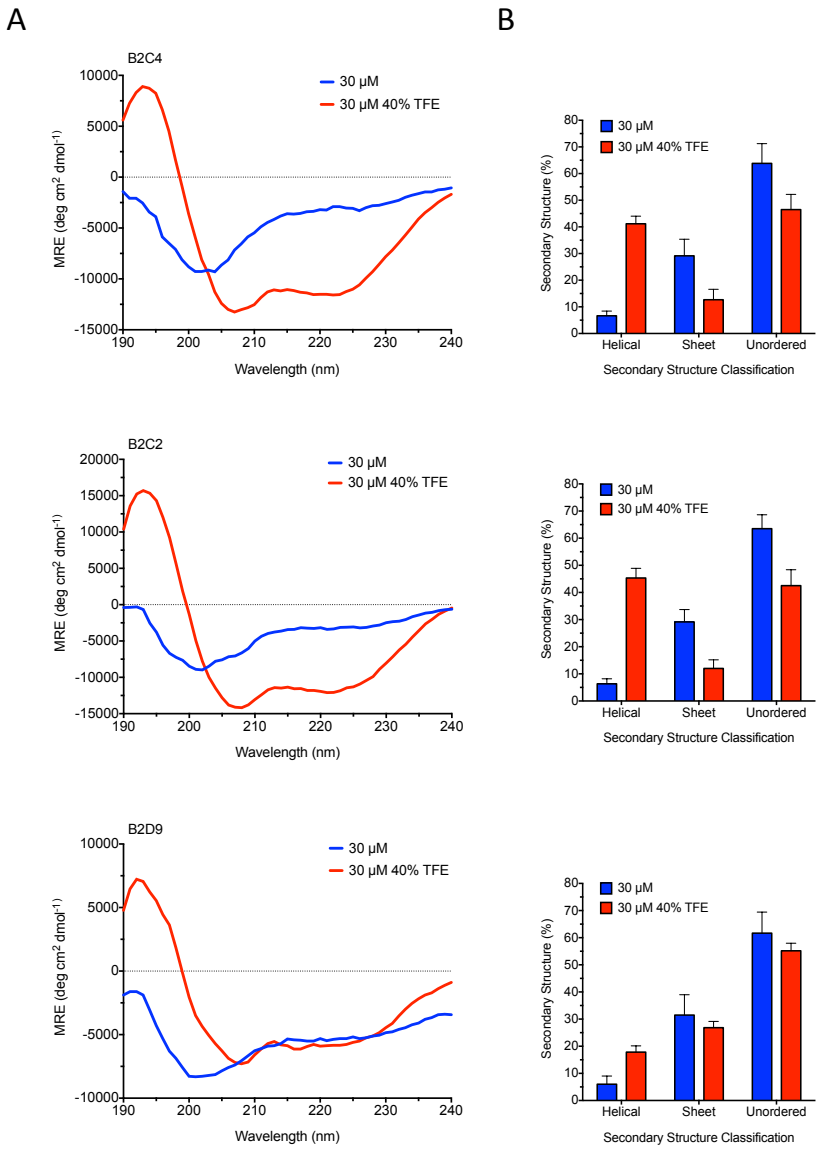


Figure 5.12 CD spectra and DichroWeb deconvolution for selected N-terminal Claspin fragments in the presence of TFE.

(**A**) CD spectra of recombinant Claspin protein fragments; from top to bottom: B2C4, B2C2 and B2D9 measured with and without 40% (v/v) TFE. The protein concentration and the TFE concentration are shown in the associated key. Samples were placed in a 0.1 mm quartz cuvette, and spectra were measured at wavelengths between 190 and 240 using a JASCO J-715 spectropolarimeter set at 20 °C controlled by a JASCO PTC-384W peltier temperature control system. Data represents the average of three scans, from which the spectra of the buffer alone has been subtracted. (**B**) Percentage of predicted secondary structure. The protein concentration as well as the TFE concentration is shown in the associated key. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm using six individual data sets, and then averaged. Error bars show one standard deviation (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).

Liquid Handling System. A wide range of different commercial crystallisation screens were used (PEG/Ion, SaltRx, Index, Structure screen I + II, PACT premier, JCSG-plus, ProPlex, Morpheus, and MIDAS), and duplicate screens were also incubated at different temperatures 4, 14 and 20 °C. As predicted, none of these sparse matrix crystallisation-screens yielded crystals. In summary, approximately >50% of the conditions tested resulted in some form of phase separation, which constitutes a separation of the protein from the mother liquor. Often, these conditions can be optimised using a screen around the original condition, which may result in crystallisation. However, optimisation of a variety of different phase separation-generating conditions failed to yield any crystals.

5.12 Small Angle X-ray Scattering

SAXS can determine the shape/volume and size of macromolecules in solution, and can be used to generate a low-resolution 3-dimensional 'bead model' which best represents the 'shape envelope' of a protein in solution (section **1.11.4**). SAXS can also be of great use when studying proteins with intrinsically disordered regions (Receveur-Brechot and Durand, 2012). Here, SAXS was used to determine the hydrodynamic volume of the Claspin protein fragments, and to determine if they possessed any level of 'foldedness'.

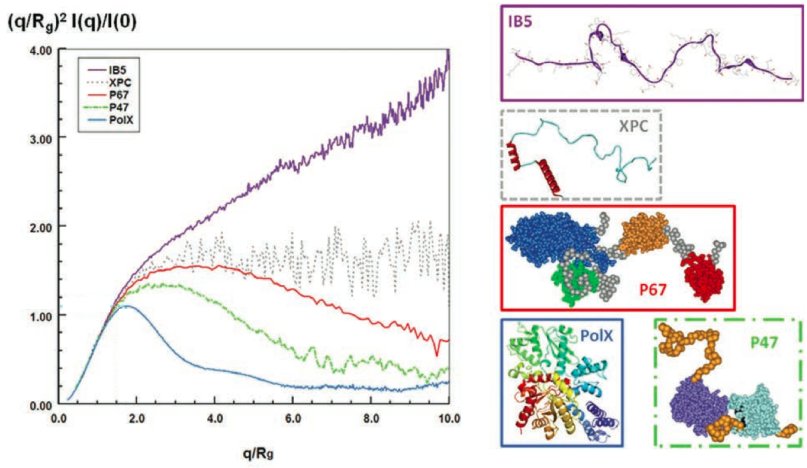
Seven Claspin protein fragments were expressed and purified as previously described (Section **5.3**) and then concentrated to roughly 1, 2.5, 5, 7.5 and 10 mg/ml. Samples were then frozen by flash freezing in liquid nitrogen and stored at -80 °C until required. Analysis of these samples was conducted at the ESRF synchrotron (Grenoble, France) on the BM29 BioSAXS beamline (section **2.23**). The Kratky plots and Guinier plots used for the analysis of the measured scattering data, were produced by the automated data processing software on the BM29 beamline, or with the software package PRIMUS (Konarev et al., 2006).

A Kratky plot [$I \times s^2$ versus s (where I = scattering intensity and s = scattering vector; where s can also be designated as q)] provides a convenient method to examine the 'foldedness' of a protein, as determined by its scattering curve. For a folded, globular protein, the Kratky plot is a bell-shaped curve with a distinct

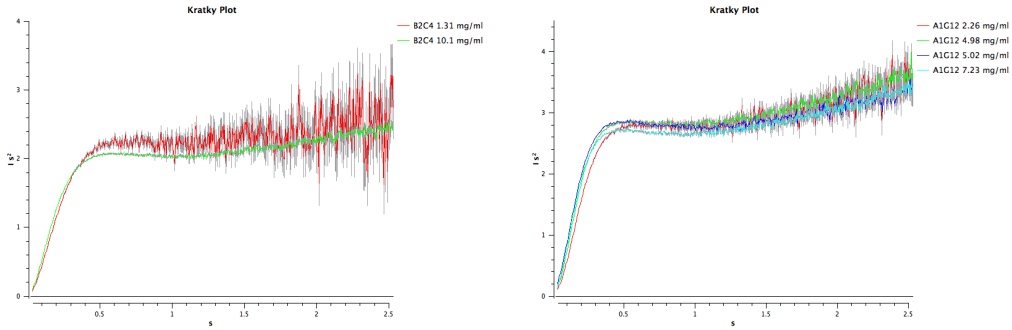
maximum. Increasing flexibility / elongation in the protein chain (with globular regions) results in a less distinct bell-shaped curve, whilst for randomly orientated proteins there is no bell-shaped curve and the plot shows a plateau or further increase at high q values (Receveur-Brechot and Durand, 2012) (**Figure 5.13A**). The Kratky plots for protein fragments B2C4 and A1G12, show a slight distinct maxima, but not a bell-shaped curve (**Figure 5.13B**); these plots are remnant of the grey plot in **Figure 5.13A**, potentially indicating some local secondary structure, but not a compact globular protein. The Kratky plots for B2C2 and B2D6, did not show a distinct maxima, but showed a plateau (**Figure 5.13C**); these plots are remnant of the purple plot in **Figure 5.13A**, indicating these protein fragments are highly disordered. This was also true for the Kratky plots of B2D9, B1F8 and A1G6D; although these plots showed more severe disorder and sample aggregation as the individual plots did not overlay with increasing protein concentrations (**Figure 5.13D**).

A Guinier plot (*ln[I(s)]* versus s^2) was also used to analyse the scattering data obtained for each protein fragment. This can be used to determine an apparent molecular mass and radius of gyration, and to ascertain if there is any aggregation present in a sample. The Guinier plots indicated that all the protein samples, with the exception of B2C2, showed varying levels of sample aggregation. Scattering from aggregates is most detectable at small *s* values, leading to spikes on the Guinier plots that are not linear; linear Guinier plots are required for accurate extrapolation of the data (Receveur-Brechot and Durand, 2012). As a result the B2D9, B1F8 and A1G6D fragments could not be analysed further (B2D9; **Figure 5.14A**). However, For B2C2 (**Figure 5.14B**) and three further samples B2C4, A1G12 and B2D6, the molecular mass and radius of gyration could be extracted from the Guinier plot for each of the protein concentrations measured. This data is summarised in **Table 5.2**. For both B2C4 and A1G12, there was a concomitant increase in both the radius of gyration and molecular mass with increasing protein concentration, thus indicating aggregation or multimerisation at the higher protein concentrations. At the lowest concentration tested (0.72 mg/ml) the molecular mass determined for A1G12 was identical to its calculated molecular mass; i.e. the monomeric form of protein. For B2C2 and B2D6, the molecular mass

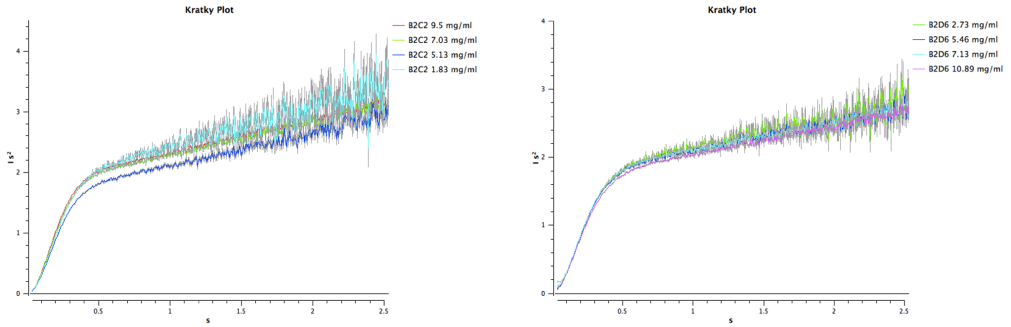
A


(q/Rg)² I(q)/I(0)
IB5
XPC
P67
P47
PolX
q/Rg
IB5
XPC
P67
PolX
P47

B


Kratky Plot
B2C4 1.31 mg/ml
B2C4 10.1 mg/ml
Kratky Plot
A1C12 2.26 mg/ml
A1C12 4.98 mg/ml
A1C12 5.02 mg/ml
A1C12 7.23 mg/ml

C


Kratky Plot
B2C2 9.5 mg/ml
B2C2 7.03 mg/ml
B2C2 5.13 mg/ml
B2C2 1.83 mg/ml
Kratky Plot
B2D6 2.73 mg/ml
B2D6 5.46 mg/ml
B2D6 7.13 mg/ml
B2D6 10.89 mg/ml

D

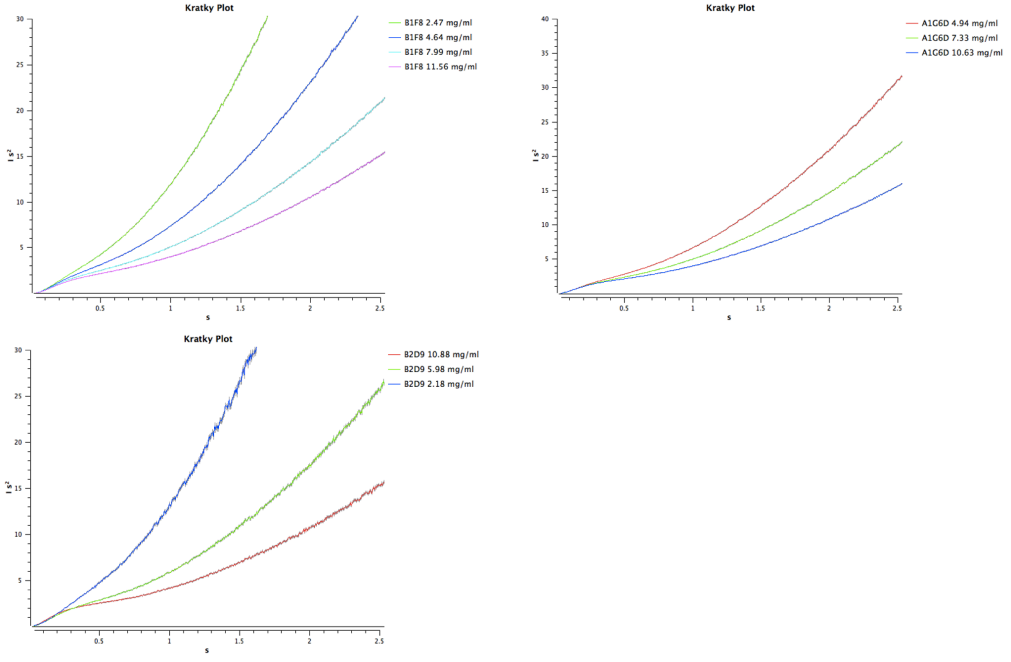

Kratky Plot
B1F8 2.47 mg/ml
B1F8 4.64 mg/ml
B1F8 7.99 mg/ml
B1F8 11.56 mg/ml
Kratky Plot
A1C6D 4.94 mg/ml
A1C6D 7.33 mg/ml
A1C6D 10.63 mg/ml
Kratky Plot
B2D9 10.88 mg/ml
B2D9 5.98 mg/ml
B2D9 2.18 mg/ml

Figure 5.13 bioSAXS Kratky plot analysis for the Claspin protein fragments. (**A**) Example Kratky plots. Scattering curves from a compact globular protein showing a bell-shaped curve (blue line / box), for multidomain proteins with compact globular regions but with an extended conformation (red and green lines / boxes), for an unstructured protein with short regions of secondary structure elements (grey line / box), or for a completely disordered protein in an extended conformation (purple box / line). Adapted from Receveur-Brechot and Durand, (2012). (**B**-**D**) Kratky plot analysis from scattering data collected using 20 µl sample with continuous flow at 4 °C using 100% transmission of the X-ray beam on the BM29 BioSAXS beamline at the ESRF synchrotron, Grenoble. Data was collected using 10 frames and averaged, and the buffer scattering was subtracted. Plotting the SAXS data as $I \times s^2$ versus s (Kratky plot) using Primus (Konarev et al., 2006). Protein concentration is shown top right on each plot. Where '**B**' (left) B2C4 and (right) A1G12, '**C**' (left) B2C2 and (right) B2D6, and '**D**' (top left) B1F8, (top right) A1G6D, and (bottom left) B2D9

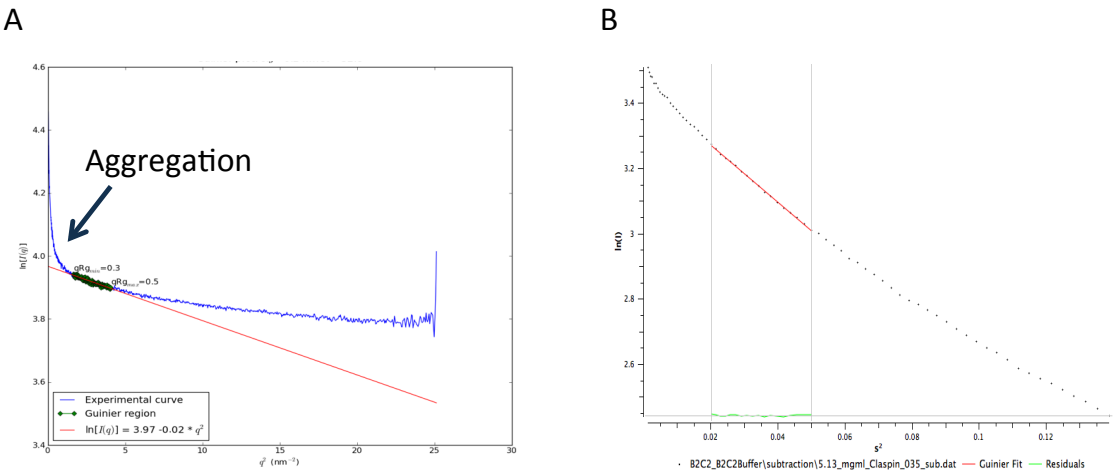


Figure 5.14 bioSAXS Guinier plots for Claspin protein fragments B2D9 and B2C2.

Gunier plot analysis for scattering data collected using 20 µl sample with continuous flow at 4 °C using 100% transmission of the X-ray beam on the BM29 BioSAXS beamline at the ESRF synchrotron, Grenoble. Data was collected using 10 frames and averaged, and the buffer scattering was subtracted. (**A**) Plotting the bioSAXS data as *ln[I(q)]* versus q^2 (Guinier plot) using automated in house software (ESRF, Grenoble) for B2D9 protein at 0.47 mg/ml. The scattering curve is represented by the blue curve, the linear Guinier region is represented by the green dots, and the red line shows the extrapolation for molecular mass determination; refer to the legend in the figure. Arrow is indicative of a non-linear Guinier plot. (**B**) Plotting the bioSAXS data as *ln[I(s)]* versus s^2 (Guinier plot) using Primus (Konarev et al., 2006) for B2C2 protein at 5.13 mg/ml. The scattering curve is represented by grey dots, the linear Guinier region and extrapolation is represented by the red line; refer to legend in the figure. *s* and *q* are interchangeable.

and radius of gyration remained fairly constant with increasing protein concentration, and the determined molecular mass was similar to the predicted masses of the monomer forms of each protein. However, an increase in the calculated molecular mass and radius of gyration was seen at the highest concentration for B2C2 (9.5 mg/ml), potentially indicating some aggregation or association of the sample.

Protein Sample	Molecular Weight (kDa)	Concentration (mg/ml)	Guinier Plot		
			Aggregation (Y/N)	Rg (Å)	Io (kDa)
B2C4	31.08	0.33	Y	5.66	40.39
		1.31	Y	6.38	48.10
		10.1	Y	7.70	63.10
A1G12	36.52	0.72	Y	4.49	36.28
		1.04	Y	5.27	49.47
		2.26	Y	5.88	61.50
		4.98	Y	6.74	81.60
		7.23	Y	7.67	95.71
B2C2	26.41	1.83	N	5.19	33.78
		5.13	N	5.12	31.39
		7.03	N	5.04	34.58
		9.5	N	5.52	37.55
B2D6	23.67	2.73	Y	4.68	27.85
		5.46	Y	5.03	29.20
		7.13	Y	4.93	29.20
		10.89	Y	4.73	26.92

Table 5.2 Summary of Guinier plot calculations.
Aggregation: Y=yes, N=no. Rg=radius of gyration, Io=calculated molecular mass.

The radius of gyration of each fragment is very large for the determined molecular mass. The radius of gyration for a number of proteins standards [as determined by bioSAXS (Mylonas and Svergun, 2007)] previously used to calibrate the ASEC column include; 7.05 Å for ferritin (440 kDa), 3.51 Å for aldolase (158 kDa) and 2.66 Å for ovalbumin (43 kDa) (Mylonas and Svergun, 2007). From SAXS analyses, the four Claspin protein fragments (at the lowest protein concentration) had calculated radii of gyration of between 4.49 and 5.66 Å, and each protein fragment has a predicted relative molecular mass of below 36.52 kDa. However, the four Claspin proteins eluted at or before aldoase (3.51 Å), thereby the predicted radius of gyration is larger than would have been expected from the ASEC retention volumes. This potentially indicated the samples were not stable after flash freezing; which is discussed in section **5.15**.

Together, these data indicate that these protein fragments are predominantly unstructured in solution. The SAXS data is therefore consistent with previous experiments.

5.13 Nuclear Magnetic Resonance Spectroscopy

NMR was also used to study the A1G12 (36.52 kDa) and B2C2 (26.41 kDa) protein fragments; again in order to examine their degree of 'foldedness' in solution. For these studies, A1G12 and B2C2 were expressed in ^{15}N-labelled auto-induction medium. For A1G12, there was almost no detectable recombinant protein expression in this medium. B2C2 showed relatively low levels of recombinant protein, but this could be purified as previously described (see section **5.3**). Purified B2C2 was concentrated to 260 μM in Claspin GF Low Buffer and supplemented with 10% (v/v) deuterium oxide; this was somewhat below the typical concentration of ~500 μM that is normally used for this type of experiment. NMR data was collected on a Varian VNMRS 600 MHz with the assistance of Dr Iain Day (University of Sussex).

A ^{1}H–^{15}N HSQC spectrum was collected for B2C2, but this showed a very compact spectrum with very few distinct peaks (**Figure 5.15**); a well-dispersed spectrum, with distinct peaks is indicative of a folded protein, whereas condensed peaks overlapping in the centre of the plot is indicative of an unstructured protein.

5.14 Tryptic digests

Limited proteolytic digestion can often identify the boundaries of domains from larger protein entities. In order to investigate whether this technique would yield any smaller folded regions from the N-terminal Claspin fragments, tryptic digests were carried out. Trypsin cleaves the peptide backbone on the carboxyl-side of a solvent accessible (elongated or unfolded) lysine or arginine (except when followed by proline). Examination of the lysine and arginine amino acids within the N-terminal region of Claspin showed that they were regularly spaced, and if all theoretical cleavage sites were cut, the largest peptide left would be around 6 kDa (aa 470-526) (ExPASy PeptideCutter tool; Gasteiger et al., 2005); thereby the

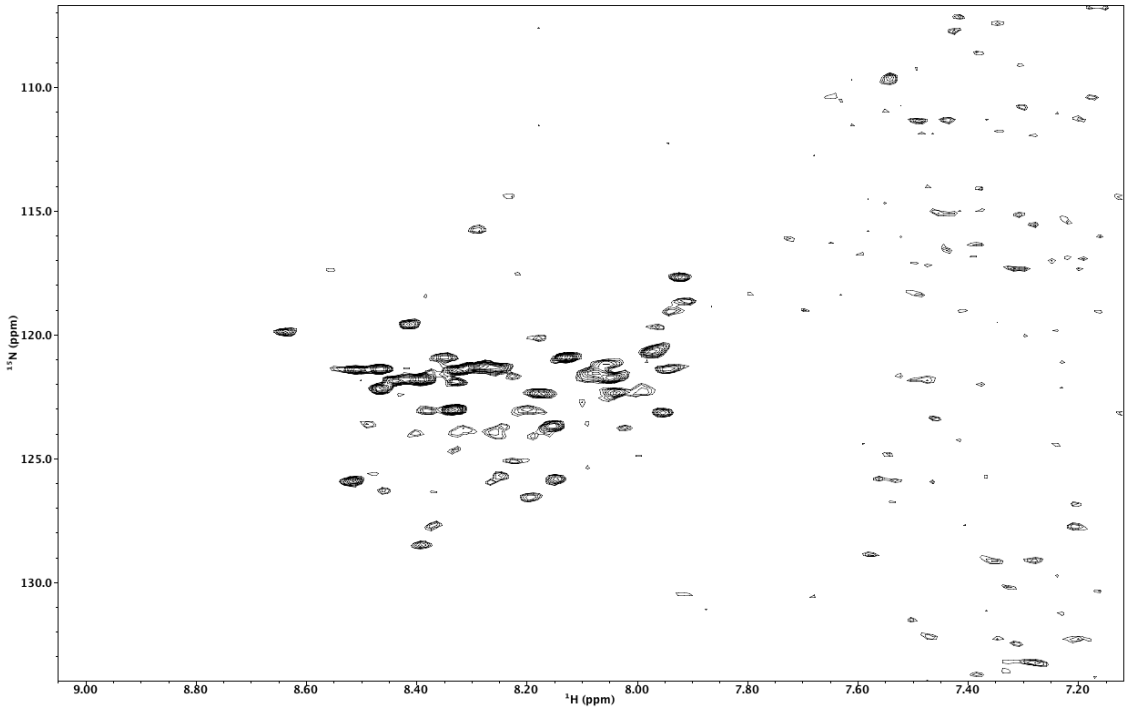
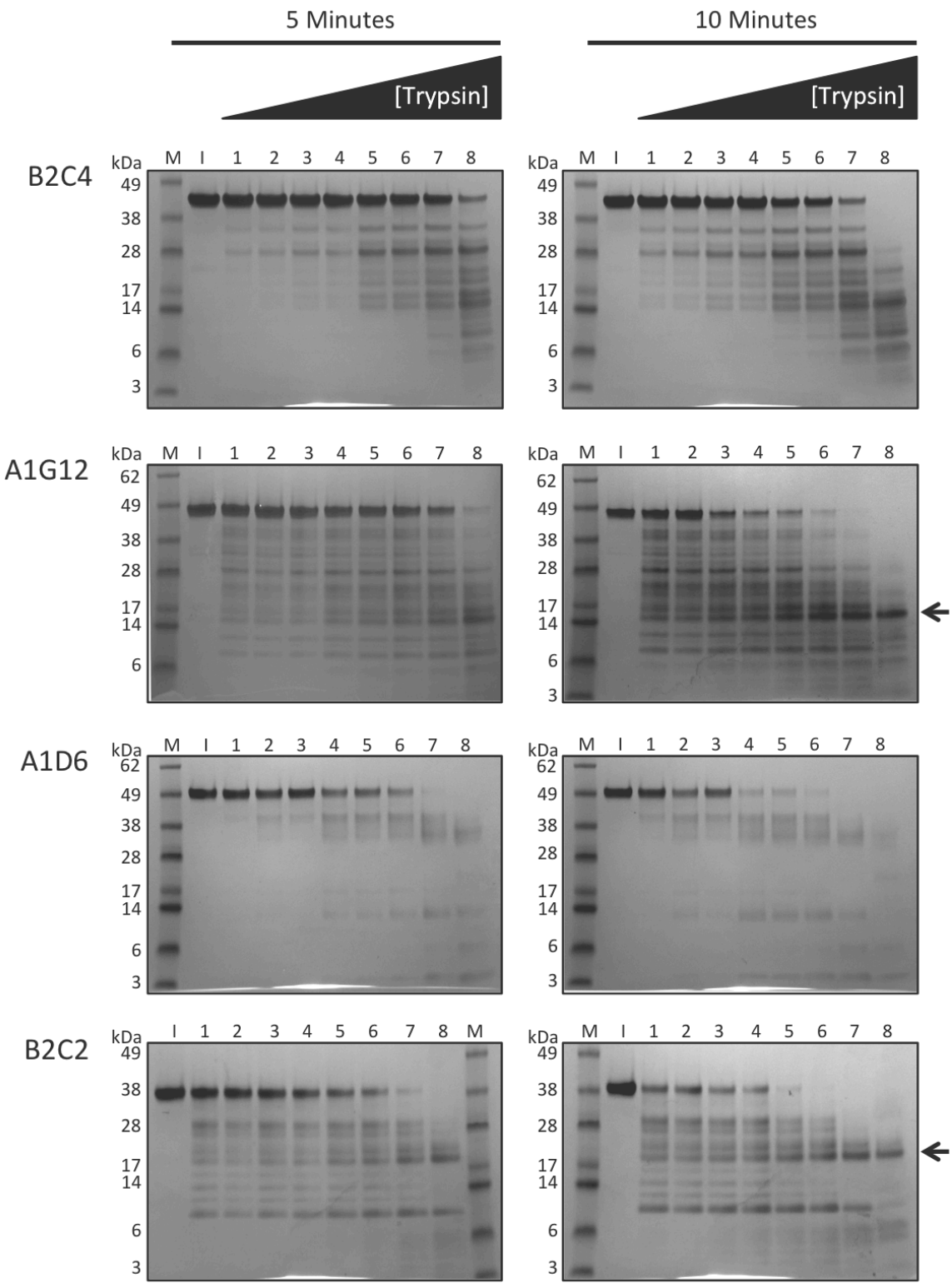


Figure 5.15 ^{1}H–^{15}N HSQC spectrum for the Claspin protein fragment B2C2.
^{1}H–^{15}N HSQC NMR spectrum for ^{15}N labelled purified B2C2 (260 μM) in 10% (v/v) deuterium oxide, was collected at 600 MHz using Varian VNMRS 600 MHz for 3 hours at 20 °C. NMR data collection and spectrum processing was by Dr Iain Day (University of Sussex).

identification of a proteolytically resistant band by SDS-PAGE, might indicate a region of folded structure.

Digests were carried out at different concentrations of trypsin (generated by serial dilution) for a fixed period of time, and subsequently analysed by SDS-PAGE. Initial examination of the gels showed there had been digestion of the protein, with a clear difference between the five and ten minute incubation points (**Figure 5.16**). Fragments B2C4, A1G12, A1D6, B2C2 and B1F8, appeared to be the most resistant to digestion by trypsin, whereas for A1G6D, B2D6 and B2D9 no obvious resistant species could be identified. Curiously, when the fragment B2D6 was digested, a product migrating on SDS-PAGE with a larger relative molecular mass than the full-length fragment was seen. Western blot analysis of this band indicated that the His_6 affinity tag was not present, indicating it did not represent the C-terminus of the Claspin fragment, thereby the slower migration of this fragment, might be due to changes in the net charge, relative to the full-length B2D6 protein.

For digests of A1G12 and B2C2, species resistant to digestion were identified (~14 kDa for A1G12 and ~17 kDa for B2C2) (marked by arrows in **Figure 5.16**), which remained stable for 15 minutes (marked by arrows in **Figure 5.17A**). Of note, these were found to not follow the linear elution relationship by ASEC found for the further Claspin protein fragments as identified in section **5.5**. A1G12 and B2C2 contain a significant region of overlap (aa 218-413), therefore a trypsin resistant fragment may indicate the presence of a region of folded structure. An anti-His western blot showed that these proteolytic fragments retained the C-terminal His_6 affinity tag (**Figure 5.17B**). B2C2 has an additional 30 amino acids at its C-terminus in comparison to A1G12, which taken with the observed differences in migration by SDS-PAGE, indicated that these fragments might in fact share a common N-terminal sequence. Edman degradation was therefore used to determine the N-terminal sequence for both protease-produced fragments. No sequence was obtained from the B2C2-derived band, whilst for the A1G12-derived species, two potential N-terminal sequences were determined – 'MPENK' (aa 308) and 'YQS' (aa 339). Both these sequences are located in the predicted HTH motifs, and would result in the loss of this motif in the observed degradation products [aa


5 Minutes
10 Minutes
[Trypsin]
[Trypsin]
B2C4
A1G12
A1D6
B2C2
kDa
M I 1 2 3 4 5 6 7 8
49
38
28
17
14
6
3
62

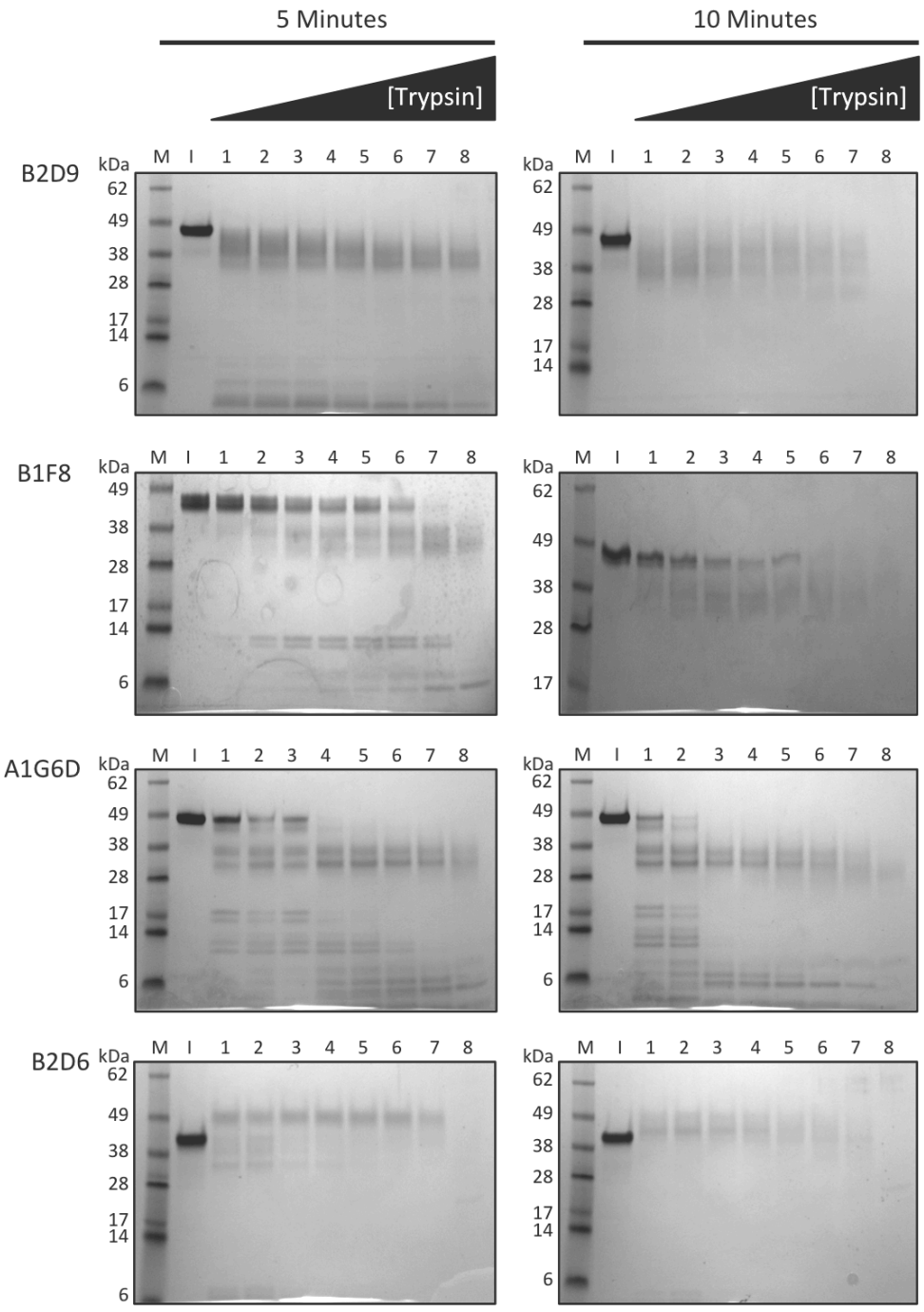
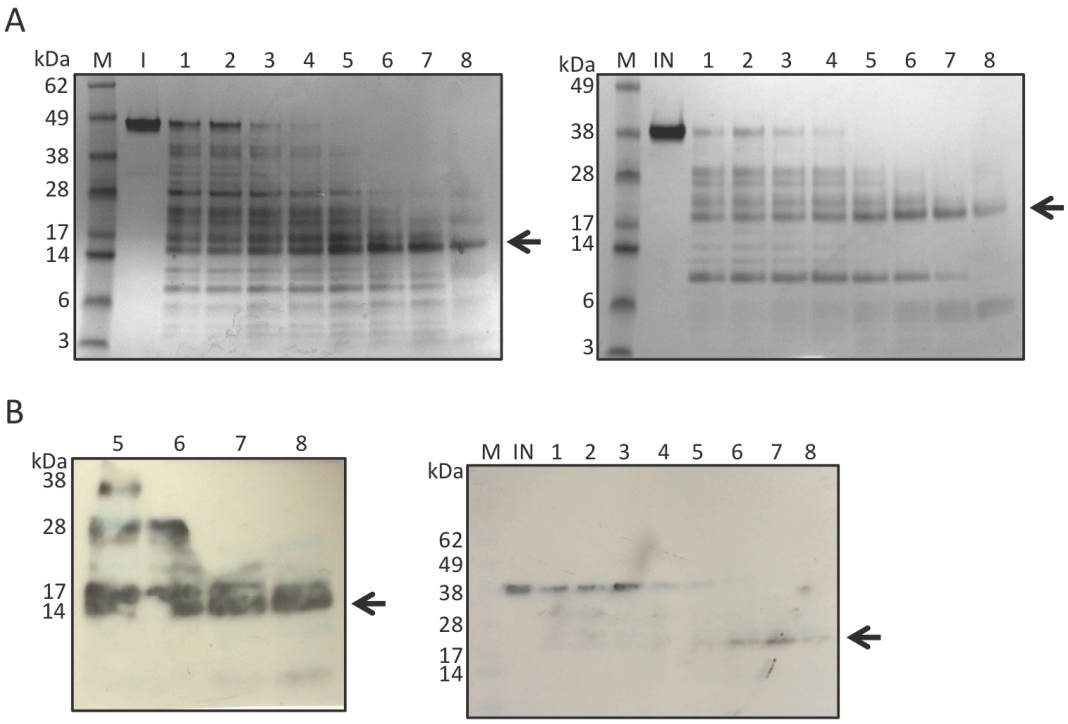


Figure 5.16 Limited proteolysis of the Claspin protein fragments.
SDS-PAGE analysis of limited proteolysis of purified Claspin protein fragments (10 μM), after incubation with dilutions of trypsin protease for (left) 5 or (right) 10 minutes. Samples are as labelled on the left of the gels, and the increase in trypsin protease concentration is shown. The arrows indicate stable protein fragments. M=molecular mass marker, I=input sample (no added trypsin), 1 to 8=trypsin protease concentration: 1=0.01 nM, 2=0.14 nM, 3=1.67 nM, 4=2 nM, 5=2.5 nM, 6=3.33 nM, 7=5 nM, 8=10 nM. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



C

YKTVADSDESYMEKSLYQENLEAQVKPCLELSLQSGNSTDFTTDRKSSKKHIHDKE
GTAGKAKVKSKRRLEKEERKMEKIRQLKKKETKNQEDDVEQPFNDSGCLLVDKDLF
ETGLEDENNSPLEDEESLESIRAAVKNKVKKHKKKEPSLESGVHSFEEGSELSKGT
TRKERKAARLSKEALKQLHSETQRLIRESALNLPYH**MPENK**TIHDFFKRKPRPTCH
GNAMALLKSSK**YQS**SHHKEIIDTANTTEMNSDHHSKGSEQTTGAENEVETNALPVV
SKETQIITGSDESCRKDLVKNEELEIQEKQ

Figure 5.17 Identification of the trypsin protease resistant protein fragments.
(**A**) SDS-PAGE analysis of limited proteolysis of purified (left) A1G12 and (right) B2C2 (10 μM), after incubation with dilutions of trypsin protease for 15 minutes. M=molecular mass marker, IN=input sample (no added trypsin), 1 to 8=trypsin protease concentration: 1=0.01 nM, 2=0.14 nM, 3=1.67 nM, 4=2 nM, 5=2.5 nM, 6=3.33 nM, 7=5 nM, 8=10 nM. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue. (**B**) Identification of the breakdown products of (left) A1G12 and (right) B2C2 using a colourmetric anti-His western blot (anti-His primary and AP-conjugated secondary antibodies). The arrows indicate stable protein fragments. (**C**) Stable trypsin protease fragment identification. Full-length A1G12 sequence without the His_6 affinity tag. The regions identified from mass spectrometry are in red type, and the two identified N-termini peptides from Edman degradation are underlined and in bold font.

279-313 (Zhao and Russell, 2004) and aa 276-342 (SwissModel; Schwede, 2003)]. Mass spectrometry was also used to confirm the sequence of the two stable proteolytic species derived from A1G12 and B2C2. This method identified a peptide containing a continuous region of the C-terminus for both A1G12 and B2C2, starting at 'YQS' (aa 339) and 'EIID' (aa 346) respectively. Additional peptides N-terminal to this sequence were also identified (**Figure 5.17C**). Curiously, the 'YQS' sequence was previously identified as the cleavage point for the 'initial hit' A1G6, which yielded the breakdown product and subsequently re-cloned fragment A1G6D (section **5.3**).

5.14.1 Expression and purification of the proteolytically stable fragments

To further investigate the proteolytic resistant fragments of both A1G12 and B2C2, the DNA encoding these regions was sub-cloned into the pDXV4 vector; using the N-terminal 'MPENK' (aa 308) or 'YQS' (aa 339) sequences, combined with either the C-terminal sequence of A1G12 (aa 413) or B2C2 (aa 441) to create four expression constructs; MPENK-1 (aa 308-413), MPENK-2 (aa 308-441), YQS-1 (aa 339-413) and YQS-2 (aa 339-441). These were expressed and purified as described in section **5.3**.

No detectable expression could be seen for either the YQS-1 or YQS-2 constructs. An example for the purification of the MPENK-1 and MPENK-2 fragments is shown in **Appendices 2.19** and **2.20** respectively. After the Talon-IMAC chromatography step, both proteins were found to run at a higher than expected relative molecular mass, when compared to the original proteolytically-generated species. Additional purification steps using Heparin chromatography and preparatory SEC, revealed the presence of a degradation product, which migrated on SDS-PAGE at roughly the molecular mass as expected for the original proteolytically-generated species.

Optimal expression constructs for the proteolytically stable protein fragments were not identified during this process. Whilst a proteolytic stable protein fragment can appear by SDS-PAGE, it is not always possible to express the same fragment as a recombinant construct in a stable manner. There can be a requirement for N- and/or C-terminal regions to the construct identified, which

are cleaved and therefore not identified; this is a disadvantage of the limited proteolysis technique.

5.15 Summary

In this Chapter, eight soluble N-terminal protein fragments of human Claspin, identified by CDH, were characterised by a variety of biochemical and biophysical techniques. Computer-based secondary structure predictions indicated that this region of Claspin is mostly α-helical in nature, but with a high content of disorder. Molecular modelling techniques identified a potential HTH motif that had previously been predicted to have DNA binding properties (Zhao and Russell, 2004) (see **Chapter 6** for further discussion). Subsequent biochemical and biophysical studies included ASEC, CD, thermal denaturation, chemical crosslinking and AUC. Together the data from these studies indicated that the Claspin fragments are elongated in shape and contain a high degree of intrinsic disorder. Formation of secondary structure (α-helix) was inducible, to some extent, by incubation of the protein fragments with the solvent TFE; indicating that there may be some disorder-to-order transition for this part of Claspin when bound to its cognate partner or substrate.

Previously a low-resolution EM image of Claspin bound to branched DNA was published (Sar et al., 2004), but as the protein fragments generated here are too small to be studied by EM, it was not possible to use this technique; therefore alternative structural biology-based techniques were used. Exhaustive crystallographic trials failed to produce any crystals. SAXS was employed to analyse each construct and sample aggregation was observed for samples, except for B2C2; this may have caused by the necessity to freeze-thaw the samples for transport to the bioSAXS beamline at ESRF, France. Subsequently protein fragments A1G12, B2C2, B2D9 and B1F8 were found to be stable upon storage at 4 °C for a minimum of two weeks, by SDS-PAGE analysis (see **Appendices 2.3F** and **2.9F**) and ASEC (data not shown), however the effects of freezing the samples was not fully investigated. From the scattering data collected, this indicated that the N-terminal region of human Claspin is not globular in solution and is likely to have a high degree of intrinsic disorder (from Kratky plots), and has an elongated shape,

as shown from the radius of gyration and molecular mass calculation (Guinier plot). The BM29 bioSAXS beamline has recently had an FPLC added to the experimental setup, which allows an ASEC column to be connected in-line. This facilitates the separation and analysis of distinct elution peaks, and enables the removal of sample aggregates (these typically elute at the void volume of the column) (Pernot et al., 2013). However, as this facility was unavailable during the time of this study, no further SAXS experiments were carried out. Furthermore, attempts to characterise selected protein fragments by NMR proved unsuccessful due to the very low yields of recombinant protein expression from in the ^{15}N-labelled medium, although some indications as to the large amount of intrinsic disorder present in the B2C2 protein fragment was obtained.

As a final attempt to identify any structurally ordered region or sub-domains within the N-terminal Claspin fragments, limited proteolysis experiments were undertaken. This method identified a putatively proteolytic-resistant species derived from two overlapping fragments. Future work could involve experiments to test if these proteolytically-derived fragments truly represent a folded region / sub-domain of Claspin.

The results presented in this chapter, as a whole, indicate that the N-terminal region of human Claspin is intrinsically disordered. Claspin is described as an 'adaptor' protein, and has been found to make a large number of protein-protein interactions, with both DNA-replication and DNA-repair associated proteins, but very few of these interactions mapped directly on the Claspin protein (Kumagai and Dunphy, 2000, Lee et al., 2005, Brondello et al., 2007, Kim et al., 2008, Gold and Dunphy, 2010, Nakaya et al., 2010, Uno and Masai, 2011, Serçin and Kemp, 2011, Rainey et al., 2013, Broderick et al., 2013).

It is known that proteins that are involved in a large number of different interactions have a higher probability of intrinsic disorder, which actually helps mediate such interactions and thereby provide functionality (Cortese et al., 2008). Segments of full-length proteins that lack a specific tertiary structure contain so-called intrinsically disordered regions (IDRs), where complete structural disorder

these can be further classified as intrinsically disordered proteins (IDPs). IDRs can exist as part of a multi-domain protein as the large flexible linkers between globular domains, or as part of proteins that exist in a pre-molten globule state, or in proteins that are either partially or completely unfolded. Extensive work is being carried out by several different laboratories, looking at IDRs and IDPs in order to ascertain their function within a cellular environment and many reviews on this topic have recently been published (Dyson, 2011, Tompa, 2012, Uversky, 2013, Liu and Huang, 2014, Dunker et al., 2015). IDRs or IDPs can be: conformationally plastic, i.e. becoming ordered upon binding to a partner protein, nucleic acid or metal ions; exist to facilitate the interaction of multiple binding partners; or simply function as a region capable of being PTM and thus to allow subsequent interaction with binding partners (Liu and Huang, 2014, Dunker et al., 2015). These types of region or protein have been identified in many organisms and are often associated with highly regulated processes such as the cell cycle and transcriptional programmes (Dunker et al., 2015). Furthermore, many IDPs have been found to be tightly regulated and have implications in a number of human diseases, including cancer (Babu et al., 2011).

Structural characterisation of this type of protein / region can be achieved by NMR and / or SAXS methodologies; but to truly understand their function, a study of their various interactions is also required. Experimental work that could follow this Thesis might involve the identification and mapping of the regions of Claspin required to interact with its various protein partners *in vivo* and verifying these *in vitro*. Co-crystallisation trials could then be carried out with these various complexes.

Chapter 6

DNA-binding ability of the N-terminal region of human Claspin

6.1 Introduction

The N-terminal region of human Claspin is reported to have DNA binding functionality, in particular a region termed the RFID (aa 265-605) has been shown to interact with chromatin, which contains the two basic sequence patches BP1 (aa 265-331) and BP2 (470-600) (Lee et al., 2005). Deletion of either basic region, or mutation of selected conserved residues, directly affected chromatin binding in Xenopus extracts, whereas deletion of the intervening region (aa 376-425) had no effect (Lee et al., 2005). In addition, a region of human Claspin, dubbed the DBD (aa 149-340) has been shown to interact with DNA *in vitro*, by both gel-shift and by pull-down assay using an immobilised DNA molecule as bait (Sar et al., 2004, Serçin and Kemp, 2011, Uno and Masai, 2011). A previous amino acid sequence conservation analysis of both Claspin and MRC1, published by Zhao and Russell, (2004), identified the presence of a potential HTH motif (aa 279-313) (Zhao and Russell, 2004) – a motif commonly associated with nucleic-acid binding (Brennan and Matthews, 1989). Furthermore, using the SwissModel homology-modelling server (Schwede, 2003) this also identified a HTH motif (aa 276-342), which was modelled on part of a bromodomain fold from a number of different templates (section **5.4**). For alignment of the N-terminal Claspin protein fragments with selected identified sequence regions refer to **Figure 5.2**.

The propensity of Claspin to bind DNA, shown in the literature by EMSAs and DNA pull downs, indicates that it has a preference for dsDNA over ssDNA, and a stronger preference for branched-DNA that includes overhangs, Y-shaped structures and fork structures, whilst having no sequence specificity (Sar et al., 2004, Tanaka et al., 2010, Serçin and Kemp, 2011, Uno and Masai, 2011, Yilmaz et al., 2011). Similar DNA-binding preferences were also shown for the yeast homologue Mrc1 (Zhao and Russell, 2004). It has also been reported that there is a preference for the binding of Claspin to DNA containing bulky adducts (Yilmaz et al., 2011).

Whilst DNA-binding capability has been crudely mapped to the N-terminus of Claspin, the precise amino acids / range of amino acids involved has not been determined. As the CDH fragment library contained a number of constructs that

expressed high levels of soluble protein and overlapped the regions with predicted DNA binding propensity, these fragments were investigated for their ability to bind DNA, using a range of techniques including EMSA, FP and ASEC, with the aim of identifying a smaller, sub-region of the N-terminus that encapsulated the DNA-binding functionality, for subsequent structural characterisation.

6.2 EMSAs

EMSAs or 'gel-shifts' can be used to test if a particular protein is capable of binding to nucleic acid. Typically DNA is labelled (either radioactively or fluorescently) incubated with increasing concentration of protein, before application of the sample to a Native-PAGE gel. Binding of protein to the labelled DNA, results in a retarded migration with respect to the unbound DNA. This method can be used to investigate many parameters of the protein:DNA interaction including stoichiometry, substrate preference and effects of protein mutation. As a preliminary assay to determine if any of the eight Claspin N-terminal fragments bound to DNA, EMSA experiments were prepared at a 10:1 protein:DNA molar ratio. This was to ensure that the DNA was fully saturated, and that presence of a protein:DNA complex could be readily visualised. Blunt-ended ds- or ssDNA of different lengths (20, 25, 30 and 35 bp / nt) were used as binding substrates, in order to probe if Claspin had any size-preference; as this had not previously been examined. All DNA oligonucleotides were commercially synthesised with a 5'-Cy5 fluorescent label to enable detection by fluorimetry.

The EMSA experiments indicated the protein from four expression constructs (B2C4, A1G12, A1D6 and B2C2) of the eight tested, bound to DNA, with a clear preference for dsDNA over ssDNA; in particular binding to DNA duplexes of longer length, with only relatively weak interactions with ssDNA (**Figure 6.1**). The protein from the BD29 construct did show a very slight shift with longer lengths of dsDNA, but this was significantly weaker interaction that those observed for the other four proteins. Notably, each of these five expression constructs encodes at least part of the predicted DBD (Sar et al., 2004) (see **Figure 5.2A**). A complete shift of DNA, into a complex, was not observed in any of the experiments even at a 10-fold molar excess of protein. Potentially the protein-DNA complexes may not be

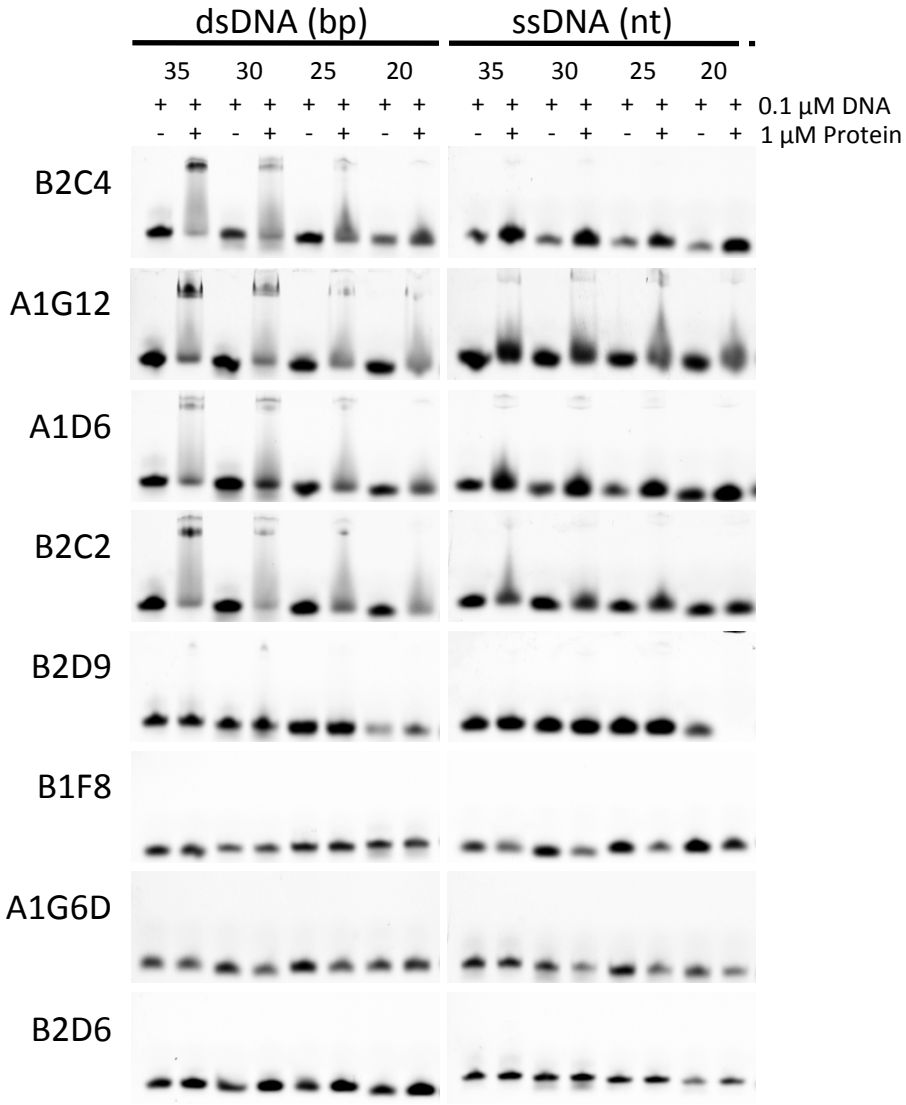


Figure 6.1 EMSAs for the Claspin protein fragments.
EMSAs were carried out using purified protein fragments against varying lengths of fluorescently labelled (5'-Cy5) (left) dsDNA (35, 30, 25 or 20 bp) or (right) ssDNA (35, 30, 25 or 20 nt). The protein fragments (2 µM) were incubated at a 10:1 ratio with the DNA (0.2 µM) in EMSA binding buffer for 20 minutes at room temperature, before being applied to a 5% acrylamide Native-PAGE gel. Gels were imaged using the FLA-5100 fluorescent imager (FujiFilm). The identity of each expression construct tested is given on the left-hand side and the DNA is indicated above.

stable under the conditions used for the EMSA experiments and sample resolution in the gel matrix could potentially disrupt the interaction. Therefore in order to further characterise the Claspin-DNA interactions, the alternative technique of FP was used, in order to determine binding constants for each interaction.

6.3 Fluorescence Polarisation

FP is an equilibrium technique, which enables the study of a particular interaction in solution and a determination of the equilibration dissociation constant (K_d) for that interaction (refer to section **1.11.1**). This enables homogenous samples to be studied without separation, and binding is detected by a change in the fluorescence signal, which allows for a more accurate comparison of the binding affinities of the different protein fragments with the DNA substrates; but the stoichiometry of the interaction is unable to be determined. To determine if any DNA binding was occurring in solution by the Claspin N-terminal fragments, the smallest ss- and dsDNA oligonucleotides previously used in EMSA were used. These DNA oligonucleotides were labelled at the 5'-end with FITC, instead of Cy5 as a label compatible with the multimode plate reader used to measure polarisation in this assay format. Each of the eight protein fragments were prepared as serial dilutions and then incubated with 100 nM 5'-FITC-DNA for 15 minutes at room temperature. Polarisation was measured using a POLARstar Omega multimode microplate reader using a fluorescence polarisation optic set. Measurements were collected in triplicate, averaged and the signal from the unbound fluorophore control subtracted. Data were plotted, and then fitted by non-linear regression, to a one-site specific binding with Hill slope equation using GraphPad Prism. In the analysis of FP data, we consider any signal below 25 mP, after baseline correction, to be non-specific.

It was possible to fit binding curves, and determine binding constants for dsDNA- and ssDNA-binding, for the data generated with the protein fragments A1G12, B2C2 and B2D9, as they each had typical rectangular hyperbola binding curves that reached saturation (plateau) at the higher protein concentrations tested (dsDNA; **Figure 6.2**, and ssDNA; **Figure 6.3**). For the five further protein fragments, this was not possible. The K_ds determined by FP, for the fragments

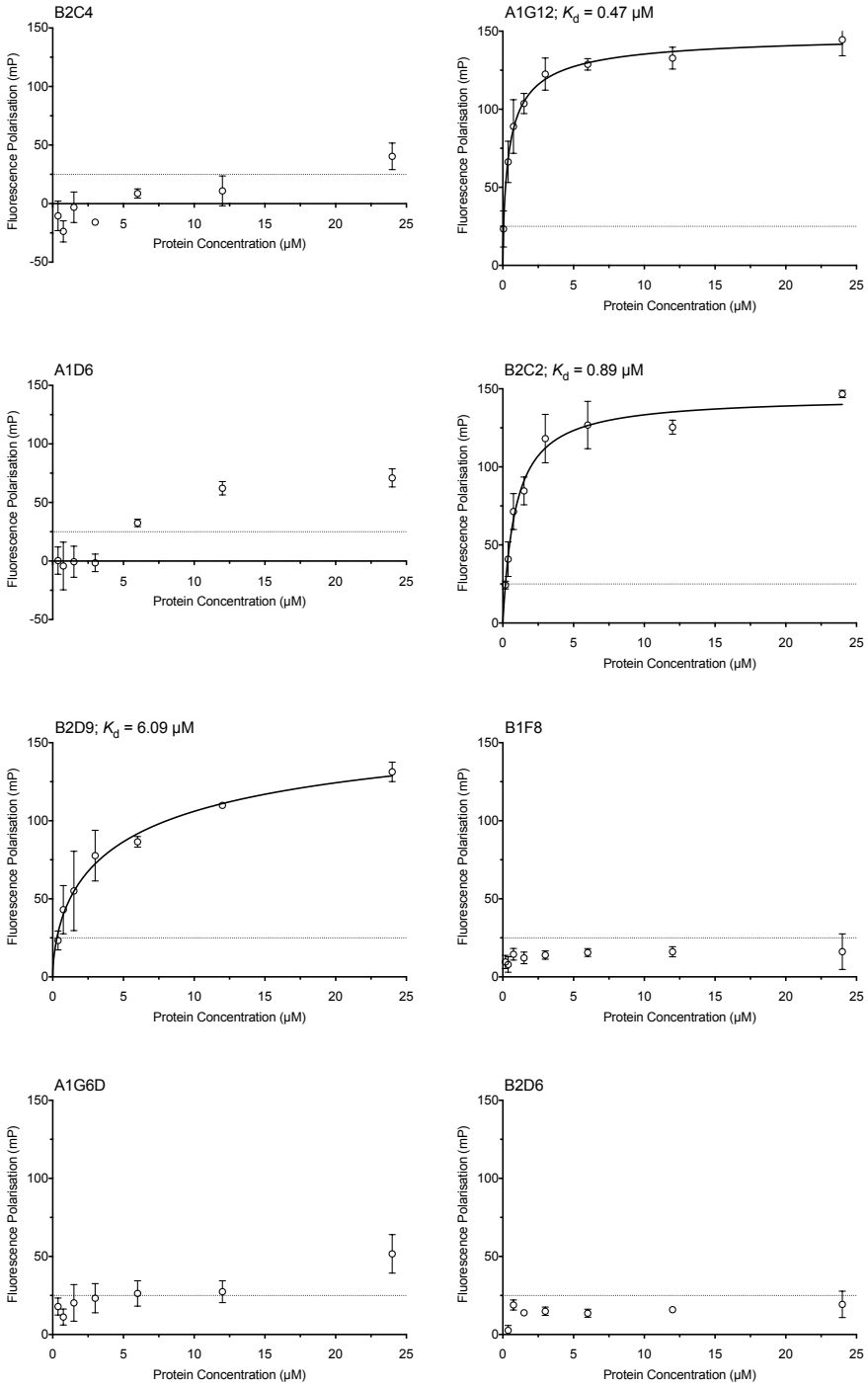


Figure 6.2 Identifying dsDNA binding from the Claspin fragments.
Binding of the Claspin protein fragments to 5'-FITC-20 bp dsDNA was tested by FP. In each case, the dsDNA was at a final concentration of 100 nM. Each data point is the mean of three individual experiments, where the error bars represent one standard deviation. The identity of each N-terminal protein fragment is indicated in the top left hand corner of each graph. Binding constants (K_d) for each experiment are shown, where they could be calculated.

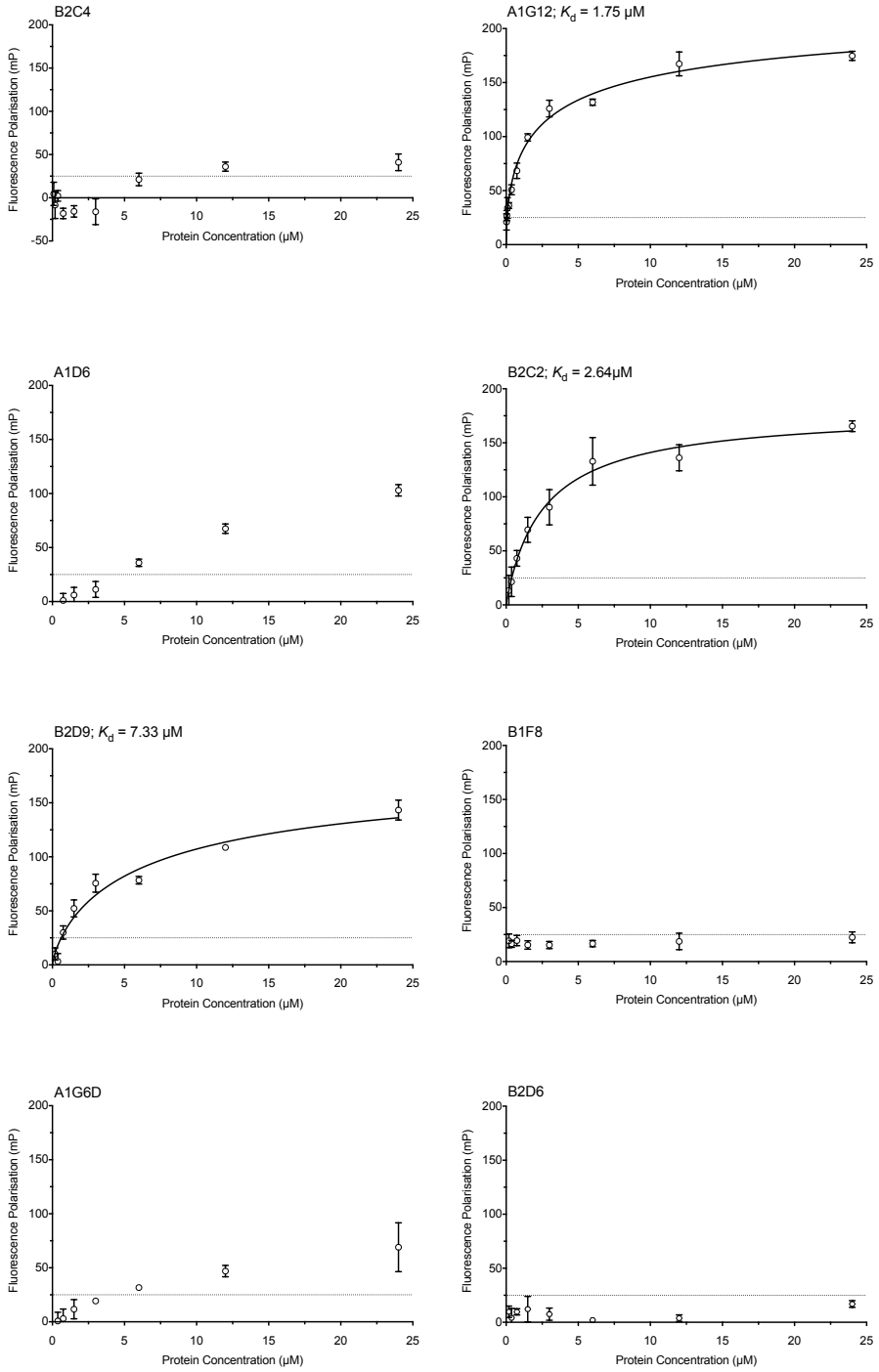


Figure 6.3 Identifying ssDNA binding from the Claspin fragments.
Binding of the Claspin protein fragments to 5'-FITC-20 nt ssDNA was tested by FP. In each case, the ssDNA was at a final concentration of 100 nM. Each data point is the mean of three individual experiments, where the error bars represent one standard deviation. The identity of each N-terminal protein fragment is indicated in the top left hand corner of each graph. Binding constants (K_d) for each experiment are shown, where they could be calculated.

A1G12, B2C2 and B2D9, were all in the micromolar range when tested against 20 bp / nt ds- and ssDNA templates; with increasing K_d for DNA binding in the order A1G12 < B2C2 < B2D9 (**Table 6.1**). One protein fragment, A1D6, showed a weak interaction with both ds- and ssDNA but the K_ds could not be accurately determined. No interaction with either ds- or ssDNA was identified for B2C4, B1F8, A1G6D or B2D6. Of note, the previously determined K_ds by Sar et al., (2004) for the interaction of full-length Claspin with DNA as determined by 50% EMSA shift found the K_d for a 50 bp dsDNA was K_d = > 0.5 μM and for 50 nt ssDNA was K_d = > 1.0 μM. These K_ds are almost identical to the K_ds determined by FP for the binding of A1G12 protein fragment to 20 bp dsDNA (K_d = 0.47 μM) and to 20 nt ssDNA (K_d = 1.75 μM).

Protein Fragment	K_d (μM)	
	dsDNA	ssDNA
B2C4	-	-
A1G12	0.47	1.75
A1D6	-	-
B2C2	0.89	2.64
B2D9	6.09	7.33
B1F8	-	-
A1G6D	-	-
B2D6	-	-

Table 6.1 Binding constants for the Claspin fragments with DNA.
Binding constants (K_d) for the N-terminal Claspin protein fragments for 20 bp / nt ds- and ssDNA, as shown in **Figures 6.2** and **6.3**.

From FP analysis, DNA binding capability has now been demonstrated for each of the Claspin fragments, which contain a predicted HTH motif (aa 279-313) (Zhao and Russell, 2004) and BP1 (aa 265-331) (Lee et al., 2005); these are A1G12, A1D6, B2C2 and B2D9, although the strength of the interaction varies with the construct.

6.3.1 Examining DNA length as a requirement for binding

The length of DNA required for the interaction of Claspin with DNA has not previously been examined. To address this question, the protein fragment with the highest affinity for DNA (A1G12) was first tested in additional FP experiments,

using both ss- and dsDNA of different lengths of between 10 and 40 nt / bp ss- and dsDNA (as in section **6.3**).

When longer than 18 bp, the calculated binding constants for A1G12 binding to dsDNA were highly similar (0.43 to 0.56 μM) (**Figure 6.4A**). However, when the DNA length was reduced to less that 18 bp, a slight rise in the K_ds was observed. It was not possible to accurately fit the data once the length of dsDNA had been reduced to 15 bp or less (**Figure 6.4B**). This was similarly observed for ssDNA when longer than 18 nt, the binding constants were calculated and similar, although these were not as consistent as for dsDNA binding (0.74 to 1.75 μM) (**Figure 6.4C**). As was identified for dsDNA, when the ssDNA length was reduced to less that 18 nt, a slight rise in the K_ds was observed, and it was not possible to accurately fit the data once the length of ssDNA was 15 nt or less (**Figure 6.4D**). The calculated binding constants for A1G12 binding to varying lengths of ds- and ssDNA are presented in **Table 6.2**.

DNA length (bp / nt)	K_d (μM)	
	dsDNA	ssDNA
40	0.56	0.74
30	0.53	0.90
20	0.47	1.75
19	0.43	1.43
18	0.55	1.14
17	0.77	1.93
16	0.95	1.82
15	-	-
10	-	-

Table 6.2 Binding constants for A1G12 protein with DNA.
Binding constants for A1G12 protein fragment for varying lengths of ds- and ssDNA, as determined from **Figure 6.4**.

These FP experiments were repeated using the B2C2 protein fragment, which showed the affinity of the B2C2 protein fragment for DNA was generally weaker than for A1G12, and the K_ds were not as consistent as for ds- or ssDNA binding. Furthermore, the same reduction in DNA binding affinity was shown for both ds- and ssDNA less than 15 bp / nt in length (**Figures 6.5A** and **6.5B**), as was found with A1G12. The calculated binding constants for B2C2 binding to varying lengths of ds- and ssDNA are presented in **Table 6.3**.

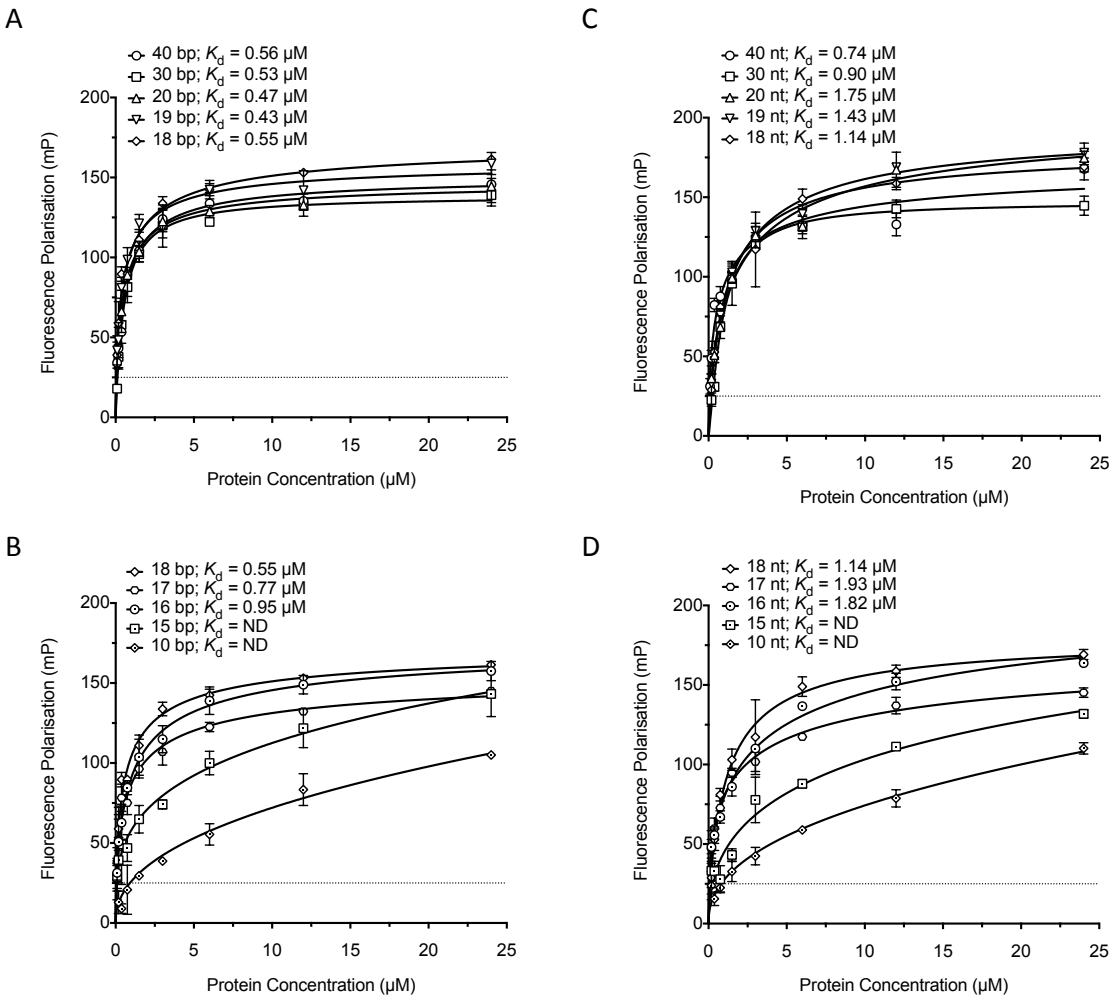


Figure 6.4 Protein fragment A1G12 binds to both ds- and ssDNA.
Binding of A1G12 to 5'-FITC ds- and ssDNA of varying lengths was tested by FP. In each case, the DNA was at a final concentration of 100 nM. Each data point is the mean of three individual experiments and each error bar represents one standard deviation. (**A**) A1G12 protein with 40, 30, 20, 19 and 18 bp dsDNA. (**B**) A1G12 protein with 18, 17, 16, 15 and 10 bp dsDNA. (**C**) A1G12 protein with 40, 30, 20, 19 and 18 nt ssDNA. (**D**) A1G12 protein with 18, 17, 16, 15 and 10 nt ssDNA. Binding constants (K_d) for each experiment are shown, where they could be calculated. ND=K_d not determined.

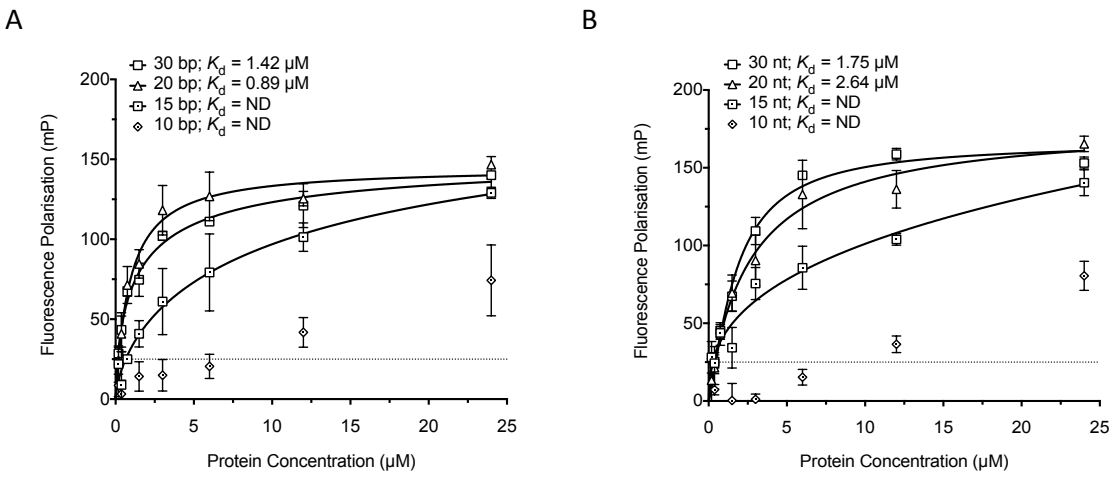


Figure 6.5 Protein fragment B2C2 binds to both ds- and ssDNA.
Binding of B2C2 to 5'-FITC ds- and ssDNA of varying lengths was tested by FP. In each case, the DNA was at a final concentration of 100 nM. Each data point is the mean of three individual experiments and each error bar represents one standard deviation. (**A**) B2C2 protein with 30, 20, 15 and 10 bp dsDNA. (**B**) B2C2 protein with 30, 20, 15 and 10 nt ssDNA. Binding constants (K_d) for each experiment are shown, where they could be calculated. ND=K_d not determined.

DNA length (bp / nt)	K_d (μM)	
	dsDNA	ssDNA
30	1.42	1.75
20	0.89	2.64
15	-	-
10	-	-

Table 6.3 Binding constants for B2C2 protein with DNA.
Binding constants for A1G12 protein fragment for varying lengths of ds- and ssDNA, as determined from **Figure 6.5**.

This set of binding data, indicates a slight preference for the binding of N-terminal Claspin fragments to dsDNA over ssDNA, which had previously been described (Sar et al., 2004, Zhao and Russell, 2004, Serçin and Kemp, 2011, Uno and Masai, 2011 and Yilmaz et al., 2011). Furthermore, there is a minimum length requirement for this interaction, namely DNA greater than 16 bp / nt in length, which accounts for roughly 1.5 turns of dsDNA.

6.4 Complex formation of Claspin with dsDNA

Previous experiments showed that selected N-terminal Claspin fragments bound preferentially to dsDNA, however the 'stability' of such complexes was not examined. ASEC was therefore used to investigate the monodispersity of each of the complexes formed by the N-terminal fragments with dsDNA. Initially complex formation of constructs A1G12 and B2C2 with various lengths of dsDNA was investigated. Samples of A1G12 and B2C2 protein at a concentration of 5 mg/ml were incubated with a 1:1.2 molar excess of dsDNA (10, 15, 20 and 30 bp). The respective chromatographs were overlaid, and selected fractions from across each elution peak were analysed by both SDS- and Native-PAGE gels.

These analyses indicated that the A1G12 protein did not interact with the 10 bp dsDNA, however some indication of complex formation was seen at 15 bp dsDNA, but robust interaction was only seen when the dsDNA was longer than 20 bp. On the chromatographs there was a shift in the elution volume of the peak with longer lengths of dsDNA indicating an increase in the molecular mass of the protein (**Figure 6.6A**). DNA despite having an absorbance maxima at 260nm, still absorbs

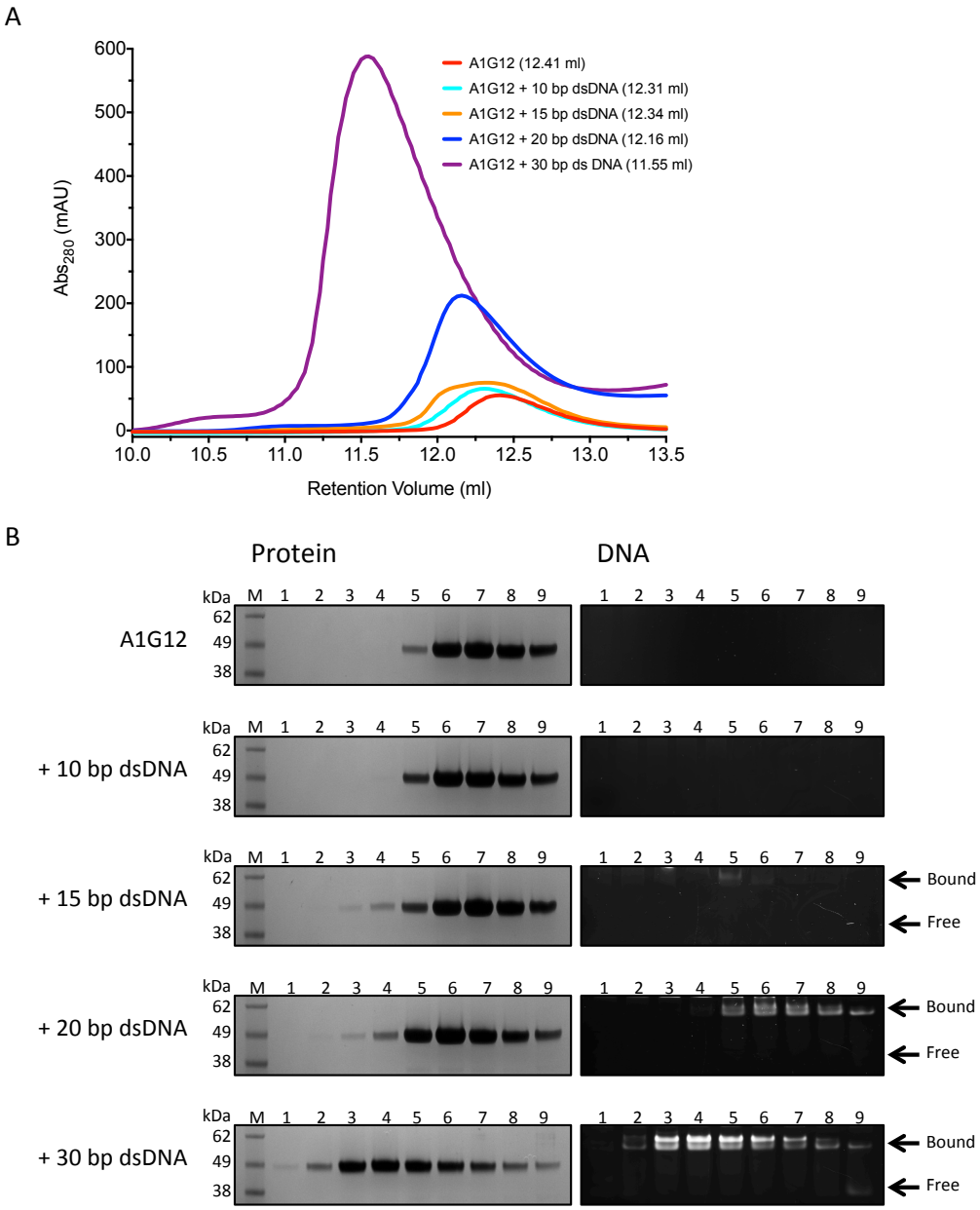


Figure 6.6 Analysis of A1G12 - dsDNA complexes by ASEC.
(**A**) Representative chromatograms for the A1G12 fragment incubated with increasing lengths of dsDNA. A1G12 at a final concentration of 5 mg/ml was either applied directly to the column [10/300 SD200 Increase], or incubated with increasing lengths of dsDNA at a 1:1.2 ratio, for 30 minutes on ice, prior to sample injection. The identity of each of the A1G12 / dsDNA complexes is indicated in the top right hand corner of the trace, and the elution volume is shown in brackets. (**B**) Analysis of selected fractions by SDS-PAGE (Left; Instant Blue-stained 4-12% Bis-Tris SDS-PAGE gel) and Native-PAGE (Right; EtBr-stained 5% acrylamide Native-PAGE gel). The migration position of bound and free DNA are indicated by arrows. M=molecular mass marker, 1 to 9=selected consecutive elution fractions.

appreciably at 280nm; and thus will contribute to the overall size of each peak on the chromatograph; resulting in an increase in the peak height as seen. The shift in the elution peak is visible by SDS-PAGE when the dsDNA was longer than 20 bp (**Figure 6.6B**). Native-PAGE showed co-elution of dsDNA with protein when the dsDNA was longer than 20 bp as indicated by the shifted "bound" DNA band (**Figure 6.6C**); these fractions correlate with the protein containing fractions from SDS-PAGE. The dsDNA length required for a stable interaction with A1G12 protein corroborates with the dsDNA length requirements shown by FP (section **6.3.1**).

For B2C2, these analyses indicated that the B2C2 protein did not interact with the 10 or 15 bp dsDNA, there was indication of complex formation seen with 20 bp dsDNA, but a robust interaction was only seen for 30 bp dsDNA. On the chromatograms there was a shift in the elution volume of the peak for the longest length of dsDNA (**Figure 6.7A**). The shift in the elution peak is visible by SDS-PAGE for the 30 bp dsDNA incubated protein sample (**Figure 6.7B**). Native-PAGE showed the slight presence of co-elution of dsDNA with protein for 20 bp dsDNA, but a strong dsDNA and protein co-elution was identified for the 30 bp dsDNA as indicated by the shifted "bound" DNA band (**Figure 6.7C**); these fractions correlate with the protein containing fractions from SDS-PAGE. The dsDNA length required for a stable interaction with B2C2 protein corroborate with the DNA length requirements shown by FP (section **6.3.1**), although a more stable interaction with 20 bp dsDNA would have been expected by ASEC based on the FP data.

6.4.1 Characterization of weaker Claspin complex formation with dsDNA

ASEC was also carried out for five additional protein fragments; two which had weaker DNA binding activity (A1D6 and B2D9) and three that did not show any DNA binding activity (B2C4, B1F8 and B2D6) as previously determined by FP (section **6.3**). The experiment was carried out as before, but only using the 20 bp dsDNA duplex. The protein fragment A1G6D was not analysed due to the presence of a shoulder on the peak visualised on the ASEC chromatograph indicating this was not monodisperse in solution (see **Figure 5.4**). Incubation of A1D6 and B2D9 with the 20 bp dsDNA produced a small shift in the elution volume and a concomitant increase in the peak height (due to absorbance from the DNA)

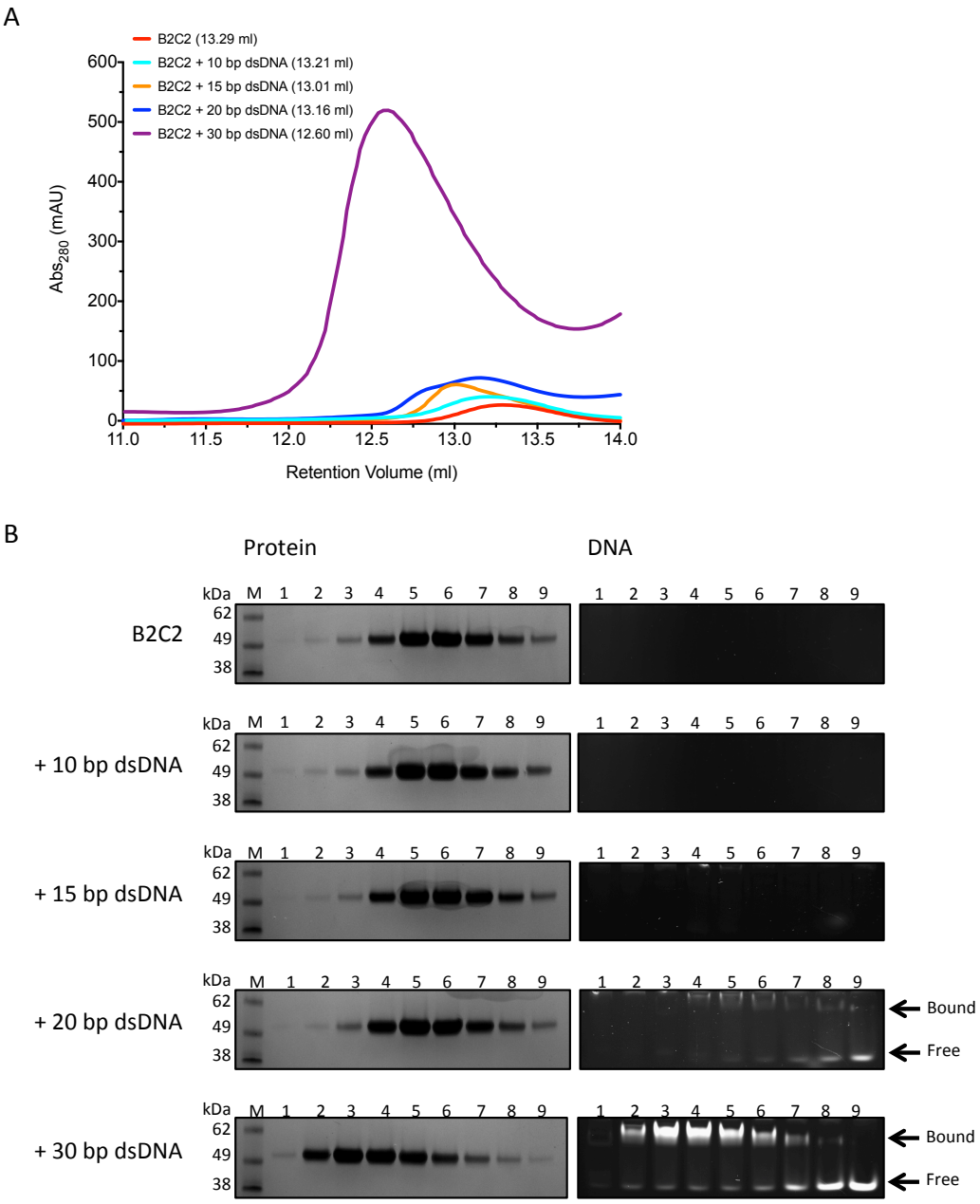


Figure 6.7 Analysis of B2C2 - dsDNA complexes by ASEC.
(**A**) Represenative chromatograms for the B2C2 fragment incubated with increasing lengths of dsDNA. B2C2 at a final concentration of 5 mg/ml was either applied directly to the column (10/300 SD200 Increase), or incubated with increasing lengths of dsDNA at a 1:1.2 ratio, for 30 minutes on ice, prior to sample injection. The identity of each of the B2C2 / dsDNA complexes is indicated in the top right hand corner of the trace, and the elution volume is shown in brackets. (**B**) Analysis of selected fractions by SDS-PAGE (Left; Instant Blue-stained 4-12% Bis-Tris SDS-PAGE gel) and Native-PAGE (Right; EtBr-stained 5% acrylamide Native-PAGE gel). The migration position of bound and free DNA are indicated by arrows. M=molecular mass marker, 1 to 9=selected consecutive elution fractions.

(**Figure 6.8A**), where B2D9 showed a more robust interaction with DNA than A1D6, as was previously indicated by the FP experiments (section **6.3**). The fragments that did not show an interaction by FP, also showed no shift in elution volume or increase in the peak height, confirming that there was no interaction between these fragments and the DNA tested (**Figure 6.8B**).

6.5 The degradation product of B2D9 loses DNA binding capability.

During a purification of B2D9 (as described in section **5.3**) (aa 261-534), two separate species (peaks) were found to elute from the Heparin column (**Figure 6.9A**). SDS-PAGE analysis of fractions from both elution peaks, found that the first peak contained a slightly smaller protein than expected for the full-length B2D9 protein, which was found to elute in the second peak (**Figure 6.9B**). Previous purifications of B2D9 using Heparin chromatography found the smaller species usually eluted in the void 'unbound' fraction from the Heparin column (**Appendix 2.9C**). The two Heparin elution peaks were pooled separately, and then applied to a SEC column individually. Each species eluted from the SEC column as a single peak and were found to be stable, as determined by SDS-PAGE (**Figure 6.9C**). The smaller protein was identified as a breakdown product of B2D9 by Edman degradation (N-terminal sequencing), which begins at amino acid 317, revealing loss of the 56 residues encoding the predicted HTH motif (Zhao and Russell, 2004) (**Figure 6.9D**); this is hereon in referred to as B2D9D (aa 317-534). Previously, DNA binding capability has now been demonstrated for each of the Claspin fragments, which contain a predicted HTH motif (aa 279-313) (Zhao and Russell, 2004) and BP1 (aa 265-331) (Lee et al., 2005). This fortuitous breakdown product enabled testing of the DNA-binding capability of this motif.

Firstly, using FP, any differences in the DNA binding capability of the B2D9D protein compared to B2D9 were examined, using the same 20 nt / bp ss- and dsDNA templates as previous (section **6.3**). Here, no robust increase of FP signal was seen for the B2D9D protein fragment, indicating that it was no longer able to bind to DNA (**Figure 6.9E**). Secondly, ASEC was used to investigate any interaction of B2D9D with a 20 bp dsDNA (section **6.4**). Chromatographs indicated no change

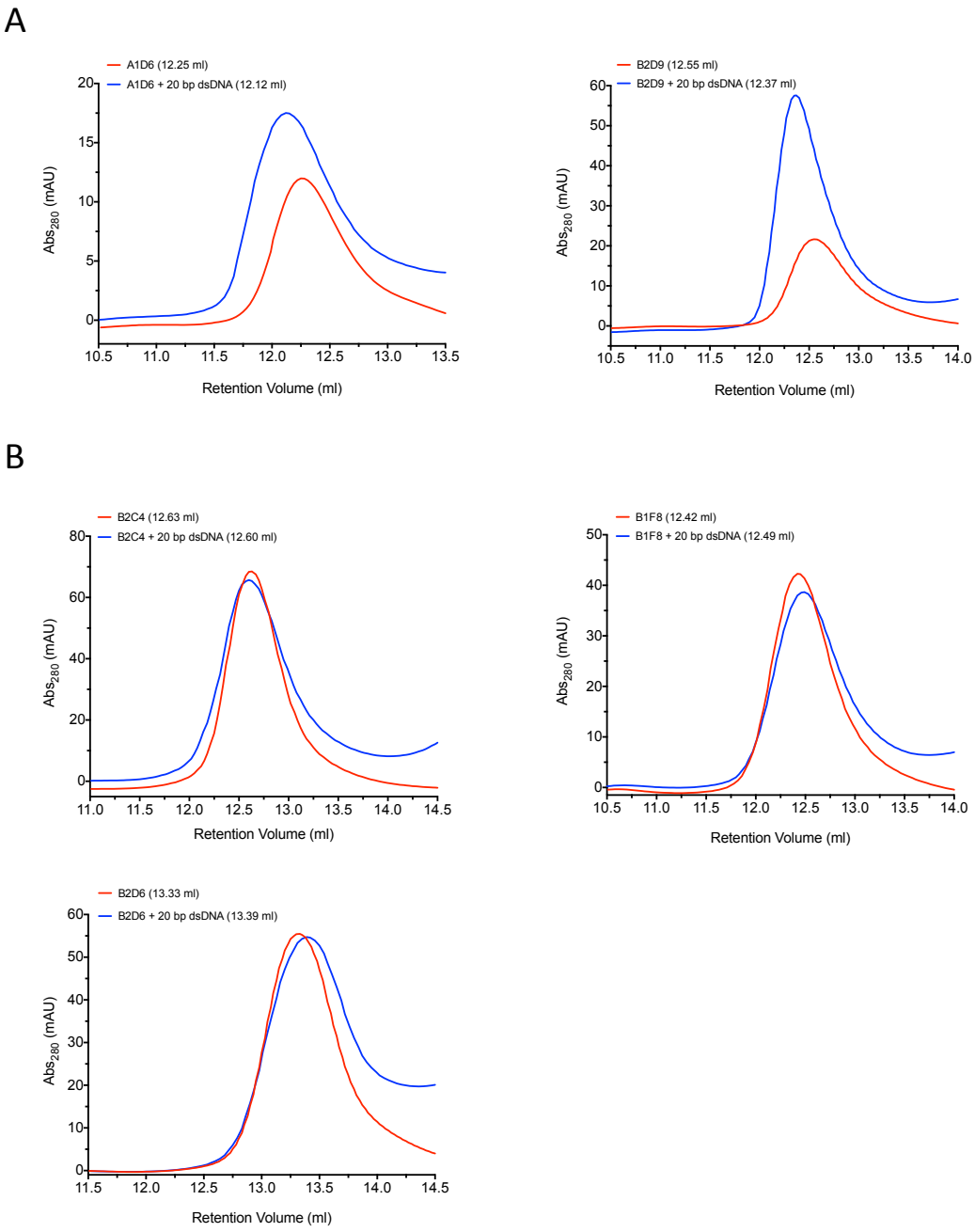


Figure 6.8 ASEC of weakly or non-interacting Claspin fragments with dsDNA. Represenative chromatograms for the Claspin protein fragments incubated with increasing lengths of dsDNA. Claspin protein fragments at a final concentration of 5 mg/ml were either applied directly to the column (10/300 SD200 Increase), or incubated with 20 bp dsDNA at a 1:1.2 ratio, for 30 minutes on ice, prior to sample injection. Red lines represents the protein fragment elution profile, while the blue lines represent the protein fragment incubated with 20 bp dsDNA. The identity of each of the protein fragments is indicated in the top right hand corner of the trace, and the elution volume is shown in brackets. (**A**) Chromatographs for weak DNA interacting protein fragments A1D6 and B2D9 (**B**) Chromatographs for DNA non-interacting protein fragments B2C4, B1F8 and B2D6.

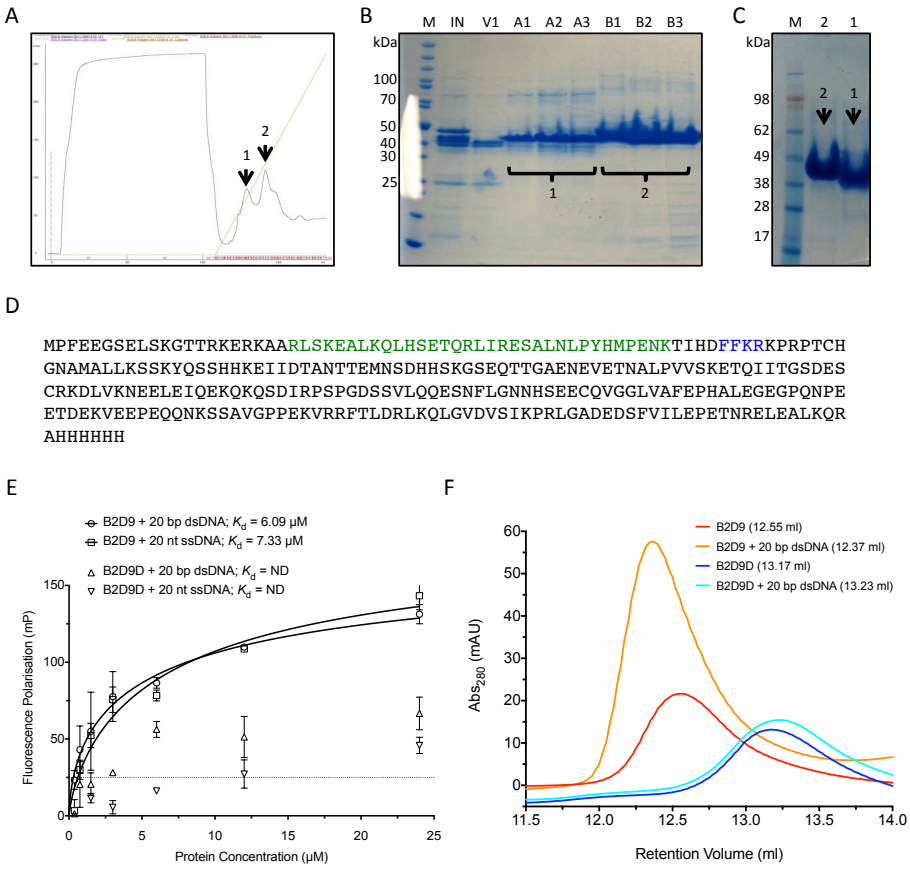


Figure 6.9 The predicted HTH motif sequence is required for DNA binding. (**A**) Two different species were found to elute from a Heparin column, marked '1' and '2' on the representative chromatograph which correspond to B2D9D and B2D9 respectively [showing absorbance at 280 nm (blue line) and the applied NaCl gradient (green line)]. (**B**) SDS-PAGE analysis of fractions collected across peak 1 and 2 as indicated. 4-20% Tricine SDS-PAGE gel, stained with Instant Blue. M=molecular mass marker, IN=input, V=fractions from the void volume, A1 to A3=fractions of peak 1, B1 to B3=fractions of peak 2. (**C**) SDS-PAGE analysis of purified B2D9 ('2') and B2D9D ('1'). 4-20% Tricine SDS-PAGE gel, stained with Instant Blue. (**D**) Edman degradation indicates that B2D9D has lost the predicted HTH motif (Zhao et al., 2004) (coloured green) and starts at 'FFKR' amino acid 317 (coloured blue). (**E**) Binding of B2D9 and B2D9D to 5'-FITC 20 bp / nt ds- and ssDNA was tested by FP. In each case, the indicated DNA was at a final concentration of 100 nM. Each data point is the mean of three individual experiments and each error bar represents one standard deviation. Binding constants (K_d) for each experiment are shown, where they could be calculated. ND=K_d not determined. (**F**) DNA interaction analysis by ASEC. B2D9 / B2D9D at a final concentration of 5 mg/ml was either applied directly to the column (10/300 SD200 Increase), or incubated with increasing lengths of dsDNA at a 1:1.2 ratio, for 30 minutes on ice, prior to sample injection. The identity of each of the B2D9 / B2D9D with dsDNA complexes is indicated in the top right hand corner of the trace, and the elution volume is shown in brackets. In both assays formats, B2D9 interacts with a 20bp dsDNA, whereas B2D9D does not.

in either peak height or elution volume when the protein was pre-incubated with the dsDNA template, which is in contrast to that previously observed for the B2D9 protein fragment (**Figure 6.9F**). Collectively, these analyses indicated that the HTH motif and some surrounding amino acids sequence are required for the DNA binding capability of human Claspin.

6.6 Additional experiments to characterise the complexes between Claspin protein fragments and DNA

In addition to the experiments described above, both limited proteolysis and crystallisation trials were carried out using the DNA binding Claspin protein fragments. However, as these were unsuccessful, only a brief summary is given below.

6.6.1 Limited proteolysis of protein:DNA complexes

Limited proteolysis experiments using trypsin, were carried out as described in section **5.14**, but using purified A1G12 and B2C2 in complex with ss- or dsDNA. Purified protein fragments were prepared at a final concentration of 10 μM and then mixed at a ratio of 1:1.2 (protein:DNA) with either a 20 nt / bp ss- or dsDNA, and incubated with a trypsin dilution series. For both of the protein fragments tested, the 'full-length' protein was somewhat stabilised against trypsin digestion in the presence of DNA, but there was no significant difference in the pattern of breakdown products when compared to the unliganded protein (**Figure 6.10**).

6.6.2 Crystallography trials with Claspin protein fragments in complex with DNA

In addition to the experiments described above crystallisation trials were carried out. Fragments A1G12 and B2C2, at 5, 10 and 15 mg/ml were incubated with either 20 bp dsDNA or 20 bp dsDNA + 2 nt overhang DNA template at either a 2:1 or 5:1 DNA:protein ratio. Typically these were mixed at a high concentration, and then incubated for 20 minutes on ice. These were set up in sitting drop format in plates, as a 0.2 μl protein with 0.2 μl mother liquor drops, against a well volume of 40 μl, using an ARI Crystal Phoenix Liquid Handling System. A wide range of different commercial crystallisation screens were used (Natrix, PEG/ION, JCSG-

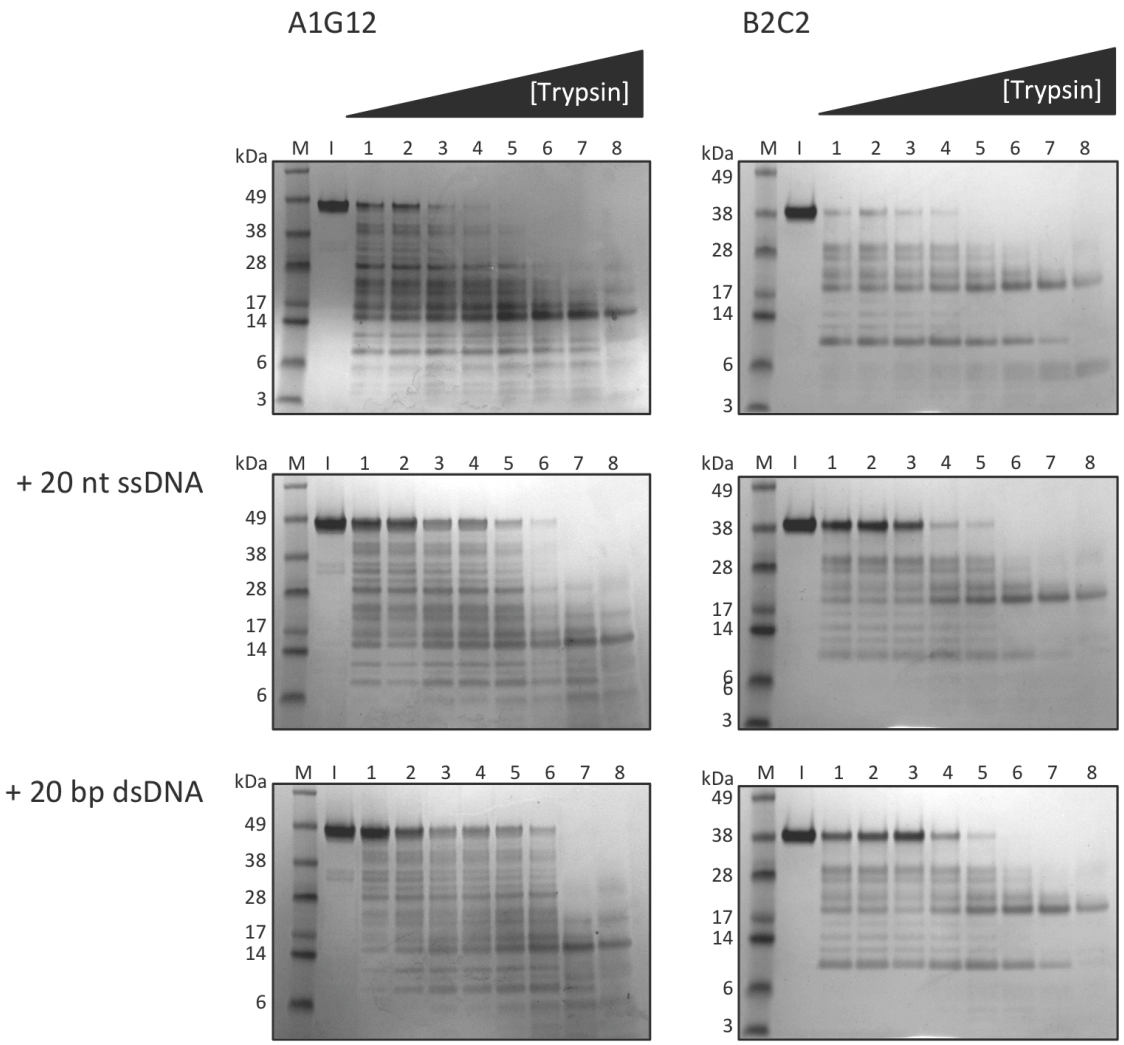


Figure 6.10 Limited proteolysis of protein:DNA complexes.
SDS-PAGE analysis of limited proteolysis reactions. Purified A1G12 (right) and B2C2 (left) at a final concentration of 10 μM was incubated for 15 minutes with trypsin (at varying concentrations), either alone (top) on in complex with 20 nt ssDNA (middle) or 20 bp dsDNA (bottom), for 20 minutes at room temperature. M=molecular mass marker, I=input sample (no added trypsin), 1 to 8=trypsin protease concentration: 1=0.01 nM, 2=0.14 nM, 3=1.67 nM, 4=2 nM, 5=2.5 nM, 6=3.33 nM, 7=5 nM, 8=10 nM. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

plus, ProPlex and PACT premier), and duplicate screens were also incubated at temperatures of 4 and 14 °C. However, as with the crystallography trials of the Claspin protein fragments (section **5.11**), no protein crystals were identified and typically >50% of the conditions tested would produce phase separation. The conditions that resulted in phase separation were not explored or further optimised.

6.7 Summary

The objective of the experimental work reported in this chapter, was to further characterise the N-terminal region of human Claspin, in terms of its ability to bind to DNA, using the eight N-terminal soluble fragments identified from the initial CDH screen (from **Chapters 4** and **5**). A summary of the DNA binding assay results can be found in **Table 6.4**.

From these cumulative results, two protein fragments (A1G12 and B2C2) were found to possess strong DNA-binding capability, whilst one protein fragment (B2D9) possessed a slightly weaker interaction and a very weak interaction was identified in a further protein fragment (A1D6); with increasing K_d in the micromolar range for DNA binding in the order A1G12 < B2C2 < B2D9. ASEC of the protein:DNA complexes made by A1G12 and B2C2 with dsDNA, showed a single monodisperese, protein-DNA species could be visualised, and the complexes were verified by SDS- and Native-PAGE analyses. Two further Claspin fragments (A1D6 and B2D9) showed a slight interaction with dsDNA by ASEC. The stoichiometry of these complexes could not be verified as both the Claspin protein fragments and DNA are elongated molecules and resolve aberrantly on SEC. Previous reports on the stoichiometry of full-length Claspin with DNA by Sar et al., (2004) using Electron Microscopy (EM), identified a 1:1 ratio of protein:DNA (Sar et al., 2004). These experiments identified for A1G12 and B2C2 a minimal length of DNA templates required for an interaction of 16 nt ssDNA or 16 bp dsDNA were required for a robust interaction, with further increases in DNA length not affecting the calculated binding constants.

Protein Fragment	Method				
	EMSA		FP		ASEC
	dsDNA	ssDNA	dsDNA	ssDNA	dsDNA
B2C4	Y	W	N	N	N
A1G12	Y	W	Y	Y	Y
A1D6	Y	W	W	W	W
B2C2	Y	W	Y	Y	Y
B2D9	W	N	Y	Y	W
B1F8	N	N	N	N	N
A1G6D	N	N	N	N	-
B2D6	N	N	N	N	N

Table 6.4 Claspin protein fragments DNA binding summary.
Assessment of Claspin protein fragments DNA binding functionality from **Chapter 6**. Y=strong interaction, W=weak interaction, N=no interaction, -=not tested.

Three protein fragments (B1F8, A1G6D and B2D6) did not bind showing DNA binding propensity by these methods. B2C4 could be shown to interact with DNA by EMSA, but not by any further method. It is worth noting that for B2C4, the FP signal was consistently and strongly negative at the low protein concentrations for both ds- and ssDNA (see **Figure 6.2** and **Figure 6.3**). This indicates the fluorophore-DNA in the B2C4 protein sample is tumbling faster than in the fluorophore-DNA in the background sample. Whilst this remains an anomaly, this could indicate there is interference from B2C4 protein to the fluorophore-DNA molecule, but there was no further evidence from EMSAs or ASEC that could explain this phenomenon. The identification of a N-terminal degradation product of B2D9, termed B2D9D suggested that the sequence spanning the predicted HTH motif (Zhao and Russell, 2004) was necessary for DNA binding activity, as determined by FP and ASEC. This also corroborates with data for the protein fragment B1F8, which contains a small part of the HTH motif, and does not show DNA binding functionality.

Further corroboration of these DNA binding results, was during the purification of the protein fragments. The second purification step for all eight Claspin protein fragments was application to a Heparin column (section **5.3**). This resin is somewhat unusual as it acts as both an affinity matrix, as well as a weak cation exchange resin. The 'affinity' arises from the fact that the Heparin functional group

mimics the sugar moiety of the deoxyribose chain. One observation is that each of the fragments A1G6D, B1F8 and B2D6 (found to not bind DNA), required a dilution step (to reduce the overall concentration of sodium chloride), before they were retained on the Heparin column. Furthermore, these all eluted at significantly lower sodium chloride concentrations, compared to the other 'DNA-binding' fragments.

Limited proteolysis experiments using trypsin, identified for both of the protein fragments tested, the 'full-length' protein was somewhat stabilised against trypsin digestion in the presence of DNA, but there was no significant difference in the pattern of breakdown products when compared to the unliganded protein. However, tryptic digests with DNA were not carried out with a protein fragment that did not interact with DNA, therefore the effect of DNA binding on overall protein stability, as measured by sensitivity to trypsin, could not be determined.

Crystallisation of different protein fragments in complex with DNA was attempted, but unfortunately, despite exhaustive screening efforts, crystals of a Claspin protein fragment in complex with dsDNA were not obtained. In **Chapter 5**, I found the protein fragments had a high degree of structural disorder, it is not known if DNA binding results in the structural ordering of the protein fragments or part of the protein fragments, and in this thesis, this was not determined.

Whilst DNA binding activity has been reported in the scientific literature for a number of different fragments of human Claspin, to date there has been no consensus as to how much of the N-terminal region of Claspin is required to support DNA-binding (Lee et al., 2005, Sar et al., 2004, Zhao and Russell, 2004, Serçin and Kemp, 2011, Uno and Masai, 2011 and Yilmaz et al., 2011). The two soluble N-terminal fragments of Claspin, A1G12 and B2C2, identified by CDH, both interact tightly with DNA, and in fact contain a significant region of sequence overlap (aa 218-413). Significantly, this region of overlap contains the previously predicted HTH motif (aa 279-313) (Zhao and Russell, 2004) and BP1 (aa 265-331) (Lee et al., 2005). Furthermore, the fragment A1G12 also contains the entire region of previously experimentally identified DNA-binding domain (DBD; aa 149-340)

(Sar et al., 2004), whilst B2C2 only contains part of this. A1G12 has consistently shown a tighter interaction with DNA throughout this work indicating a region of sequence found within A1G12 but not B2C2 is required for enhancing DNA binding. The protein fragments A1G6D, B1F8, B2C4 and B2D6, found not to interact with DNA, do not contain the full amino acid sequence of the HTH motif or BP1, again re-enforcing the idea that both these regions are required for DNA-binding. Interestingly, the protein fragments A1D6 and B2D9 have only weak DNA-binding, despite containing significant sequence overlap with A1G12 and B2C2, (containing both the HTH motif and the DBD), which suggests that the additional (C-terminal amino acids) found in these constructs (but not A1G12 or B2C2) may in fact be inhibitory to DNA-binding. These amino acids could potentially fold back over the DNA-binding region, occluding the DNA binding functionality of the protein and thereby mediate the interaction of Claspin with DNA. However, to determine if this or if something else was restricting DNA binding activity, the tertiary structure of the protein fragments would need to be determined.

DNA binding proteins can interact with DNA in a number of different modes; generally sequence- or structure-specific. Claspin does not appear to have any sequence specificity and therefore is likely to interact with DNA through simple recognition of the phosphodiester backbone, explaining the general lack of discrimination between both ss- and dsDNA templates. This does not however, exclude the possibility that other parts of the Claspin protein have differing modes of DNA-binding activity; this would require further characterization through structural studies.

For further work on the Claspin protein fragment-DNA complex, the stoichiometry of the interaction and/or structural studies could be carried out using ITC, AUC, NMR, or bioSAXS. Preliminary ITC experiments were carried out (data not shown), but this generated very large heats of dilution on injection of DNA templates into the sample chamber, which would not be suitable for an ITC experiment, potentially through buffer incompatibilities in the samples. No further attempts were made at ITC for the protein:DNA interaction. AUC experiments for three protein fragments determined that they were all monomeric at the protein

concentrations tested (section **5.9**). However, due to restricted access to the ultracentrifuge, experiments on the protein:DNA complexes were not carried out. These types of experiments could have confirmed if a stable protein-DNA interaction formed in solution by the identification of a single sedimenting species that sedimented faster than protein alone. A change in the sedimentation profile for protein:DNA in comparison to free protein would enable the calculation of the molecular mass of the complex and subsequently the stoichiometry of the complex. Alternatively, bioSAXS experiments would allow the calculation of complex stoichiometry and an investigation of any structural changes in the protein fragments upon binding to DNA. The original bioSAXS data for the protein fragments was not of sufficient quality for extensive data analysis, which was likely due to the storage conditions (section **5.12**). For investigation of the protein fragments with DNA, this would require further optimisation initially for the optimal bioSAXS data collection for the protein fragments, prior to analysis with DNA, due to time and access restraints this was not possible. Another alternative methodology could be to use NMR titration experiments, with ^{1}H-^{15}N HSQC analysis – if DNA binding to the protein fragment results in a structural alteration of the protein, this would result in peak shifts on a ^{1}H-^{15}N HSQC spectra. However, the production of sufficient isotope labelled recombinant protein to facilitate this type of approach was difficult, due to the relatively poor levels of recombinant protein expression (section **5.13**), thereby preventing the study of the protein:DNA complex by NMR. Furthermore, as previously described for the Claspin protein fragments alone, the study of the structure of the protein-DNA complex by EM would not a viable technique, due to the relatively small overall size of the protein:DNA complexes.

Chapter 7

Characterisation of the human Chk1-Claspin interaction

7.1 Introduction

Claspin is proposed to interact with a number of proteins that are essential for both DNA replication and DNA replication-coupled repair, however, only a few sites of interaction have been accurately located on Claspin, one of which is Chk1 (Kumagai and Dunphy, 2003, Clarke and Clarke, 2005, Chini and Chen, 2006). In response to replication stress and DNA damage, Claspin functions as a so-called 'mediator' and is required for the activation of Chk1 by ATR, and subsequent cell cycle arrest (Kumagai and Dunphy, 2000, Chini and Chen, 2003, Kumagai and Dunphy, 2003, Lin et al., 2004, Clarke and Clarke, 2005, Lindsey-Boltz et al., 2009) (refer to section **1.10.3**). The putative CKBD in human Claspin (aa 909-985), is a highly conserved segment, containing three repeats of a 10 amino acid CKB motif comprised of ExxxLC(S/T)GxF (motif 1: aa 910-919, motif 2: 939-948 and motif 3: 976-985) – where phosphorylation on the central Ser / Thr is essential for interaction by Chk1 (Kumagai and Dunphy, 2003, Clarke and Clarke, 2005, Chini and Chen, 2006) (refer to section **1.10.4**). The protein kinase CK1γ1 has been found to be responsible for phosphorylating the CKBD (Meng et al., 2011). Chk1 can interact with the phosphorylated CKBD through its N-terminal kinase domain, where a positively charged pocket on the surface of Chk1 formed from Lys54, Arg129, Thr153 and Arg162, is thought to bind the phosphorylated CKB motif (Chen et al., 2000, Jeong et al., 2003). The interaction of Chk1 with Claspin brings it into the proximity of ATR, which then phosphorylates Chk1 on Ser317 and Ser345 (Lopez-Girona et al., 2001, Zhao and Piwnica-Worms, 2001, Capasso et al., 2002, Gatei et al., 2003, Katsuragi and Sagata, 2004); the activation of Chk1 is described in section **1.9**. Activated Chk1 has been shown to have a lower binding affinity for Claspin, resulting in its disassociation and dispersal throughout the nucleus (Jeong et al., 2003, Smits et al., 2006).

Whilst the interaction site of Chk1 on Claspin has been identified, the structural underpinning of this interaction has not yet been determined. Here, experiments to probe the interaction between Chk1 and phosphorylated CKB motifs were undertaken, prior to attempts to co-crystallise a suitable synthetic peptide representing the Claspin-CKB with the kinase domain of human Chk1.

7.2 Recombinant protein expression of human Chk1 in *Sf*9 cells

7.2.1 Baculovirus expression constructs

Chk1-KD has been successfully expressed and purified from *Sf*9 insect cell culture, as a C-terminal His_6 affinity tagged fusion (Chen et al., 2000, Zhao et al., 2002, Foloppe et al., 2005, Vanderpool et al., 2009, Oza et al., 2010) – it was previously reported that there were difficulties in expressing the recombinant protein in *E. coli* (Chen et al., 2000, Arlander et al., 2006). The structure of the human Chk1 kinase domain (Chk1-KD) has previously been determined using recombinant protein (aa 1-289) expressed in the baculovirus / insect cell system. Here, for recombinant protein expression in *Spodoptera frugiperda* cells (cell-line *Sf*9), three different expression constructs were made: His-Chk1 (full-length Chk1 construct), Chk1-KD^{1-270}-His, and Chk1-KD^{1-289}-His. Note, by their terminology, that the first construct has an N-terminal His_6 affinity tag, whereas the latter two are tagged at the C-terminus. These expression constructs were first generated and confirmed by either colony PCR or restriction digests (**Appendix 3.1A**). These plasmids were used to transform the *E. coli* DH10MultiBac cell strain, in order to transpose each constructs DNA into a bacmid suitable for transfection into insect cells; transposition was confirmed by PCR (**Appendix 3.1B**).

7.2.2 Transfection

*Sf*9 insect cells were transfected with the transposed bacmids in order to generate recombinant baculovirus suitable for infection of *Sf*9 cells for large-scale expression cultures. Protein expression levels during viral amplification stages were examined by western blot (anti-His) (**Appendix 3.2A**). Protein doublets were observed in some cases, potentially due to some level of N-terminal degradation (C-terminal tagged constructs). Recombinant baculovirus was amplified by the infection of increasing volumes of *Sf9* cells, and the resulting viral titre (plaque forming units; PFU per ml) was determined by plaque assay to calculate the appropriate volume of baculovirus needed to uniformly infect *Sf*9 cells in order to produce recombinant protein.

7.2.3 Recombinant protein expression

A small-scale expression test (0.5 litre) was carried out for the baculovirus expressing Chk1-KD$^{1-270}$-His. Samples of infected *Sf*9 cells were taken at 24, 48 and 62 hour time points, post-infection, and subsequently analysed by western blot (**Appendix 3.2B**). Expression of recombinant protein was readily detected and showed increased levels of protein expression with time. Furthermore, the protein doublet visible at the 48 hour time point (and also seen during the viral amplification; indicative of N-terminal degradation), was not present at the 62 hour time point. For large-scale expression cultures, 4-8 litres of *Sf*9 cells were infected with recombinant baculovirus. Infected cells were harvested by centrifugation around 62 hours post infection or once the viability of cells had dropped to between 92 and 95% (seemingly optimal for high yields of recombinant Chk1-KD protein), these were frozen in liquid nitrogen prior to storage at -80 °C.

7.3 Purification of Chk1

Recombinant Chk1 protein was typically purified from the soluble fraction resulting from 2 litres of *Sf*9 cell culture, and was purified in a two-step purification protocol consisting of: Talon-IMAC and SEC (section **2.15.3**). Selected fractions from these purification steps were analysed by SDS-PAGE.

The initial IMAC step achieved substantial purification for each of the recombinant Chk1 proteins showing a modest level of expression, but some contaminating species remained in all cases; Chk1-KD$^{1-270}$-His (**Figure 7.1A**), His-Chk1 (**Appendix 3.3A**) and Chk1-KD$^{1-289}$-His (**Appendix 3.4A**). Each of the proteins was applied to an SEC column and eluted within the column resolving volume. SDS-PAGE analysis showed that the purity of His-Chk1 was not significantly improved, as shown by the number of other bands, and also indicated by the multiple peaks present on the chromatograph (**Appendix 3.3B**). However, the purity of the two constructs expressing the Chk1-KD domain had been significantly improved, as shown by the single migrating species by SDS-PAGE; Chk1-KD$^{1-270}$-His (**Figure 7.1B**) and Chk1-KD$^{1-289}$-His (**Appendix 3.4B**). Fractions containing the recombinant protein were pooled and then concentrated. Re-analysis of the

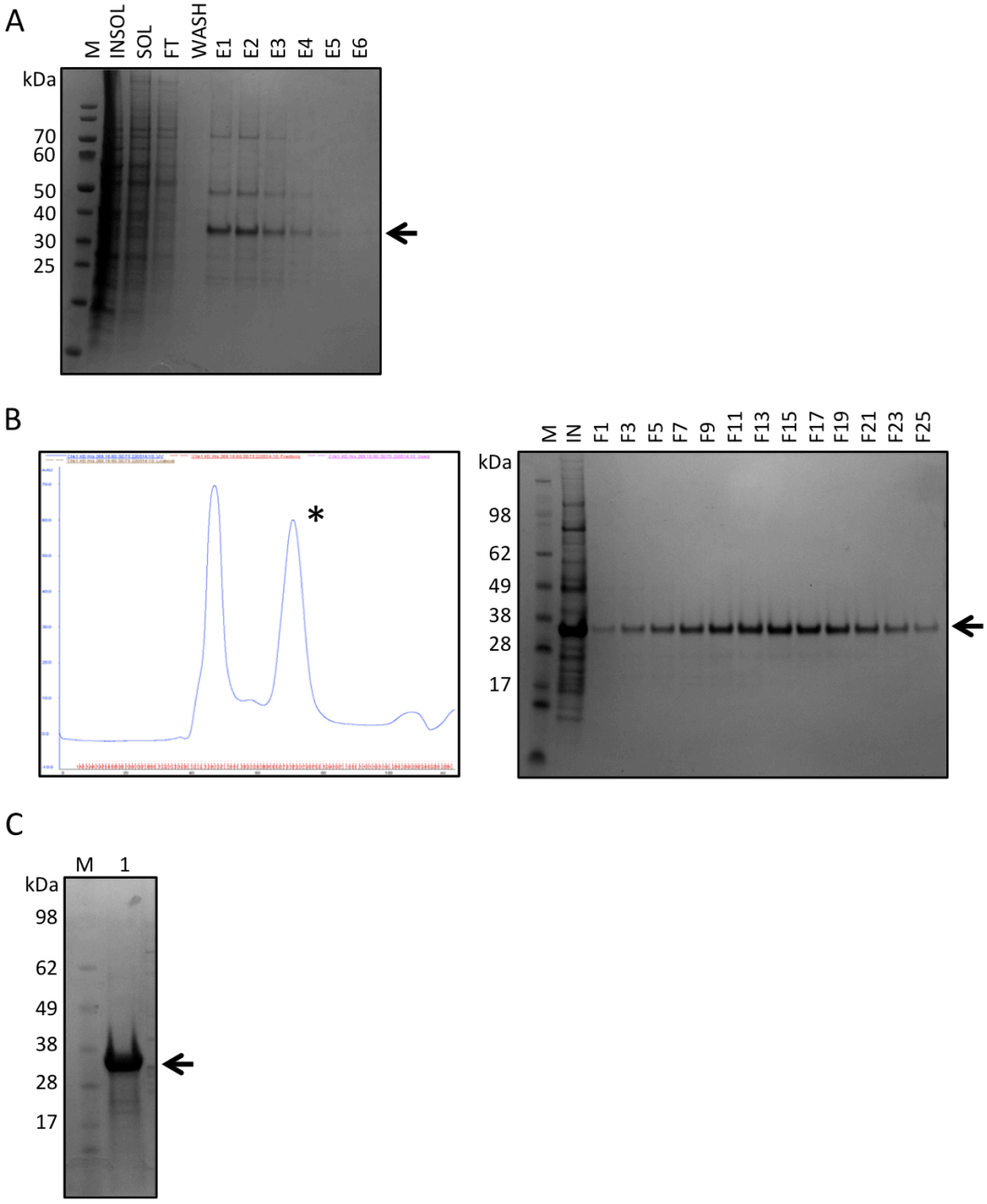


Figure 7.1 Purification of Chk1-KD$^{1-270}$-His: IMAC and SEC steps.
(**A**) SDS-PAGE analysis of Sf9 cell lysate, and IMAC purification steps. M=molecular mass marker, INSOL= insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**B**) SEC using a 16/60 SD75 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, F=indicated fraction. (**C**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

concentrated protein samples again by SDS-PAGE, showed that His-Chk1 protein was highly contaminated with co-purifying species of lower molecular mass (**Appendix 3.3C**). Again, a single species represented both Chk1-KD$^{1-270}$-His (**Figure 7.1C**) and Chk1-KD$^{1-289}$-His (**Appendix 3.4C**), but with some relatively minor level of impurities. The Chk1$^{1-270}$-His and Chk1-KD$^{1-289}$-His constructs were both readily purified and used for further study, whilst the His-Chk1 protein was not amenable to further study and was not taken further. Typically though, only relatively low yields of both Chk1-KD proteins (<0.5 mg of purified Chk1-KD per litre of *Sf*9 cell culture) were obtained; a phenomenon previously reported and attributed to the intrinsic kinase activity of the Chk1-KD protein (Foloppe et al., 2005). Additionally, concentration of both proteins proved challenging, requiring regular mixing throughout the concentration process, and it was not possible to concentrate either of these above 6.5 mg/ml due to precipitation of the sample. Both proteins were used to investigate the binding of the Chk1-KD to Claspin CKB-motif phosphopeptides and for co-crystallisation trials.

7.4 Chk1-KD binds to CKB-motif phosphopeptides

It was previously shown that Chk1 was able to interact with a single composite CKB-motif phosphopeptide (based on sequence homology) (Clarke and Clarke, 2005). However, the affinity of the Chk1-KD protein for individual CKB motifs had not been determined. Therefore, FP was used to study these individual interactions. Synthetic phosphopeptides were based on each of the three 10 amino acid CKB motifs described by Kumagai and Dunphy, (2003), but with additional amino acids at both the N- and C-termini based on their conservation across homologues of human Claspin (**Figure 7.2A**). The sequence of these peptides was also compared to the composite CKB phosphopeptide previously shown to interact with Chk1 by Clarke and Clarke, (2005). Fluorescently-labelled peptides containing the CKB motifs were synthesised: with and without the phosphorylated residue, and a fluorescein label at the N-terminus. The CKB phosphopeptides are termed: Flu-T916p (motif 1; aa 908-922), Flu-S945p (motif 2; aa 937-951) and Flu-S982p (motif 3; aa 974-988), whilst the CKB non-phosphopeptides are named: Flu-T916, Flu-S945 and Flu-S982 respectively (**Figure 7.2B**). FP assays were carried out and analysed as described in section **6.3**.

A

```
Q9HAW4 CLSPN_HUMAN  903  AMDANMDELLDLCTGKFTSQAE-KHL-PRKSDKKENMEELLNLCSGKFTSQDASTPASSE
Q9DF50 CLSPN_XENLA  851  AMDANMDELLDLCSGQFKESLSGTSQAAESDAKKQPMDELLELCSGKFVSQADCSTQDSS
Q80YR7 CLSPN_MOUSE  880  AMDANMDELLDLCTGQFTSQPEEKCQ-PRKNDKKENMEELLNLCSGKFPTQDASPVAPLG
D2XSJ5 D2XSJ5_CHICK 898  AMDANMDELLDLCSGQFSSQAEH-VP-STSSNKKQNMEELLNLCSGKFVTQNSPTWASSV
                         *************:*:*... .       .. **: *:***:****** :*

Q9HAW4 CLSPN_HUMAN  961  LNKQEKESSMGDPMEEALALCSGSFPTDKEEEDE----EEEFGDFRLVSNDNEFDSDEDE 1
Q9DF50 CLSPN_XENLA  911  ASAKDRSTAVKKDISDEVATVSSSFLTEREQEED---EEEEFGEFKLLPCDDSESENEEQ
Q80YR7 CLSPN_MOUSE  939  LRSQEKESSTEDPMEEALALCSGSFPTDREEEGE----EEEFGDFQLVSKENGFASDEDE
D2XSJ5 D2XSJ5_CHICK 956  SSKAEKDSDIEDPMAEALALCSGSFPTDREEEDEEHEEHEELGDFQLVTDDNAFDSEEDE 1
                             ::.:   . : : :*  *.** *::*:* :    .**:*:*:*:  ::   .:*::
```

B

```
Flu-T916p  Flu-GGMDELLDLCTGKFTSQ
Flu-S945p  Flu-GGMEELLNLCSGKFTSQ
Flu-S982p  Flu-GGMEEALALCSGSFPTD
                    *:* * **:*.* ::
```

Figure 7.2 Claspin CKBD sequence alignments.
(**A**) Amino acid sequence alignment of the CKBD sequence (CKB motifs are highlighted pink) of Claspin from human (Uniprot ID: Q9HAW4) (aa 903-958), *Xenopus* (Uniprot ID: Q9DF50) (aa 851-967), mouse (Uniprot ID: Q80YR7) (aa 880-994) and chicken (Uniprot ID: D2XSJ5) (aa 898-1015). The alignment was prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010). (**B**) Amino acid sequence alignment of human Claspin CKB motifs (Uniprot ID: Q9HAW4) as synthesised peptides, where yellow highlight indicates the phosphorylation site. Flu=Fluorescein. The alignment was prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010).

For the interaction of Chk1-KD$^{1-270}$-His with each of the phosphopeptides it was possible to fit binding curves, and determine disassociation constants (K_d), as these had the expected hyperbola-shaped binding curves (**Figure 7.3**); each of which were in the low micromolar range, with an observed preference for binding of Flu-S945p (motif 2) < Flu-T916p (motif 1) < Flu-S982p (motif 3) (**Table 7.1**). The weakest interaction was found with motif 3, which is the least conserved across the Claspin homologues and is not found in the sequence of *Xenopus* Claspin (Kumagai and Dunphy, 2003). Longer incubations of Chk1-KD$^{1-270}$-His with the phosphopeptides, resulted in decreased levels of signal and eventually, all binding was lost. However, signal could be restored by the addition of fresh phosphopeptides. This strongly indicated the purified protein was potentially contaminated with low levels of a protein phosphatase. For all subsequent experiments PhosStop tablets (a phosphatase inhibitor cocktail) were used in all experimental buffers; which eliminated the observed decrease of binding signal with time. For each of the non-phosphopeptides no robust interaction was observed (data not shown).

Peptide	Motif	$K_{d\ (\mu M)}$
Flu-T916p	1	1.05
Flu-S945p	2	0.67
Flu-S982p	3	1.66
Flu-T916	1	-
Flu-S945	2	-
Flu-S982	3	-

Table 7.1 Binding constants for Chk1-KD$^{1-270}$-His binding to Claspin CKB motifs. Binding constants calculated from **Figure 7.5**.

As a negative control for phosphopeptide binding, the assay was repeated using the purified Chk2-KD (Oliver et al., 2006), which does not interact with the CKBD of Claspin (Chini and Chen, 2003). As expected, no interaction was observed by FP for Chk2-KD binding any of the CKB peptides tested (**Appendix 3.5**).

7.4.1 Staurosporine-bound Chk1-KD can still bind to CKB phosphopeptides

The structure of Chk1-KD with STU, shows Chk1-KD is able to interact with STU and the sulphate ion simultaneously (PDB: 1NVR) (**Figure 7.4**). We subsequently

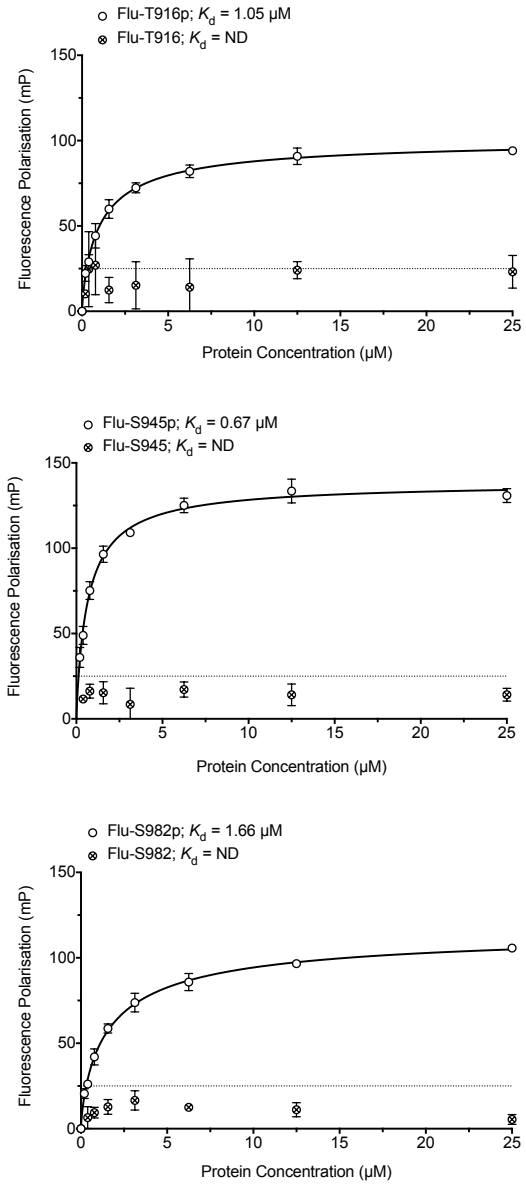


Figure 7.3 Recombinant Chk1-KD$^{1\text{-}270}$-His binds to phosphorylated CKB motifs.

Binding of Chk1-KD$^{1\text{-}270}$-His to 5'-fluorescein synthetic peptides (representing the CKB motifs of human Claspin) was tested by FP. In each case, the indicated peptide was at a final concentration of 100 nM. Data were measured using a POLARstar Omega multimode microplate reader (BMG Labtech). Each data point is the mean of three individual experiments and error bars represent one standard deviation. The identity of each CKB peptide is indicated in the top left hand corner of each graph. Binding constants (K_d) for each experiment are shown, where they could be calculated. ND= K_d not determined.

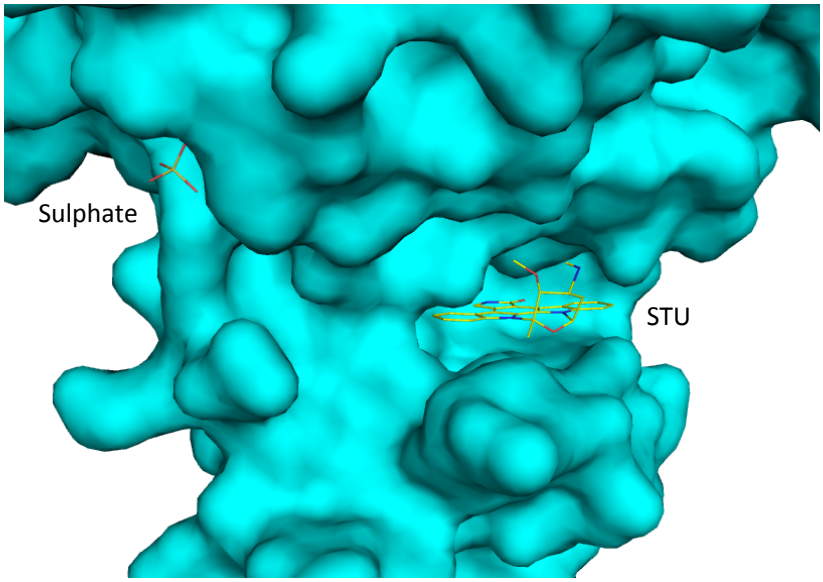


Figure 7.4 Structure of Chk1-KD bound to STU and a sulphate ion.
The Chk1-KD tertiary structure (PDB: 1NVR) (cyan), showing the interaction with a sulphate ion and staurosporine (STU) simultaneously.

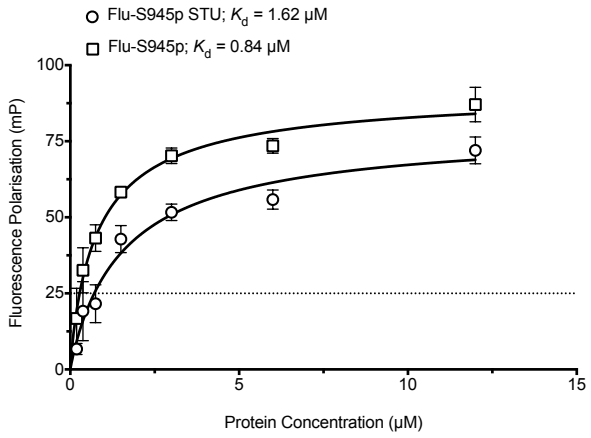


Figure 7.5 STU inhibited recombinant Chk1-KD$^{1\text{-}270}$-His binds to the S945p CKB phosphopeptide.
Binding of STU inhibited Chk1-KD$^{1\text{-}270}$-His to Flu-S945p was tested by FP. In each case, the indicated peptide was at a final concentration of 100 nM. Data were measured using a POLARstar Omega multimode microplate reader (BMG Labtech). Each data point is the mean of three individual experiments and each error bar represents one standard deviation from this mean. Binding constants (K_d) for each experiment are shown.

tested if the CKB phosphopeptide that bound with the highest affinity (Flu-S945p) was affected by the binding of Chk1-KD to STU. Chk1-KD$^{1-270}$-His was pre-incubated with STU at a ratio of 1:1.2 protein:STU prior to setting up the FP assay as in section **6.3**.

STU-inhibited Chk1-KD$^{1-270}$-His retains the ability to bind to the Flu-S945p phosphopeptide. It was possible to fit binding curves, and determine disassociation constants (K_d) for both the uninhibited and inhibited protein, as these had the expected hyperbola-shaped binding curves. There was a good agreement between the K_d obtained for Chk1-KD$^{1-270}$-His binding to the Flu-S945p phosphopeptide (K_d = 0.84 μM) in this experiment and the previous one (K_d = 0.67 μM). The relatively small differences could be attributed to the binding curves not completely reaching saturation (due to low protein yields), resulting in the concomitant inaccuracy in the K_d values obtained, or could be due to slight batch-to-batch differences in the purity of the recombinant protein. There was however a 2-fold reduction in the K_d of Chk1-KD$^{1-270}$-His for the Flu-S945p phosphopeptide, when bound to STU (K_d = 1.62 μM) (**Figure 7.5**). STU is a relatively large compound that occupies the active site of the kinase and a large area surrounding it, whilst this does not interfere with the sulphate ion binding (refer to **Figure 7.4**), this may interfere with other interactions the CKB peptide makes with the Chk1-KD (i.e. by structural alteration or blocking the interaction), causing the reduction in the binding affinity.

7.4.2 Mutation of the CKB motif abolishes Chk1-KD binding

The sequence of the CKBD is highly conserved across the Claspin homologues and there are a number of highly conserved amino acids in the motif sequence (see **Figure 7.2B**). The four most conserved amino acids in the motif were individually mutated in the context of the previously tested Flu-S945p phosphopeptide (tightest interaction). Prior to synthesis of these peptides, it was advised to remove the N-terminal methionine due to its proximity to the Fluorescein label, which could result in the oxidation of the methionine and difficulties in obtaining sample purity (personal communication with Peptide Protein Research). Therefore, the original Flu-S945p phosphopeptide was re-synthesized without the

N-terminal methionine (Flu-937M-), as were the mutants; Lys941Gly (Flu-L941G), Lys943Gly (Flu-L943G), Cys944Ala (Flu-C944A) and Phe948Ala (Flu-F948A) (**Figure 7.6A**). Binding was measured using FP as in section **6.3**.

Binding of the (control) Flu-937M- phosphopeptide to Chk1-KD$^{1-270}$-His was greatly reduced, with an apparent 10-fold reduction in the binding affinity when compared to that for binding to Flu-S945p; K_d = 7.29 μM verses 0.67 μM respectively. (**Figure 7.6B**). This reduction in binding affinity may be through loss of secondary structure propensity; the methionine residue being directly required for the interaction; or due to the Fluorescein tag interfering with the interaction. Mutation of any of the four highly-conserved CKBD residues resulted in a loss of interaction with Chk1-KD (**Figure 7.6C**). Again these data indicate a requirement for the highly conserved amino acids within the motif, either through specific interactions between these amino acids with Chk1, or through an alteration in the secondary structure of the motif by mutation.

7.5 Chk1 KD is not further activated by binding the CKB motif

Recombinant Chk1-KD protein has been shown, *in vitro*, to be an active protein kinase (Chen et al., 2000). The interaction of Chk1 with the CKBD only requires one phosphorylated motif, however activation of Chk1 (by ATR), requires the CKBD to be phosphorylated on both Thr916 and Ser945 (Clarke and Clarke, 2005).

The commercially available ADP-Glo assay, with a Chktide peptide as a phosphorylation substrate, was used to confirm if the Chk-KD$^{1-289}$-His recombinant protein was active, and to determine if binding of a CKB phosphopeptide (Flu-S945p) resulted in increased Chk1 activity (section **2.28**). As a control, this assay was also carried out using recombinant active full-length Chk1 protein (Promega). The ADP-Glo assay determines the rate of ATP turnover over a set time period. After termination of the reaction, by the addition of a quenching agent (ADP-Glo Reagent), any remaining ATP is first depleted. Subsequently, the ADP product is converted back into to ATP (Kinase Detection Reagent), with the newly synthesized ATP detected by a luminescent (luciferase/luciferin) reaction.

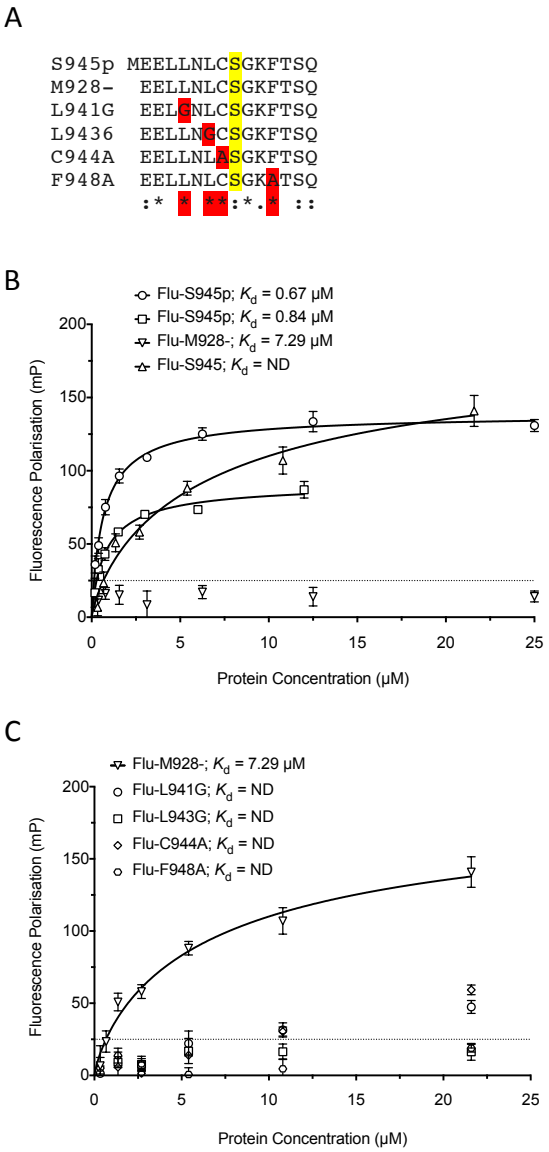


Figure 7.6 Mutation of the CKB motif inhibits Chk1-KD$^{1-270}$-His binding.
(**A**) Alignment of the human Claspin CKB motif 2 (Uniprot ID: Q9HAW4) where yellow highlight indicates the phosphorylation site and red indicates the individual phosphopeptide mutation. The alignment was prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010). (**B**-**C**) Binding of Chk1-KD$^{1-270}$-His to 5'-fluorescein synthetic peptides (representing the mutated CKB motifs of human Claspin) was tested by FP. In each case, the indicated peptide was at a final concentration of 100 nM. Data were measured using a POLARstar Omega multimode microplate reader (BMG Labtech). Each data point is the mean of three individual experiments and error bars represent one standard deviation. Where '**B**' is Chk1-KD$^{1-270}$-His binding to Flu-S945p, Flu-945 and Flu-M928- peptides, and '**C**' is Chk1-KD$^{1-270}$-His binding Flu-M928-, Flu-L941G, Flu-L943G, Flu-C944A and Flu-F948A peptides. Binding constants (K_d) for each experiment are shown, where they could be calculated. ND= K_d not determined.

Full-length Chk1 or Chk1-KD$^{1-289}$-His were pre-incubated with the Flu-S945p phosphopeptide or Flu-S945 peptide, at 1:2 molar stoichiometry (protein:peptide). The assay was carried out in triplicate using a 384-well plate, and the resulting luminescence measured using a PHERAstar FS plate reader (BMGLabTech). The assay confirmed that Chk1-KD$^{1-289}$-His was active, as has previously been described (Chen et al., 2000). Furthermore, the activity of both this protein or the full-length protein did not change after incubation with either the CKB phosphopeptide or peptide (**Figure 7.7**).

7.6 Enhancing the stability and solubility of Chk1-KD

The kinase domain of human Chk1 has previously been crystallized in a range of different conditions, typically at neutral pH, in sulphate-containing conditions, and in complex with an inhibitor (Chen et al., 2000, Zhao et al., 2002, Foloppe et al., 2009, Vanderpool et al., 2009). It was found to be necessary to have >500 mM NaCl in the buffer, when the protein was concentrated to above 2 mg/ml – as at lower salt concentrations the protein started to precipitate. Additionally, the protein was (at times) difficult to produce, so it was decided that key factors that could enhance protein stability and solubility needed to be determined in order to obtain sufficient protein for crystallographic trials. Thermal denaturation (ThermoFluor) of Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His was carried out as described in section **5.11**. Three different concentrations of protein (0.75, 1.9 and 3.75 μM) were assayed in order to determine the optimal concentration to be used in subsequent experiments. For Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His, a single cooperative transition from folded to unfolded states was observed, with calculated temperature mid-points (*T*m) of 48.6 °C (**Figure 7.8A**) and 48.1 °C (**Figure 7.8B**) respectively (**Table 7.2**). The protein concentration for subsequent assays was selected to be 1.9 μM, as this provided a suitable (single and sharp) denaturation signal.

The thermal stability of Chk1-KD$^{1-270}$-His (at 1.9 μM) was measured in six different buffer conditions: 100 mM (i.e. 5 x the concentration of the protein purification buffer) MES pH 6.5; HEPES pH 7.0 and pH 7.5; and Tris pH 7.5, pH 8.0 and pH 8.5. The assay revealed minor differences in the observed *T*m for each of the buffers

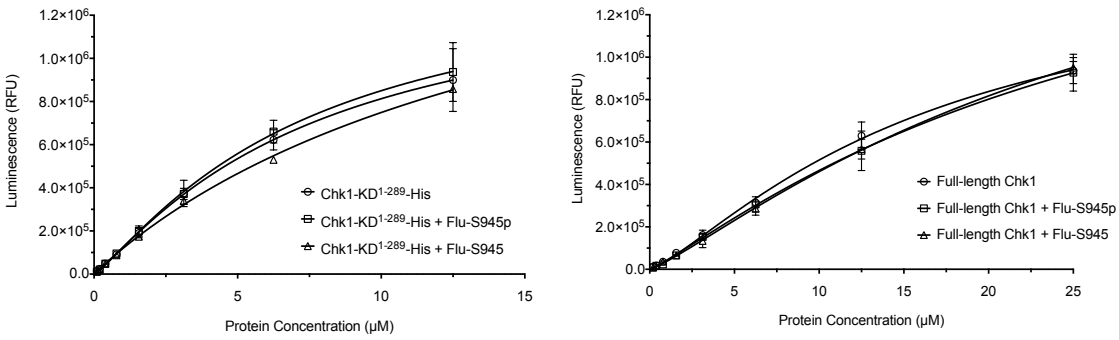


Figure 7.7 Chk1 is not further activated by the CKB phosphopeptide.
ADP-Glo assay of (left) Chk1-KD$^{1-289}$-His and (right) full-length Chk1 (Promega), or protein incubated with Flu-S945p or the Flu-S945 CKB phophopeptides at room temperature for 20 minutes. A 5 μl kinase reaction in 1x kinase reaction buffer prepared with 50 μM ATP and 200 μg/ml Chktide as the substrate. Samples were transferred into rows in a white 384-well plate and incubated at room temperature for 40 minutes. The remaining ATP was depleted using 5 μl ADP-Glo reagent by incubation at room temperature for 40 minutes. The conversion of ADP to ATP (detected by a luciferase / luciferin) was by the addition of 10 μl kinase detection reagent and incubation at room temperature for 40 minutes. Luciferase luminescence was measured using a PHERAstar FS with a luminescence optic. Each data point is the mean of three individual experiments and each error bar represents one standard deviation. The identity of each peptide with Chk1 is indicated in the bottom right hand corner of each graph.

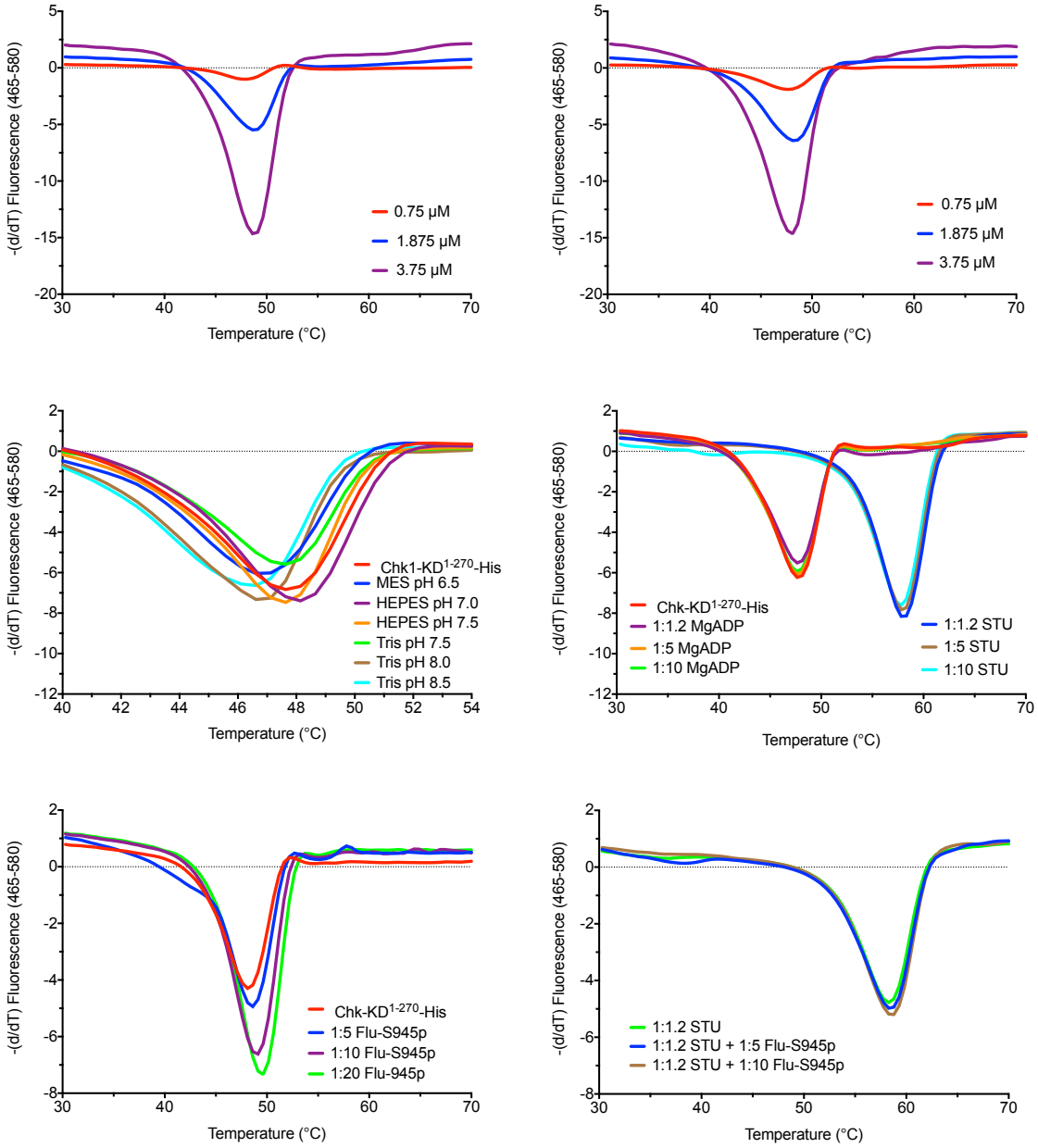


Figure 7.8 Thermal denaturation of Chk1-KD proteins.
For thermal denaturation, protein samples were mixed with 5x SYPRO orange, and melting profiles were obtained from 20 to 80 °C, with a temperature ramp of 0.02°C s^{-1} using the Roche LightCycler 480. Each additive was incubated with the protein on ice for 20 minutes prior to preparation of the assay. Each plot is the mean of three individual experiments. (**A**) Chk1-KD$^{1-270}$-His protein. (**B**) Chk1-KD$^{1-289}$-His protein. (**C**) Chk1-KD$^{1-270}$-His protein thermal denaturation buffer pH stability screen. (**D**) Chk1-KD$^{1-270}$-His thermal denaturation with MgADP or STU. (**E**) Chk1-KD$^{1-270}$-His thermal denaturation with the CKB Flu-S945p phosphopeptide. (**F**) Chk1-KD$^{1-270}$-His thermal denaturation with the CKB Flu-S945p phosphopeptide and STU. The identity of each sample is indicated at the bottom of each graph, where the ratio is given with respect to Chk1-KD$^{1-270}$-His. The melting temperatures (*T*m) were determined by the point of the minima of the negative peak of the denaturation curve (see **Tables 7.2**, **7.3**, **7.4**, **7.5** and **7.6**).

Protein	Concentration (μM)	Calculated *T*m (°C)
Chk1-KD$^{1-270}$-His	0.75	48.1
	1.875	48.6
	3.75	48.6
Chk1-KD$^{1-289}$-His	0.75	47.6
	1.875	48.1
	3.75	48.6

Table 7.2 Thermal denaturation of Chk1-KD proteins. Thermal denaturation *T*m for Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His as determined from **Figure 7.10A** and **Figure 7.10B**.

Protein	Additive	Calculated *T*m (°C)
Chk1-KD$^{1-270}$-His	-	47.6
	Mes pH 6.5	46.1
	HEPES pH 7.0	47.6
	HEPES pH 7.5	47.6
	Tris pH 7.5	46.6
	Tris pH 8.0	46.6
	Tris pH 8.5	46.1

Table 7.3 Thermal denaturation of Chk1-KD$^{1-270}$-His in alternative buffering systems. Thermal denaturation *T*m for Chk1-KD$^{1-270}$-His as determined from **Figure 7.10C**.

Protein	Additive	Calculated *T*m (°C)
Chk1-KD$^{1-270}$-His	-	47.6
	1.2x MgADP	47.6
	5x MgADP	47.6
	10x MgADP	47.6
	1.2x STU	57.8
	5x STU	57.8
	10x STU	57.8

Table 7.4 Thermal denaturation of Chk1-KD$^{1-270}$-His with chemical additives. Thermal denaturation *T*m for Chk1-KD$^{1-270}$-His as determined from **Figure 7.10D**.

tested, but with no obvious enhancement in the stability of the protein (**Figure 7.8C** and **Table 7.3**). Subsequently, the thermal stability was determined in the presence of MgADP or STU at molar ratios of 1:1.2, 1:5 and 1:10 (protein:additive). The presence of MgADP did not change the *T*m of the protein (47.6 °C), whereas incubation with STU resulted in a 10 °C positive shift (57.8 °C) (**Figure 7.8D** and **Table 7.4**). Interestingly, the structure of Chk1-KD in complex with AMP-PMP (a non-hydrolysable analogue of ATP) had only minor structural differences to the structure of the apo-enzyme as described by Chen et al., (2000); therefore binding of ADP to the protein may not result in an increased thermal stability. Assays including EDTA (a metal ion chelating agent) also did not alter the thermal stability of Chk1-KD$^{1-270}$-His, indicating that metal ions are not required for protein stability (data not shown). STU increases the thermal stability of protein kinases domains through alterations in their conformation; i.e. through the large number of contacts that STU makes with the amino acids of the kinase active site (Prade et al., 1997) (section **1.6.1**).

The thermal stability of Chk1-KD$^{1-270}$-His was also determined in the presence of the Flu-S945p CKB phosphopeptide; at molar ratios of 1:5, 1:10 and 1:20 (protein:peptide). Samples were also prepared with STU, included at a molar ratio of 1:1.2 (protein:STU). Incubation with the Flu-S945p phosphopeptide, produced only a modest increase in thermal stability (20:1 ratio peptide:protein; 49.6 °C), but this was considered not to be significant (**Figure 7.8E** and **Table 7.5**). Incubation with both the phosphopeptide and STU, produced a 10 °C increase in the *T*m (as shown previously with STU of 58.2 °C) indicating that the presence of the phosphopeptide did not further influence the stability of the protein, as judged by this thermal denaturation assay (**Figure 7.8F** and **Table 7.6**).

To potentially enhance the chances of co-crystallising the Chk1 kinase domain with a CKB phosphopeptide, the selected buffer was HEPES pH 7.5 and using high molar ratios of protein:peptide, prepared with and without the inclusion of STU in order to stabilise the protein.

Protein	Additive	Calculated *T*m (°C)
Chk1-KD$^{1-270}$-His	-	48.1
	5x Flu-S945p	48.6
	10x Flu-S945p	49.1
	20x Flu-S945p	49.6

Table 7.5 Thermal denaturation of Chk1-KD$^{1-270}$-His with the CKB Flu-S945p phosphopeptide. Thermal denaturation *T*m for Chk1-KD$^{1-270}$-His as determined from **Figure 7.10E**.

Protein	Additive 1	Additive 2	Calculated *T*m (°C)
Chk1-KD$^{1-270}$-His	1.2x STU	-	58.2
	1.2x STU	10x Flu-S945p	58.2
	1.2x STU	20x Flu-S945p	58.2

Table 7.6 Thermal denaturation of Chk1-KD$^{1-270}$-His with staurosporine and the CKB Flu-S945p phosphopeptide. Thermal denaturation *T*m for Chk1-KD$^{1-270}$-His as determined from **Figure 7.10F**.

7.7 Crystallisation trials with the Chk1 kinase domain

Purified Chk1-KD$^{1-289}$-His was concentrated to 6.2 mg/ml and set up in an optimisation screen (as previously described for Claspin, section **5.11**) that was based on the previously published crystallisation conditions: 2.5 - 8% (w/v) PEG 8000, 0.1 M HEPES pH 7.4, 32 - 36% (v/v) ethylene glycol (Vanderpool et al., 2009). These were manually set up as sitting drops, by mixing 0.5 µl of protein with 0.5µl of mother liquor, against a well volume of 100 µl, and incubated at a temperature of 14 °C. Microcrystals were visible after 3 days, and after 2 weeks, small crystals could be harvested from a crystallisation condition containing 8% (w/v) PEG 8000, 0.1 M HEPES pH 7.5, 36% (v/v) ethylene glycol. These were then flash-frozen in and stored under liquid nitrogen.

Diffraction data was collected from a single crystal at a temperature of 100K on beamline I02 at the Diamond Light Source (DLS, Didcot, UK). Diffraction data was automatically processed at the beamline using the Xia2 pipeline (Winter, 2010). Using programs from the CCP4 suite (Collaborative Computational Project, Number 4; Winn et al., 2011), phases were solved by molecular replacement using PDB: 1IA8 as a search model (Chen et al., 2000). Iterative model refinement indicated that the structure of Chk1-KD$^{1-289}$-His was essentially identical to the previously solved structure by Chen et al., (2000) (**Figure 7.9A**).

7.8 Co-crystallisation of Chk1-KD with CKB phosphopeptides

In the crystal structure of Chk1-KD (PDB: 1IA8), there is a sulphate ion near the catalytic centre of the kinase domain (**Figure 7.9A**), which is coordinated in a positively charged pocket, formed from the side chains of four conserved amino acid residues: Lys54, Arg129, Thr153 and Arg162 (**Figure 7.9B**) (Chen et al., 2000). Individual mutation of these four residues in Xenopus Chk1 prevented the stable association of the protein with Claspin (Jeong et al., 2003). The interaction of Chk1 with the Claspin CKBD is required for the activation of the replication checkpoint during replication stress, but the tertiary structure of the complex is unknown. The minimal interaction region for Chk1 binding to the Claspin CKBD is an individual CKB motif; however for activation of Chk1 this requires two

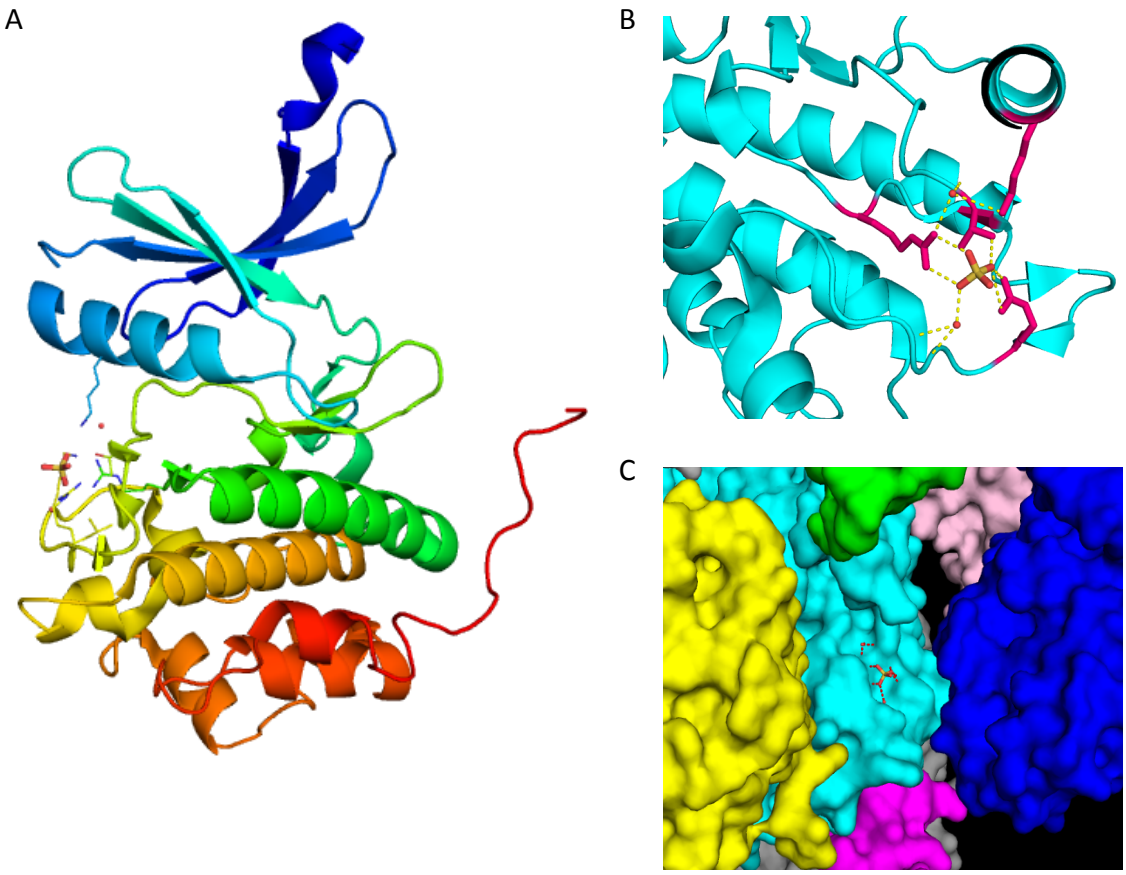


Figure 7.9 Structure of Chk1-KD.
(**A**) Structure of Chk1-KD (aa 1-289) (PDB: 1IA8). Structure is shown in molecular cartoon representations, coloured from blue to red, from the N- to C-terminus respectively (Jones' Rainbow). (**B**) Structure of Chk1-KD (PDB: 1IA8) (coloured cyan), showing Lys54, Arg129, Thr153 and Arg162 (coloured pink) coordinating a sulphate ion. (**C**) Chk1-KD crystal contacts (PDB: 1IA8). Individual molecules are coloured and the sulphate ion is shown (bound to the cyan molecule).

phosphorylated motifs, potentially required for bringing two Chk1 molecules together for dimerisation and cooperative activation by phosphorylation. An individual CKB motif is small, which is ideal for co-crystallisation studies, as this would potentially fit into the current crystal form, without disruption of the lattice. However, a number of crystal contacts are made in the lattice surrounding the proposed CKB interaction site, which may hinder the complex co-crystallisation in the current lattice form (**Figure 7.9C**). Whilst a larger CKB peptide, containing two phosphorylated CKB motifs, could be used for crystallography, this would be expensive to produce (discussed in section **7.9**).

7.8.1 Chk1-KD$^{1-289}$-His optimisation co-crystallisation

The optimisation conditions that yielded Chk1-KD$^{1-289}$-His crystals (section **7.7**) were used for initial co-crystallisation screens. Chk1-KD$^{1-289}$-His protein was incubated with the S945p phosphopeptide (at a 1:5 molar ratio, protein:peptide), with and without the addition of STU (at a 1:1.2 molar ratio, protein:STU), and were put into crystallisation trials as described in section **7.7**. However, no crystals or signs of crystal growth were obtained.

7.8.2 Soaking of Chk1-KD$^{1-289}$-His crystals

Additionally, soaking the Chk1-KD$^{1-289}$-His protein crystals with the phosphopeptides in reservoir solution was attempted, in order to incorporate the phosphopeptide into the crystal lattice. Crystals of Chk1-KD$^{1-289}$-His were harvested and then placed into mother liquor that was supplemented with 0.8 mM of either the Flu-S945p or S945p phosphopeptides, for a period of 16 hours at 4 °C. However, after incubation, the visual quality of the protein crystals was found to have deteriorated substantially. The crystals were harvested and then frozen in liquid nitrogen by Dr Mark Roe (University of Sussex). Diffraction data were collected for a soaked crystal on beamline I04 at DLS. However, the observed diffraction pattern was streaky and smeary, indicating the presence of several overlaying crystal lattices. The resolution of the diffraction data collected was also severely reduced.

7.8.3 Sparse matrix co-crystallisation screening of Chk1-KD$^{1-270}$-His and T916p CKB phosphopeptide complexes

Sparse matrix co-crystallisation screens were set up with Chk1-KD$^{1-270}$-His at 6 mg/ml mixed with the T916p phosphopeptide at a molar ratio of 1:5 protein:peptide and incubated on ice for 2 hours. These were set up, in sitting drop plates, as 0.2 µl protein with 0.2 µl mother liquor drops, incubated against a 50 µl well volume, using an ARI Crystal Phoenix Liquid Handling System with different commercially available crystallisation screens (PEG/Ion, Structure Screen I + II and JCSG-plus), and incubated at a temperature of 20 °C.

Small protein crystals and crystalline precipitates were detected in a number of different conditions including: PEG/Ion A2 (**Figure 7.10A**) and C9 (**Figure 7.10B**), and Structure Screen I + II (**Figure 7.10C**). These conditions were then used to inform subsequent optimisation screens, to (a) reproduce the initial crystals, (b) improve crystal quality and (c) to hopefully generate single crystals that would diffract to high resolution. Screens were set up manually, in sitting drops as 0.5µl protein with 0.5µl mother liquor drops, against a well volume of 100 µl, and incubated at 20 °C. Crystals of varying visual quality were detected after five days – these included; condition '1E6' (**Figure 7.10D**), condition '2C2' (**Figure 7.10E**) and condition '2B6' (**Figure 7.10F**). Single crystals from both initial and subsequent optimisation screens were harvested after ~2 weeks by Dr Mark Roe (University of Sussex). These were cryo-protected using glycerol, by step-wise changes of buffer with increasing concentrations of glycerol [to 30% (v/v) glycerol] and then frozen and stored under liquid nitrogen.

Diffraction data were collected for 14 different crystals on beamline I04 at DLS, and these were processed as before (section **7.7**). The resolution of diffraction data ranged from 2.51 to 3.95 Å. In each of these crystals, the unit cell dimensions were larger that those previously found for crystals of Chk1-KD alone (Chen et al., 2000), and could be sub-categorised into three groups; large orthorhombic, small orthorhombic, and monoclinic (**Table 7.7**). In each case, phases were determined using molecular replacement using PDB: 1IA8 as previously described (section **7.7**). No density for the N-terminal half of the protein was visible in any of the

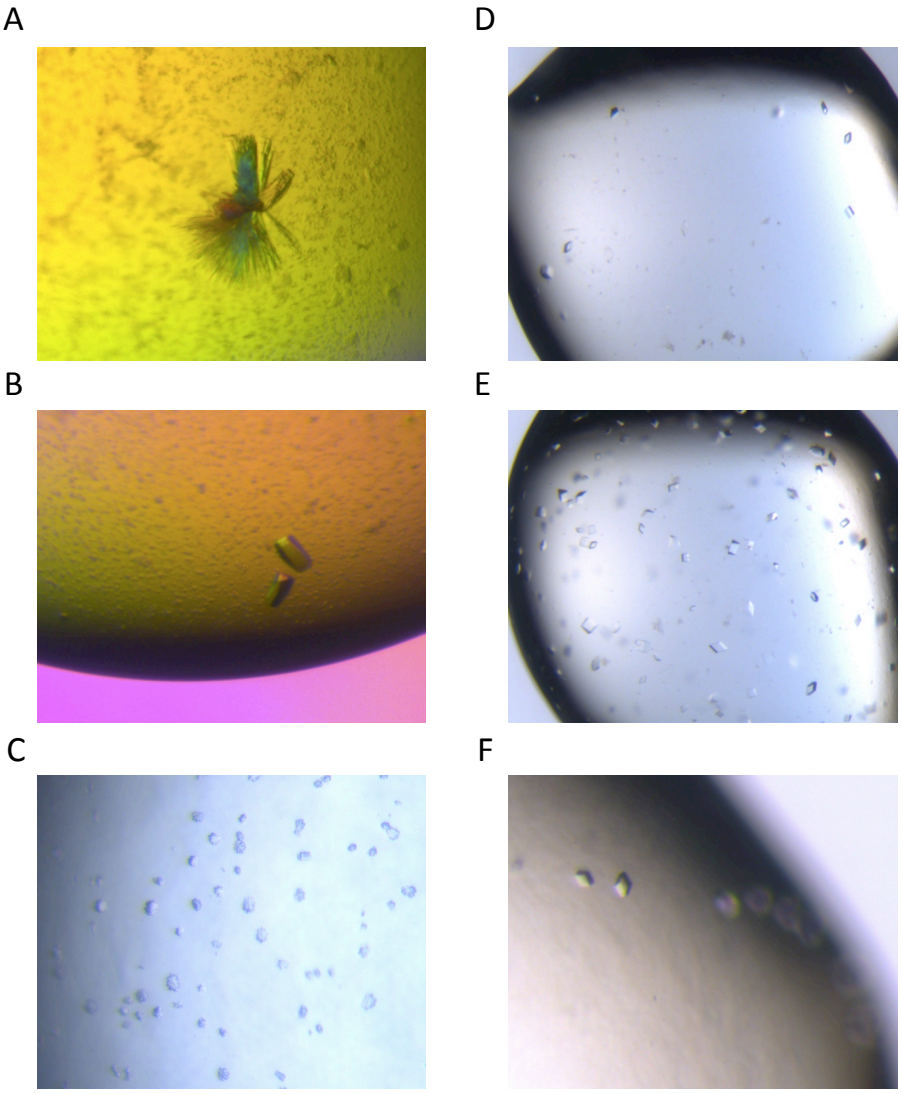


Figure 7.10 Chk1-KD$^{1-270}$-His protein crystals.
Protein crystals grown from Chk1-KD$^{1-270}$-His incubated with the T916p phosphopeptide at a 5:1 peptide:protein ratio at 20 °C. (**A**) PEG/Ion screen (Hampton Research) condition A2 [0.2 M Potassium fluoride pH 7.3, 20% (v/v) PEG 3350]. (**B**) PEG/Ion screen condition C9 [0.2 M Sodium sulphate decahydrate pH 6.7, 20% (v/v) PEG 3350]. (**C**) Structure Screen I+II (Molecular Dimensions) condition D1 [0.2 M Sodium acetate pH 8.0, 20% (v/v) PEG 4000]. (**D**) Optimisation screen of '**A**', condition '1E6' [0.25 M Potassium fluoride pH 7.3, 22% (v/v) PEG 3350]. (**E**) Optimisation screen of '**B**', condition '2C2' [0.2 M Sodium sulphate decahydrate pH 6.7, 18% (v/v) PEG 3350]. (**F**) Optimisation screen of '**C**', condition '2B6' [0.1 M Tris pH 8.0, 0.2 M Sodium acetate, 25% (v/v) PEG 4000].

calculated maps (aa 1- ~36), and a number of amino acids appeared to be structurally disordered and were not able to be modelled (aa ~36- ~60). However, the remainder of the protein structure was identical to that previously determined. Crystals obtained in sulphate-containing conditions, had regions of electron density consistent with a bound sulphate ion, and were correctly liganded by the side-chain of Arg162, as previously reported in Chen et al, (2000). The highest resolution datasets were subjected to sequential rounds of model building and refinement, but did not show any difference density consistent with a bound phosphopeptide. The observed changes in unit cell dimensions were driven by a crystal contact formed between the degraded N-terminus of Chk1-KD$^{1-270}$-His protein, resulting in two protein molecules forming the asymmetric unit instead of one molecule as previously described by Chen et al., (2000). As these structures did not contain the desired phosphopeptide, these were not refined any further.

Chk1 Structure	Unit Cell	Space Group	Unit Cell Dimensions					
			a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)
1IA8[1]	Monoclinic	P 1 2_1 1	45.20	65.70	58.10	90.00	93.90	90.00
M5S2[2]	L. Orthorombic	P 2_1 2_1 2	120.52	63.56	76.44	90.00	90.00	90.00
M6S11[3]	S. Orthorombic	P 2_1 2_1 2	66.02	109.96	45.18	90.00	92.64	90.00
M5S5[4]	Monoclinic	P 1 2_1 1	65.74	120.50	76.83	90.00	91.59	90.00

Table 7.7 Unit cell dimensions of Chk1-KD$^{1-270}$-His crystals.
Parameters for M5S2, M6S11 and M5S5 datasets are shown (in comparison to PDB: 1IA8), as representative examples of the 14 different crystal datasets collected on beamline I04 at DLS. [1]Chk1-KD (PDB: 1IA8), [2]Chk1-KD$^{1-270}$-His protein crystal [0.25 M Sodium sulphate, 18% (v/v) PEG 3350], [3]Chk1-KD$^{1-270}$-His protein crystal [0.2 M Sodium acetate, 0.1 M Tris pH 8.0, 25% (v/v) PEG 4000], [4]Chk1-KD$^{1-270}$-His protein crystal [0.25 M Sodium fluoride, 24% (v/v) PEG 3350]. L=large, S=small.

To investigate protein degradation, purified Chk1-KD$^{1-289}$-His protein samples incubated with the Flu-S945p phosphopeptide were incubated at a temperature of 4 °C for a period of 6 weeks, and then analysed by SDS-PAGE. Whilst a sample incubated with STU remained stable, with no obvious signs of degradation, those that were not ran as a smaller species at a molecular mass roughly equivalent to that observed in the protein crystals (**Figure 7.11**). This analysis indicated there was likely to be problems with protease contaminants, which was not able to cleave the N-terminus of Chk1 KD in the presence of STU; likely as STU greatly

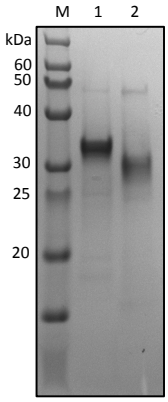


Figure 7.11 Purified Chk1-KD$^{1-289}$-His protein stability.
SDS-PAGE analysis of purified Chk1-KD$^{1-289}$-His after incubation with the Flu-S945p phosphopeptide for 6 weeks at 4 °C. M=molecular mass marker, 1=incubation with STU and 2=incubation without STU. 4-20% TruPAGE Precast SDS-PAGE gel, stained with Instant Blue.

increases the stability of the protein kinase structure. This contaminant could have been carried through during the purification of Chk1 KD (discussed in section **7.9**).

7.8.4 Additional sparse matrix co-crystallisation screening

Further attempts at co-crystallography used either Chk1-KD$^{1-270}$-His or Chk1-KD$^{1-289}$-His in complex with the CKB phosphopeptides (Flu-T916p, Flu-S945p, T916p or S945p). Crystal screens were set up at a protein concentration of between 2.5 and 6.5 mg/ml in sitting drops as 1:1, 2:1 or 3:1 drops (protein to mother liquor ratios; final drop volume of 0.4, 0.6 or 0.8 µl respectively) using an ARI Crystal Phoenix Liquid Handling System or a Douglas Instruments Oryx8 robot. Crystal screens were set up with a wide range of different commercially available crystallisation screens (PEG/Ion, Structure screen I + II, PACT premier, JCSG-plus, ProPlex, Morpheus, and MIDAS) and duplicates plates were incubated at 4, 14 and 20 °C. The plates were regularly checked, and mostly showed either light or heavy precipitate, although a number of conditions remained clear or had a granular type of precipitate. Attempts at optimising a number of these conditions did not result in the growth of protein crystals. The only protein crystals obtained from these screens, were found to be in similar conditions to those described above; and were also identified as forming as a result of Chk1-KD degradation. No protein crystals were identified from any screen where STU was additionally included.

7.9 Summary

The aim of this chapter was to first express and purify the kinase domain of human Chk1 and then further characterise the phospho-specific interaction of this protein with the CKBD of human Claspin. The kinase domain of human Chk1 was successfully expressed as a soluble protein in *Sf*9 insect cells from two different recombinant baculoviruses: Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His. Whilst it had been possible to express a full-length version of the protein (His-Chk1) is was not readily amenable to purification.

The active form of the Chk1 kinase domain was shown to interact with each of the three CKB phosphopeptide motifs found in Claspin, with the tightest interaction

seen with phosphopeptide motif 2 (S945p; 0.67 μM). However, as expected interaction of the kinase domain with a single CKB phosphopeptide motifs did not result in any increase in kinase activity. Chk1-KD can also interact with a CKB phosphopeptide in the presence of STU, but with an approximate two-fold reduction in the binding affinity. Furthermore, mutation of conserved amino acids with the CKB motif prevented interaction, indicating that either secondary structure propensity or direct sequence recognition is required for interaction, in addition to the phosphorylated residue. It has been shown that only a single phosphorylated motif is required for the interaction of Chk1 with Claspin (Clarke and Clarke, 2005), but for activation of Chk1 phosphorylation of two CKB motifs is required (Clarke and Clarke, 2005, Chini et al., 2006). Chk1-KD protein is fully active when expressed in *Sf9* culture (Chen et al., 2000), and it has been shown active full-length Chk1 has a reduced affinity for the phosphorylated CKBD (Jeong et al., 2003, Smits et al., 2006). As it was not possible to produce or source sufficient quantities of active and/or inactive full-length Chk1 protein, it was not possible to determine if Chk1 showed differences in binding to the CKB phosphopeptides or if binding the CKB phosphopeptides resulted in Chk1 activation.

The protein from the Chk1-KD$^{1-289}$-His expression construct was successfully crystallised, and the solved structure was in essence identical to that which had previously been determined (Chen et al., 2000). Attempts at co-crystallising Chk1-KD$^{1-289}$-His with a CKB phosphopeptide, using identical conditions, proved unsuccessful. Furthermore, attempts at soaking Chk1-KD$^{1-289}$-His crystals with a CKB phosphopeptide were also unsuccessful. Subsequent sparse matrix co-crystallisation studies with Chk1-KD$^{1-270}$-His and Chk1-KD$^{1-289}$-His in complex with CKB phosphopeptides also did not result in crystals containing both components. Unfortunately, a degraded form of the kinase domain resulted in the formation of a number of crystals in several different conditions (when in conditions with a CKB phosphopeptide). Crystallisation drops were prepared both manually and automatically with a robot, furthermore both commercially sourced crystallization screens and manually prepared optimisation screens yielded the proteolytic crystal fragments. Additionally, crystals of the proteolysis product took five days

for crystal growth to appear. This indicated the sample was contaminated with a protease, which had been carried through during the purification of Chk1-KD proteins. Interestingly, no protein crystals (degraded or full-length) grew from conditions where STU was incubated in the sample with the CKB phosphopeptide (Chk1-KD was not incubated alone in crystallisation screens with STU). Binding of STU stabilises many protein kinase domains, especially the N-terminal kinase lobe (where proteolysis was observed here), and may prevent access of particular amino acids to proteases, and thus inhibit proteolysis.

CKB phosphopeptides have been suggested to interact with a specific pocket on Chk1, formed by the side chains of Lys54, Arg129, Thr153 and Arg162 – which in the crystal structures reported by Chen et al., (2000), is occupied by a sulphate ion (Jeong et al., 2003). The precise orientation of the CKB phosphopeptide in relation to the rest of the kinase domain is not known, and it is possible that binding of the peptide may not be readily compatible with crystal lattice formation; in agreement with the lack of crystals in sparse matrix screening efforts. Theoretically if the phosphopeptide was not binding, the Chk1 KD protein would crystallize by itself, especially in the conditions where Chk1-KD$^{1-289}$-His and sparse matrix conditions where the Chk1 KD previously was found to crystallize alone; although the presence of the peptide may hinder this.

Future attempts at determining a crystal structure of the kinase domain of human Chk1 in complex with the CKBD region of human Claspin could involve the synthesis of a longer phosphorylated peptide (corresponding to more of the CKBD) and to use this in co-crystallization trials. Whilst it is technically possible to synthesise a peptide containing two phosphorylated CKB motifs, the overall cost is prohibitively high. An alternative approach would be to express the CKBD region as a fusion-protein – but generating the correct phosphorylation sites would be more of a challenge; the only kinase known to phosphorylate this sequence is CK1γ1 kinase (Meng et al., 2011), which can be expressed in the baculovirus insect cell system (Meng et al., 2011, Kurihara et al., 2014). However, as this point of study was carried out towards the end of the funded-period of this experimental work, there was not sufficient time to enable the gene synthesis, bacmid

generation and insect cell expression for this protein. Additional characterisation of the Chk1-KD and CKB phosphopeptide complex in solution could theoretically be carried out by bioSAXS, to determine the binding location of the phosphopeptide on the surface of the kinase domain, and also to probe any structural rearrangements made as a result of phosphopeptide binding (i.e. dimerisation).

Chapter 8

Concluding remarks

It has long been known that fragments (sub-constructs) of target genes that encode domains are often more tractable, yielding soluble, proteolytically stable protein that is correctly folded (Littler, 2010, Prodromou et al., 2007, Savva et al., 2007, Hart and Waldo, 2013). Both bioinformatics-based and fragmentation or mutation library approaches have been developed to overcome difficulties in sub-construct design as discussed in **Chapter 1**. One such methodology is 'Combinatorial Domain Hunting' (CDH) (Reich et al., 2006); in this method a random fragmentation library of constructs is created for subsequent protein expression trials. This has successfully been used to identify sub-constructs that are expressed at high levels in the heterologous host *E. coli* to investigate the domain/s from a target protein for subsequent biochemical, biophysical and structural characterisation (Reich et al., 2006, Littler, 2010, Meier et al., 2012).

In this thesis, expansion of the CDH methodology has been explored. The first approach involved the development and assessment of a functionality screen, to be implemented as a novel step during the CDH methodology, in order to identify sub-constructs that express the functional domain of interest in the correctly folded tertiary structure. Subsequently, CDH was used to identify expression sub-constructs for the protein Claspin, in order to investigate potential domain architecture in the protein. Biochemical and biophysical characterisation of the Claspin sub-constructs was undertaken, with further investigation into the DNA binding propensity of these protein fragments. Finally, the known PPI between Claspin and Chk1 was investigated.

8.1 'Super' CDH methodology development

CDH, as well as the described bioinformatics and mutation or fragmentation libraries, identifies constructs that encode well-expressed, soluble and proteolytically stable protein, regardless of the biochemical functionality of the construct. In a conventional CDH screen, the 'initial-hit' constructs have their sequence verified, and are subsequently analysed to determine if the domain has a stable tertiary structure and whether this is functional. **Chapter 3** describes the development and incorporation of an affinity selection screening methodology, to identify constructs that expressed a specified functional domain of interest;

specifically focussing on protein kinases for 'proof-of-concept' studies. This functionality screen was to be incorporated during a conventional CDH expression screen and would ideally be developed for use for high-throughput screening of cell lysates from CDH fragment protein expression libraries.

An AP (STU-PEG_4-BIOTIN) containing STU was designed and synthesized for the identification of functionally folded protein kinases on an immobilised matrix using the ForteBio Octet. Preliminary screening tests identified the optimal conditions for use of the Octet with the AP, and subsequent studies were successful in identifying specific binding of purified functional protein kinases. However, attempts to extend this methodology for screening in cell lysates containing recombinant protein, proved unsuccessful due to very large background signals that could not be eliminated. This prevented useful discrimination between cell lysates that contained functional protein kinases and those that did not. Subsequently, this methodology was assessed using one-step affinity-tag purified proteins, which had been purified through the traditional high-throughput CDH methodology. This resulted in the successful discrimination between protein kinases and non-protein kinases, with varying recombinant protein expression levels. Whilst this screening methodology could be made to work under selected circumstances using the Octet, it could not be adapted to a streamlined functionality assay that could be used to probe a CDH cell lysate, which would result in the desired reduction of parallel clone handling. Therefore, this screening methodology was not taken any further forward.

8.2 Biochemical and biophysical investigation of human Claspin protein

Chapter 4 used the CDH methodology for the identification of sub-constructs of human Claspin protein that were well expressed, soluble and proteolytically stable. These were subsequently characterized by biochemical and biophysical techniques in **Chapter 5**. The Claspin sequence only has a few bioinformatic annotations, contains no defined or known domains and has little sequence homology to all other known proteins (Kumagai and Dunphy, 2000), making it an ideal test-bed for an empirical method such as CDH. To date, the only structural information

available for human Claspin is a low-resolution EM image, which indicates that Claspin has a ring-shaped structure when bound to branched DNA (see **Figure 1.7**) (Sar et al, 2004). Bioinformatic analyses indicated that Claspin is composed predominantly of α-helices, but with a high degree of predicted structural disorder. Two regions of sequence were predicted to be coiled-coils, there is a predicted HTH motif, and a region in the C-terminus is defined as a repeat motif (CKB). Homology modelling for full-length Claspin did not identify any conserved folds / domains with significant sequence homology, whilst protein threading for both the N- and C-terminus of the protein, identified a mostly α-helical structure but with a high percentage of disordered structure. Furthermore, two existing structural models were identified for Claspin but these had a low sequence identity / homology between Claspin and the template.

CDH was used to create a DNA fragmentation library and these constructs were screened for recombinant protein expression levels, which indicated the presence of two distinct regions, one at the N-terminus of Claspin and the other at the C-terminus. Expression and purification tests found the C-terminal Claspin fragments refractory to purification, whilst N-terminal fragments of Claspin proved to be far more tractable, and these were characterised further. Eight N-terminal Claspin fragments were successfully purified and subsequently investigated using a number of biochemical and biophysical techniques. Biochemical studies showed that the N-terminal fragments of Claspin were elongated and had a high degree of disorder. However, α-helical secondary structure was readily inducible, with the incubation of the protein fragments with the co-solvent TFE. The Claspin protein fragments proved refractory to crystallisation. Furthermore, SAXS and NMR analysis confirmed the proposal that the N-terminus of Claspin has a high degree of intrinsic structural disorder.

Claspin is described as an ‘adaptor’ protein that mediates a number of interactions with other proteins. The unstructured nature of the N-terminus of Claspin and the number of interactions full-length Claspin protein is predicted to mediate, suggests that Claspin could be a protein with intrinsic disorder. This flexibility may provide Claspin with the ability to mediate the multiple protein interactions that it has

been identified to make during DNA replication and replication-coupled repair, thus providing functionality to the protein (Kumagai and Dunphy, 2003, Jeong et al., 2003, Lee et al., 2005, Brondello et al., 2007, Kim et al., 2008, Gold and Dunphy, 2010, Nakaya et al., 2010, Uno and Masai, 2011, Serçin and Kemp, 2011, Rainey et al., 2013, Broderick et al., 2013, Yuan et al., 2014).

8.3 Investigating the interaction of Claspin with DNA

In **Chapter 6**, the eight N-terminal Claspin protein fragments were used to further investigate the DNA binding propensity described in the literature for the N-terminus of Claspin. All protein fragments that bound DNA showed a strong preference for dsDNA over ssDNA, consistent with published observations, which also showed an overall preference for complex branched DNA structures (Sar et al., 2004, Zhao and Russell, 2004, Serçin and Kemp, 2011, Uno and Masai, 2011 and Yilmaz et al., 2011). In this work, the required DNA length for a strong interaction was found to ≥ 16 nucleotides or base-pairs; where increasing lengths of ss- or dsDNA did not increase the K_d for interaction.

Construct analysis identified that the sequence (aa 261-317) covering a predicted HTH motif [aa 279-313 (Zhao and Russell, 2004) or aa 276-342 (SwissModel; Schwede, 2003)] was required for DNA binding functionality. However, for a tight interaction with DNA, a larger region was required (aa 104-413); including the full region of basic patch sequence BP1 (aa 265-331) (Lee et al., 2005) and part of the RFID (aa 265-605) both shown previously to interact with chromatin (Lee et al., 2005), as well as the experimentally identified DBD (aa 149-340) previously shown as interacting with DNA (Sar et al., 2004). These DNA-binding protein fragments showed micromolar affinity for DNA, which correlated with previously calculated binding affinities for full-length Claspin with DNA (Sar et al., 2004). Furthermore, stable complex formation was observed after resolution by ASEC for selected DNA-binding protein fragments containing these sequence regions. The stoichiometry for full-length Claspin with DNA was previously calculated by EM to be a 1:1 (protein:DNA) complex (Sar et al., 2004). However, through the analysis presented here, the stoichiometry of the complex was not able to be determined.

Furthermore, a number of fragments displayed significantly lower binding affinities for DNA, but had significant sequence overlap with these high-affinity DNA-binding protein fragments; these proteins with lower DNA binding affinities had additional sequences C-terminal to the identified DNA binding regions. This could indicate inhibition of binding by these residues; potentially through folding back onto the DNA binding region thereby reducing / inhibiting the DNA binding functionality of these proteins.

Structural characterization of this complex was attempted by protein crystallography, but this did not yield protein crystals. This was likely due to the high degree of inherent structural disorder identified in the Claspin protein constructs.

8.4 Investigating the interaction between Claspin and Chk1

In **Chapter 7**, the interaction between Claspin and Chk1 was further investigated for structural characterization of the complex. Chk1 binds to Claspin in a phosphorylation dependent manner to CKB motifs, in the CKBD, in the C-terminus of Claspin; this enables the activation of Chk1 by ATR (Kumagai and Dunphy, 2003, Lee et al., 2005, Jeong et al., 2003, Clarke and Clarke, 2005, Chini et al., 2006, Lindsey-Boltz et al., 2009).

Recombinant Chk1-KD protein was previously shown to interact with a composite phosphorylated Claspin CKB motif (Clarke and Clarke, 2005), and was here found to interact with the three individual phosphorylated CKB motifs of Claspin; where the tightest interaction was seen with motif 2 (> motif 1 > motif 3) with a micromolar affinity. Furthermore, this interaction was maintained when Chk1-KD inhibited with STU, although there was a slight reduction (two-fold) in the binding affinity for the CKB phosphopeptide. Mutation of the conserved amino acids in the CKB motif prevented interaction with Chk1, indicating sequence recognition or the secondary structure propensity is required in addition the phosphorylated residue for interaction. Furthermore, the activity of active Chk1-KD or active full-length Chk1 was not altered by incubation with the CKB phosphopeptides. Thermal denaturation assays showed the binding of the Claspin CKB phosphopeptide to the

Chk1-KD protein did not further stabilize Chk1-KD protein. Co-crystallography of Chk1-KD with the Claspin CKB phosphopeptides under various conditions was attempted, but these did not yield a co-crystal structure of the Chk1-KD with the CKB phosphopeptide.

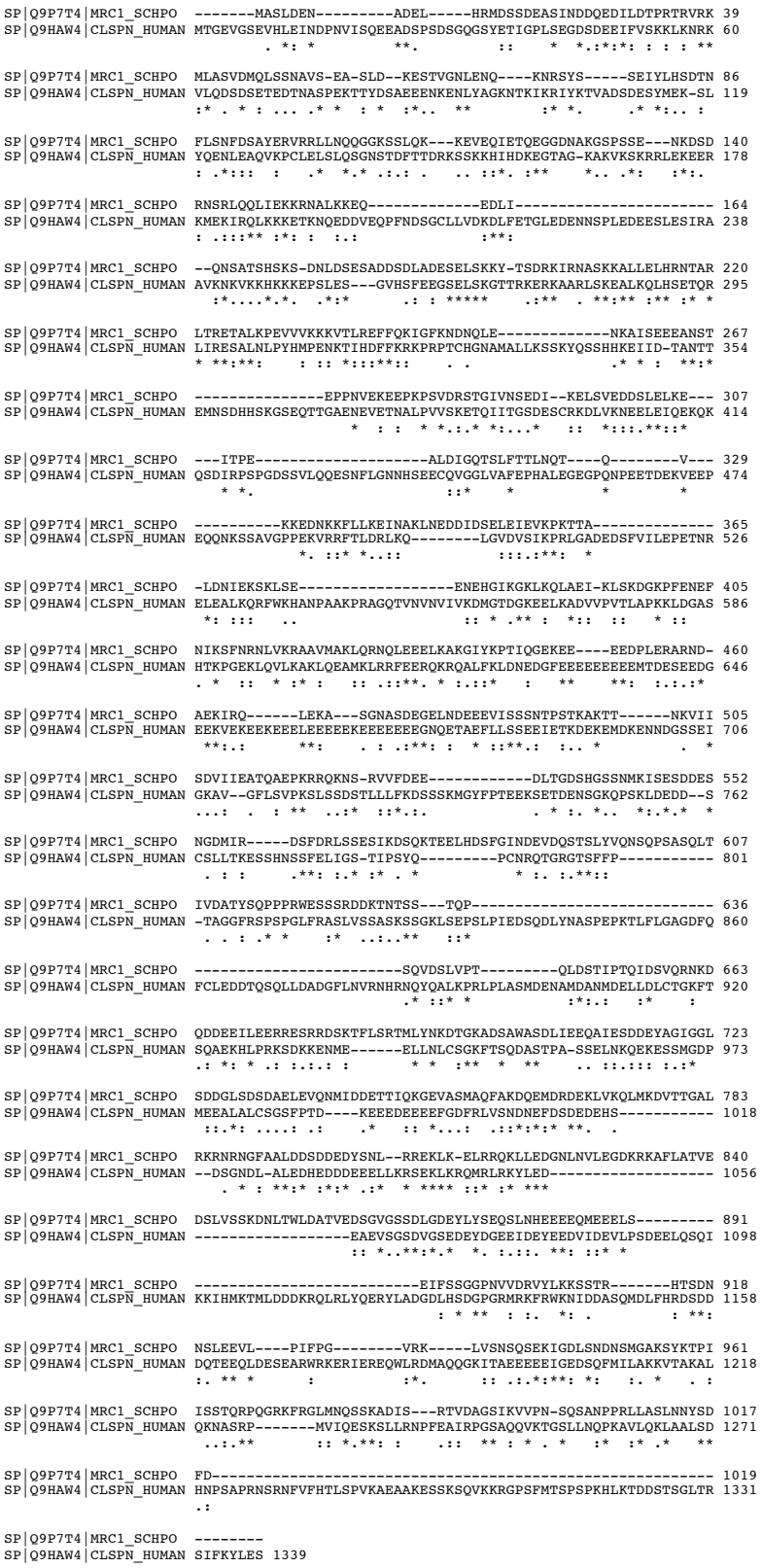
Appendix 1

Supplementary information for Chapter 4

```
SP|Q9HAW4|CLSPN_HUMAN  MTGEVGSEVHLEINDPNVISQEEADSPSDSGQGSYETIGP---------LSEGDSDEEIF 51
SP|Q9DF50|CLSPN_XENLA  MAAL--CEEEQVFLEPEDISLKIVETDSDSGQGSCEMADQNKL---LGCVEDKDTDDEIL 55
TR|D2XSJ5|D2XSJ5_CHICK MA-AAPVELQPEELDVAVAVLKAHGSDSDSGQGSAEPPSPGRPPTGSTSPQDGDSEEEIF 59
SP|Q80YR7|CLSPN_MOUSE  MTGEVGSEVNLEVNDLKLLSQEAADSPVDSGQGSFETLEP---------LSERDSDEEIF 51
                       *:     * .    :      :   :  ****** *              .: *:::**:

SP|Q9HAW4|CLSPN_HUMAN  VSKKLKNRKVLQDSDSETEDTNASPEKT-TYDSAE--EENKENLYAGKNTKIKRIYKTVA 108
SP|Q9DF50|CLSPN_XENLA  VRKKSKKKEVLVDSDSDEELEMRNF--ADNVKGHSDNEENEET-MSAYREKPRKIRSAVL 112
TR|D2XSJ5|D2XSJ5_CHICK VSRKAKGKKVLQDSESEDGEDGDSSVHNDTLGGDTENGEEKEKVTAQ-RNKKSRIRQGLL 118
SP|Q80YR7|CLSPN_MOUSE  VSKKPKSRKVLQDSDSEAEDRDDAPEKP-TYDDSA--EDTQENLHSGKSQ-SRSFPKALA 107
                       * :* * ::** **:*:            .        : :*.  :        : . :

SP|Q9HAW4|CLSPN_HUMAN  DSDESYMEKSLYQENLEAQVKPCLELS-LQSGNSTDFTTDRKSSKKHIHDKEGTAGKAKV 167
SP|Q9DF50|CLSPN_XENLA  DSDNSDHELDVQISTSQNA--AEIPESEHDSLEKETHTVKPKTSKSLKKQ-TDTNKEEIV 169
TR|D2XSJ5|D2XSJ5_CHICK DSDDSDTGDHLQIENLDTSRKSVLSENEVEEGR--PLKSGKKS-RKHKHSFEDEAAEKAV 175
SP|Q80YR7|CLSPN_MOUSE  DSDESDMEETPSQESPETQEAPSLEPG-HQTGHSVDFTTGRKLSKTLLRE--GA--EGKA 162
                       ***:*        .. :      :     :  .    .   *  :.  :.      :  .

SP|Q9HAW4|CLSPN_HUMAN  KSKRRLEKEERKMEKIRQLKKKET-----KNQEDDVEQPFNDSGCLLVDKDLFETGLEDE 222
SP|Q9DF50|CLSPN_XENLA  KNKSK---RKIPKEKIKRRTKQK------SKAVAEARPNLNDSGCLLTDGDLFDNGVENE 220
TR|D2XSJ5|D2XSJ5_CHICK GKPRRRKERERRAESIKQLIKGKKPSSEGQQVDGGEGYPFNDSGCLLDDKELFDNGLEEE 235
SP|Q80YR7|CLSPN_MOUSE  KSKRRLEKEERTMEKIRRLKKKET-----RCEESDADRPLNDSGCLLEDSDLFETGLEEE 217
                        .  :   .:   *.*::  * :               :******* * :**:.*:*:*

SP|Q9HAW4|CLSPN_HUMAN  NNSPLEDEESLESIRAAVKNKVKKHKKKEPSLESGV--HSFEEG--SELSKGTTRKERKA 278
SP|Q9DF50|CLSPN_XENLA  MDSN-EEEDSLEAIRAKMKSKLNSHSAENF--------EDFELDTEGNQESPEKRKERKA 271
TR|D2XSJ5|D2XSJ5_CHICK NNSFPGDEESIESIRAAVKEKIKKYKNKERFSEDEGYKHVFDDDNEESALKEPKRKERKA 295
SP|Q80YR7|CLSPN_MOUSE  NDSALEDEESLESIRAAVKNKVKNRKKKEPTLESEA--FSLEDG--NELSKGSARKERKA 273
                        :*   :*:*:*:*** :*.*::. . ::           ::     .  .   ******

SP|Q9HAW4|CLSPN_HUMAN  ARLSKEALKQLHSETQRLIRESALNLPYHMPENKTIHDFFKRKPRPTCHGNAMALLKSSK 338
SP|Q9DF50|CLSPN_XENLA  ARLGKEAMKQMHSETQRLIRESSVSLPYHLPEPKTIHDFFKRRPRPLCQGNAMQLIKSTK 331
TR|D2XSJ5|D2XSJ5_CHICK ARLSKEAIKQLHSETQRLIRESSVSLPYHMPEAKSVHEFFKRRSRPVYEGNAMALLKSTK 355
SP|Q80YR7|CLSPN_MOUSE  ARLSKEALKKLHSETQRLVRESALNLPYHMPESKTIHDFFKRKPRPTCQGSAMALLKSCK 333
                       ***.***:*::*******:***::.****:** *::*:****: **  .*.** *:** *

SP|Q9HAW4|CLSPN_HUMAN  YQSSHHKEIIDTANTTEMNSDHHSKGSEQTTGAENEVETNALPVVSKETQIITGSDESCR 398
SP|Q9DF50|CLSPN_XENLA  YQPCTEEKKKP---NEEICAEVPEFDYVSKEDLEISPEQPLLNTQCSHAAVLCVVQNDAR 388
TR|D2XSJ5|D2XSJ5_CHICK YEFTLNEEAAGTKTSSTDCKDGPTEGGQ---SAANEPE-----------ANLGGHTDPAA 401
SP|Q80YR7|CLSPN_MOUSE  YQSGHYKETVNPADAAGMGAEDSSRGSEQRTGAGIAAETNVLSEVSEEAGITAGSDEACG 393
                       *:    ::            :                *                  : .

SP|Q9HAW4|CLSPN_HUMAN  KDLVKNEELEIQEKQ-----KQSDIRP---SPGDSSVLQQESNFLGNNHSEECQVGGLVA 450
SP|Q9DF50|CLSPN_XENLA  -----TEGLSKSTEAVVTG-QMN-DHEDA--FSDSNIVHEQETVGLITVTE----TFQTP 435
TR|D2XSJ5|D2XSJ5_CHICK KDPLLGEGENLTEDSAEKSRKNNDDSHSAVTVTTASETEEQQSVLN----TDCSEQKE-S 456
SP|Q80YR7|CLSPN_MOUSE  KDPVRRGELEIEETE-----KHSDDRPY--SPGDRSMSQQESSIPRIEDNEGHQAGDLTE 446
                              .         : .            .  .::...

SP|Q9HAW4|CLSPN_HUMAN  FEPHALEGEGPQNPEETDEKVEEPEQQNK----------SSAVGPPEKVRRFTLDRLKQL 500
SP|Q9DF50|CLSPN_XENLA  FIPQPESVVCEQIQ------------------NDVVEMQRMPEQPTHKPKLSKLEKLKAL 477
TR|D2XSJ5|D2XSJ5_CHICK EIPLPVGGNALEQRDETA-----------PGLENSQQVGPGLAAQPEKVRKSKLDKLREL 505
SP|Q80YR7|CLSPN_MOUSE  SDPPALEGEELKTVEKTDAKEGMPEQKTQSAAAAAVAVVTAAAAPPEKVRRFTVDRLRQL 506
                         *       :                                    .* :  .:::*: *

SP|Q9HAW4|CLSPN_HUMAN  GVDVSIKPRLGADEDSFVIL-EPETNRELEALKQRFWKHANPAAKPRAGQTVNVNVIVKD 559
SP|Q9DF50|CLSPN_XENLA  GVDLSIKPRLCPDDGSFVNLDEPKPNKEFEALKERFLKHTLQKSKPRTERKVNLNIIRKE 537
TR|D2XSJ5|D2XSJ5_CHICK GIDLTIKPRICSGNESFINLDESDSNKELEALKARFLKHTLQTSKPKLERTINMSIIRKE 565
SP|Q80YR7|CLSPN_MOUSE  GVDVSSQPRLGADEDSFVILDEPKTNRELEALKQRFWRHANPAASPRACQTVNVNIIVKD 566
                       *:*:: :**:   : **: * * . *:*:**** ** :*:   :.*:  :.:*:.:* *:

SP|Q9HAW4|CLSPN_HUMAN  MGTDGKEELKADVVPVTLAPKKLDGASHTKPGEKLQVLKAKLQEAMKLRRFEERQKRQAL 619
SP|Q9DF50|CLSPN_XENLA  TTADGKEELKADVVPIVMATEKPDKSIYQKPGEKLQVLKVKLQEAMKIRRSEERLKRQAL 597
TR|D2XSJ5|D2XSJ5_CHICK TTSEGKEELKADVVPAVLAAESLDEAVHTKPGEKLQVLKAKLQEAMKLRRTEERQKRQAL 625
SP|Q80YR7|CLSPN_MOUSE  LGTNGKEELKAEVVPVTLAAEKLEGASHAKPGEKLQMLKAKLQEAMKLRRLEERQKRQAL 626
                         ::*******:*** .:* :. : : : *******:**.*******:** *** *****

SP|Q9HAW4|CLSPN_HUMAN  FKLDNEDGFEEEEEE--EEEMTDESEEDGEEKVEKEEKEEELEEEEKEEEEEEEGNQET 677
SP|Q9DF50|CLSPN_XENLA  YKLDNEDGFEDDEEE---EEMTEESEDDGD------------------------GNAET 629
TR|D2XSJ5|D2XSJ5_CHICK FKLDNEELLEEEEEEEEEEEMTDESEEEEE-------------------------GDHEA 660
SP|Q80YR7|CLSPN_MOUSE  FKLDNEDGFEEEEEE---EEMTDESE--------------------------EDGEEET 656
                       :*****: :*::***   ****:***                            *: *:
```

```
SP|Q9HAW4|CLSPN_HUMAN  AEFLLSSEEIET---------KDEKEMDKENNDGSSEIG-KAVGFLSVPKSLSSDSTLLL 727
SP|Q9DF50|CLSPN_XENLA  ADYPGGEDEEEVGDAEDD-----NDEDDTVN---DRLLGNVPEIVIPLPRPVTTDSSLML 681
TR|D2XSJ5|D2XSJ5_CHICK VESLLDEAEEDNEDLEEKQVEDGDKETDRESID-GEKVE-QAVDCASVPKPPSTESTLML 718
SP|Q80YR7|CLSPN_MOUSE  TEYLLGSEDTET---------KDEKETDKENTDTSSDIG-KSVA-LCVPKPLSSDSTLLL 705
                       .:    . : :            :.* *  .      :         :*:  :::*:*:*

SP|Q9HAW4|CLSPN_HUMAN  FKDSSSKMGYFPTEEKSETDENSGKQPSKLDEDDSCSLL--TKESSHNSSFELIGSTIPS 785
SP|Q9DF50|CLSPN_XENLA  FKDNSSKLGDSLPDESGC------KRSSRLEYE-EDSLLPQLKENSHNSSFELISSMIPS 734
TR|D2XSJ5|D2XSJ5_CHICK FKDSSSKMGYSLPDEKHESEEAANKEATKLEDDDSFSLPTPAKENSHNSSFELIGSMIPS 778
SP|Q80YR7|CLSPN_MOUSE  FKDSSSKMGYFPTEEKSETDEYLAKQSDKLDEDDSSSLL--TKESSHNSSFELIGSTIPS 763
                       ***.***:*    :*.        *.  :*: : . **    **.*********.* ***

SP|Q9HAW4|CLSPN_HUMAN  YQPCNRQTGRGTSFFPTAGGFRSPSPGLFRASLVSSASKSSGKLSEPSLPIEDSQDLYNA 845
SP|Q9DF50|CLSPN_XENLA  YQPCNKTTRVV--INSNNLGFRSPSPVHFKTSFLSSASKSSGKMSEPSLPVEDSQDLYNA 792
TR|D2XSJ5|D2XSJ5_CHICK YQPCNKQMSRGGNFLPAAGGFRSPSPGFFKTSFISSASKSSGRMSEPSLPIEDSQDLYNA 838
SP|Q80YR7|CLSPN_MOUSE  YQPCNRQIGRGASFLPTA-GFRSPSPGLFRGSLISSASKSSGKLSEPSLPVEDSQDLYTA 822
                       *****:       :     *******  *: *::********::******:*******.*

SP|Q9HAW4|CLSPN_HUMAN  SPEPKTLFLGAG-D--FQFCLEDDTQSQLLDADGFLNVRNHRNQYQALKPRLPLASMDEN 902
SP|Q9DF50|CLSPN_XENLA  SPEPKASYLCAGRNSQFQFSLEDDTQSQLLDADGFLNVGRHKS--SSAKHRLALDTMDEN 850
TR|D2XSJ5|D2XSJ5_CHICK SPEPKSLFPGAG-ESQFQFSLEDDTQSQLLDADGFLNVGQHRNKYQSSKHGLTLASMDEN 897
SP|Q80YR7|CLSPN_MOUSE  SPEPKTLFLGAG-D--FQFCLEDDTQSQLLDADGFLNIRNHRHRYQAVKPQLPLASMDEN 879
                       *****: :  ** :  ***.*****************: .*:   .: *  * * :****

SP|Q9HAW4|CLSPN_HUMAN  AMDANMDELLDLCTGKFTSQAE-KHL-PRKSDKKENMEELLNLCSGKFTSQDASTPASSE 960
SP|Q9DF50|CLSPN_XENLA  AMDANMDELLDLCSGQFKESLSGTSQAAESDAKKQPMDELLELCSGKFVSQADCSTQDSS 910
TR|D2XSJ5|D2XSJ5_CHICK AMDANMDELLDLCSGQFSSQAEH-VP-STSSNKKQNMEELLNLCSGKFVTQNSPTWASSV 955
SP|Q80YR7|CLSPN_MOUSE  AMDANMDELLDLCTGQFTSQPEEKCQ-PRKNDKKENMEELLNLCSGKFPTQDASPVAPLG 938
                       *************:*:*... .       .. **: *:***:****** :*

SP|Q9HAW4|CLSPN_HUMAN  LNKQEKESSMGDPMEEALALCSGSFPTDKEEEDE----EEEFGDFRLVSNDNEFDSDEDE 1016
SP|Q9DF50|CLSPN_XENLA  ASAKDRSTAVKKDISDEVATVSSSFLTEREQEED---EEEEFGEFKLLPCDDSESENEEQ 967
TR|D2XSJ5|D2XSJ5_CHICK SSKAEKDSDIEDPMAEALALCSGSFPTDREEEDEEHEEHEELGDFQLVTDDNAFDSEEDE 1015
SP|Q80YR7|CLSPN_MOUSE  LRSQEKESSTEDPMEEALALCSGSFPTDREEEGE----EEEFGDFQLVSKENGFASDEDE 994
                           ::.:   . : : :*  *.** *::*:* :    .**:*:*:*:  ::   .:*::

SP|Q9HAW4|CLSPN_HUMAN  HSDSGNDLALEDHEDDDEEELLKRSEKLKRQMRLRKYLEDEAEVSGSDVGSEDEYDGEEI 1076
SP|Q9DF50|CLSPN_XENLA  NEEEEE--EEDAKDDEDEEEILQ--KQQKRKLRLNDFMEDEAELSGSDVGSGDEYEGDD- 1022
TR|D2XSJ5|D2XSJ5_CHICK KSGDSDGEE---AEVSDEEELLRRREGSKKKLKLEDFMEDEAELSGSDVGSEDEYDGEDL 1072
SP|Q80YR7|CLSPN_MOUSE  HSDSNDEELA-LDLEDDEEELLKQSEKMKRQMRLKKYLEDEAEVSGSDVGSEDEYDGEEI 1053
                       .. . :         .****:*:  :  *::::*..::*****:******* ***:*::

SP|Q9HAW4|CLSPN_HUMAN  DEYEEDVIDEVLPSDEELQSQIKKIHMKTMLDDDKRQLRLYQERYLADGDLHSDGPGRMR 1136
SP|Q9DF50|CLSPN_XENLA  DEYEEEAIDEDLPSDEELQDQVNKIHMKVTMDEDQRQLRFYQERYLADGDLHSDGPGRTR 1082
TR|D2XSJ5|D2XSJ5_CHICK NEYEEEIIDEELPNEAELESQVQKLHMKAVLDDDKRQLRLYQERYLIDGDLHSDGPGRMR 1132
SP|Q80YR7|CLSPN_MOUSE  DEYEEDVIDEVLPSDEELESQIKKIHMKTMLDDDKRRLRLYQERYLADGDLHSDGPGRTR 1113
                       :****: *** **.: **:.*::*:***. :*:*:*:**:****** ********** *

SP|Q9HAW4|CLSPN_HUMAN  KFRWKNIDDASQMDLFHRDSDDDQ---TEEQLDESEARWRKERIEREQWLRDMAQQGKIT 1193
SP|Q9DF50|CLSPN_XENLA  KFRWKHLDDASQVDMFRRDSELEEVDGENEETEETELKWRKERFEREQWLREQPQGSRDN 1142
TR|D2XSJ5|D2XSJ5_CHICK RFRWKNIDDASQMDLFQRDSDNED---ENESFDETEVKWRKERFEREQWLREQKEKNKEQ 1189
SP|Q80YR7|CLSPN_MOUSE  KFRWKHIDDTSQMDLFHRDSDDDQ---VEEQLDETEAKWRKERIEREQWLREQAQQGKIA 1170
                       :****.:**:**:*:*:***: ::    :*. :*:* :*****:*******:  :  :

SP|Q9HAW4|CLSPN_HUMAN  -AEEEEEIGEDSQFMILAKKVTAKALQKNASRPMVIQESKSLLRNPFEAIRPGSAQQVKT 1252
SP|Q9DF50|CLSPN_XENLA  NEEEEEDIGEDSQFMKLAKKVTAKALQRKVSTET-NEPKKPGPRNPYEVIRPFSLPKLRT 1201
TR|D2XSJ5|D2XSJ5_CHICK EEEEEEDIGEDSQFMKLAKKVTEKSLQKRASPAVVVQDAKLLPRNPFETFRPASDIQIKN 1249
SP|Q80YR7|CLSPN_MOUSE  -AD-EEDIGDDSQFMMLAKKVTAKALQKNASHTVVVQESKSVLRNPFETIRPGGAHQLKT 1228
                         : **:**:***** ****** *:**:..*     :  *   ***:*.:** .  :::.

SP|Q9HAW4|CLSPN_HUMAN  GSLLNQPKAVLQKLAALSDHNPSAPRNSRNFVFHTLSPVKAEAAKESSKSQVKKRGPSFM 1312
SP|Q9DF50|CLSPN_XENLA  GSLLSKPKEVLQKLAAVSDLNPNAPRNSRNFVFQTVSPGKKEETTDKPRSKVRKNI--AV 1259
TR|D2XSJ5|D2XSJ5_CHICK GSLLNRPKDILQKLAAMSDLNPNAPRNSRNFVFHTLSPDKNEEAKEKSKHQVKKRGPSAA 1309
SP|Q80YR7|CLSPN_MOUSE  GSLLNQPKAVLQKLAALSDLNPSAPRNSRNFVFHTLSPTKAEAAKDSSKPQVRRRGLSSM 1288
                       ****.:** :******:** **.**********:*:** * * :.:. : :*::.

SP|Q9HAW4|CLSPN_HUMAN  TSPSPKHLKTDD-STSGLTRSIFKYLES 1339
SP|Q9DF50|CLSPN_XENLA  AMPSPKRFKRDS-TPTVKSRSIFQLLE- 1285
TR|D2XSJ5|D2XSJ5_CHICK ITSLAKRPRVDSTEQTSQSRSIFQYLES 1337
SP|Q80YR7|CLSPN_MOUSE  MSPSPKRLKTNG-SSPGPKRSIFRYLES 1315
                           *: : :       .****: **
```

Appendix 1.1 Multiple amino acid sequence alignment.
Sequence alignment for human (Uniprot ID: Q9HAW4), *Xenopus* (Q9DF50), chicken (D2XSJ5) and mouse (Q80YR7) Claspin (Consortium, 2015). The alignment was prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010).

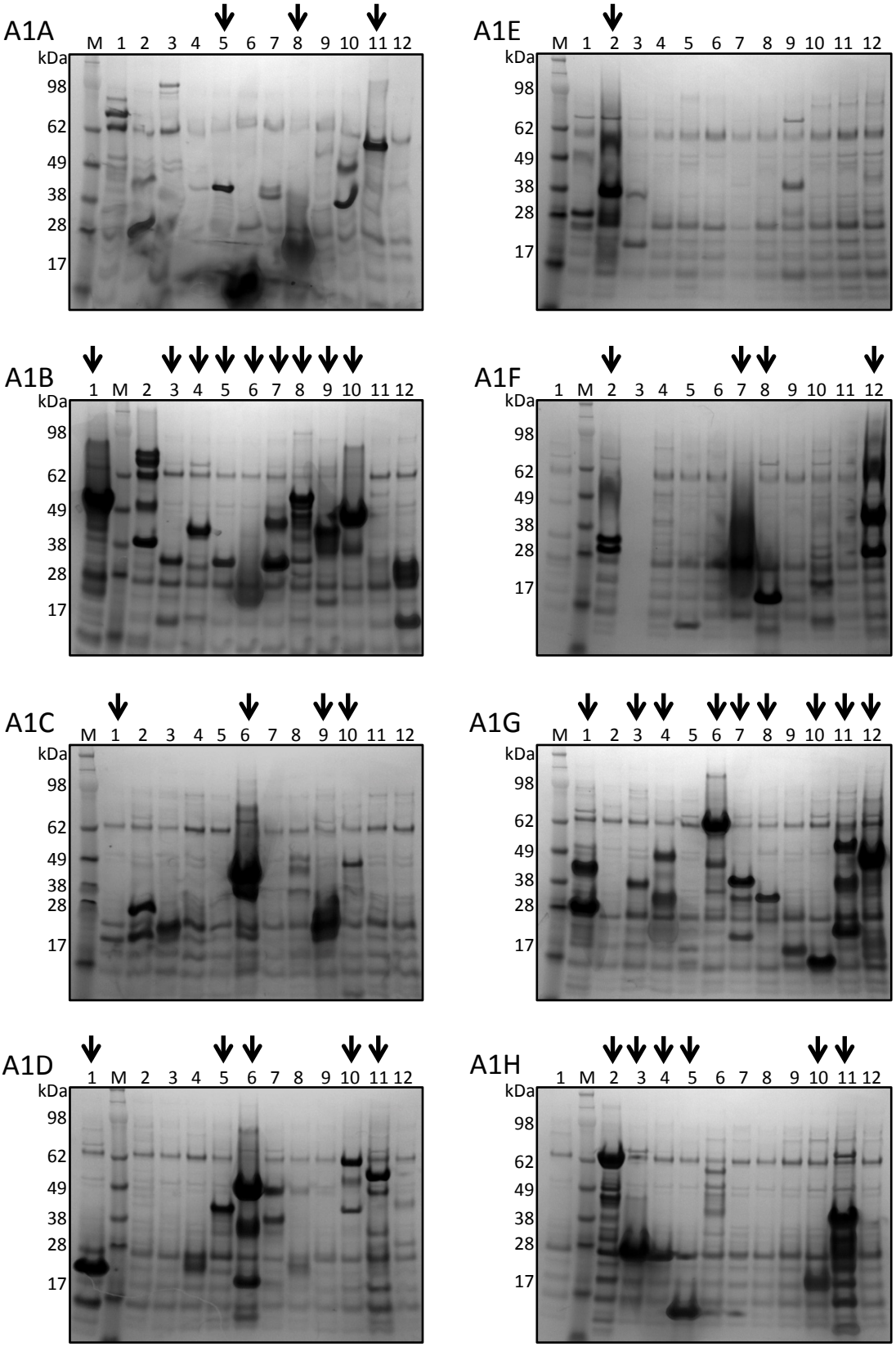


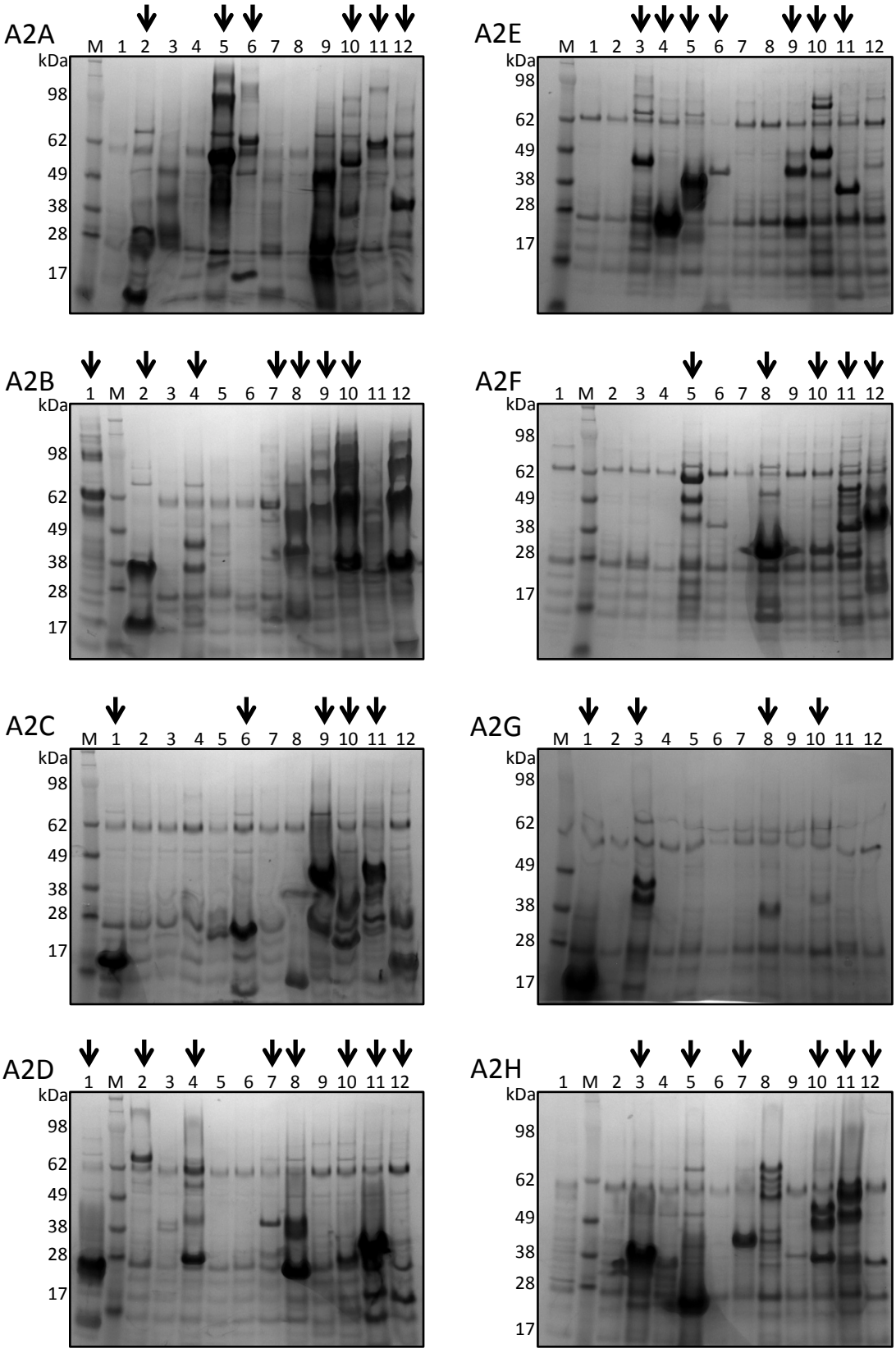
Appendix 1.2 Multiple amino acid sequence alignment of *S. pombe* Mrc1 and human Claspin.

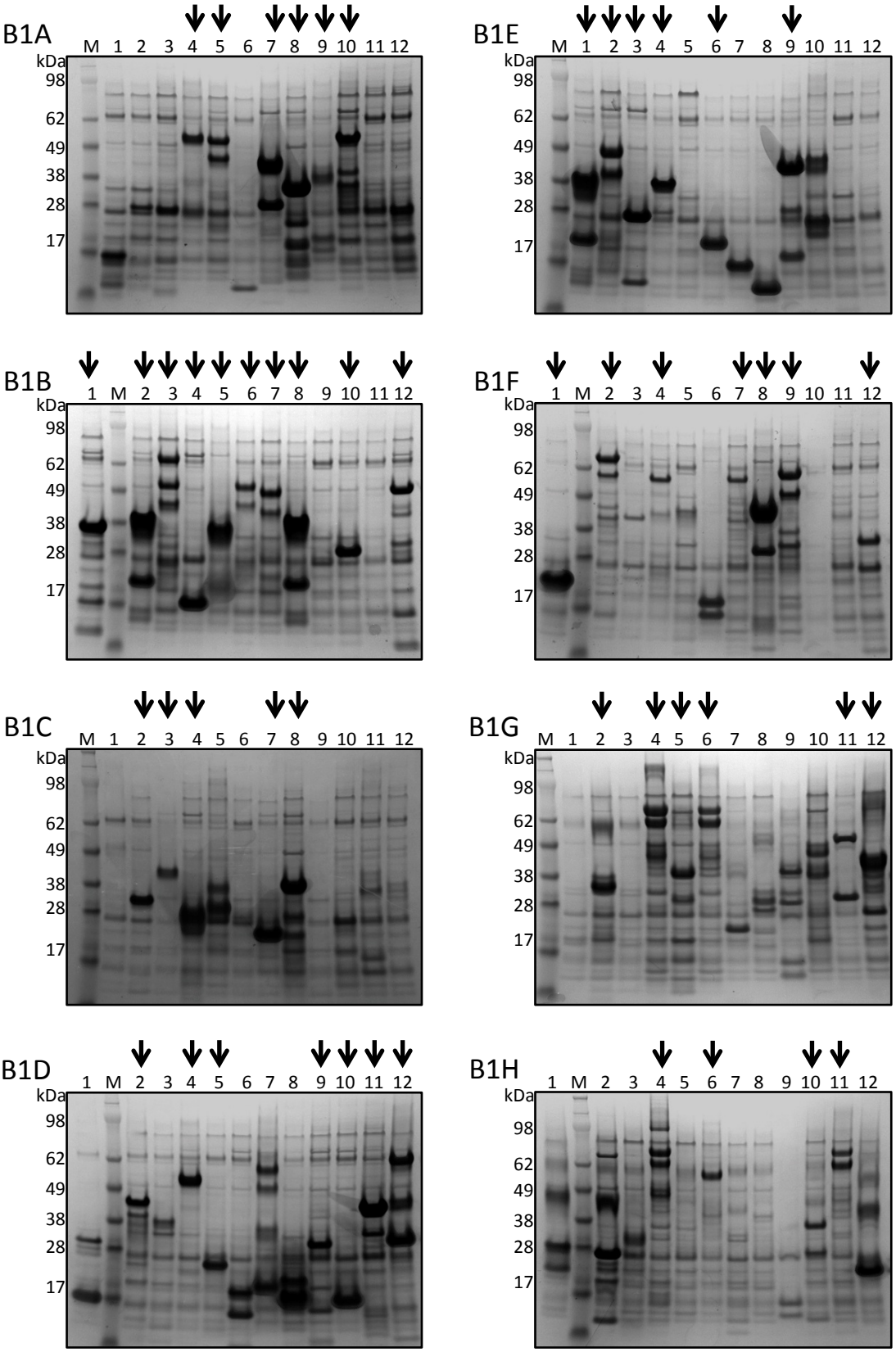
Sequence alignment for *S. pombe* Mrc1 (Uniprot ID: Q9P7T4) and human Claspin (Q9HAW4) (Consortium, 2015). The alignment was prepared using Clustal Omega (Sievers et al., 2011, Goujon et al., 2010).

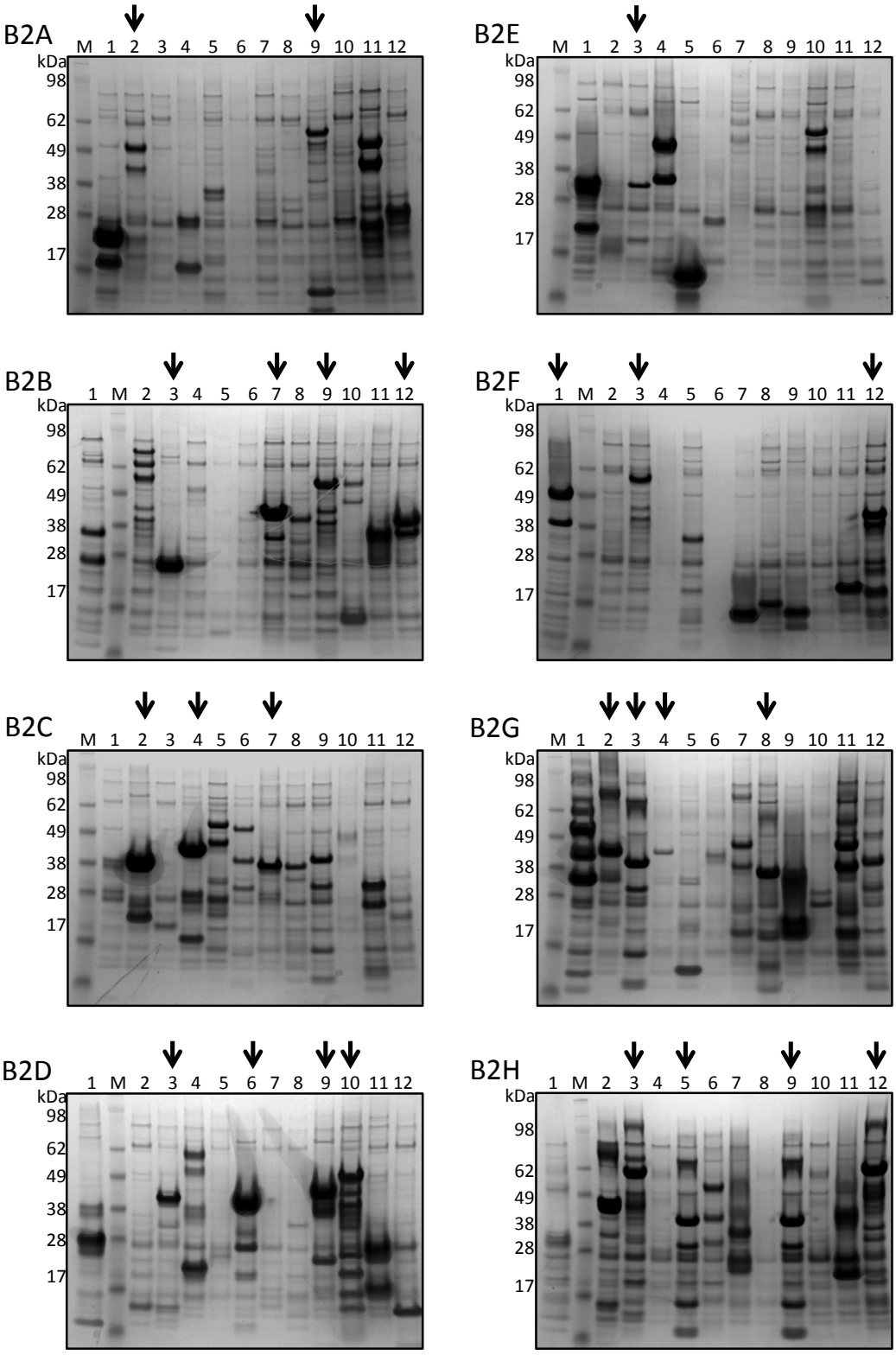
```
CATATGACGG GTGAGGTGGG TTCTGAAGTA CACCTGGAGA TCAACGATCC GAACGTAATC        60
TCTCAGGAAG AAGCGGACAG CCCGTCCGAC AGCGGTCAAG GTTCCTACGA AACTATCGGT       120
CCGCTGAGCG AGGGTGATTC AGATGAAGAA ATTTTCGTCA GCAAAAAGCT GAAAAACCGC       180
AAAGTGCTGC AAGATTCCGA TAGCGAAACG GAAGATACTA ACGCCAGCCC TGAAAAGACC       240
ACCTATGATA GTGCTGAAGA AGAAAACAAA GAAAACTTAT ATGCTGGTAA AAACACGAAA       300
ATCAAACGTA TCTACAAAAC TGTAGCTGAC TCTGATGAAT CTTACATGGA GAAGTCGCTG       360
TATCAGGAAA ATCTGGAAGC ACAGGTTAAA CCGTGTCTGG AACTGAGCCT GCAGTCCGGC       420
AATTCAACGG ATTTTACCAC CGATCGCAAG TCCTCCAAAA AGCATATTCA CGATAAAGAG       480
GGCACCGCAG GCAAAGCGAA AGTTAAGAGC AAGCGTCGCC TGGAAAAAGA AGAGCGCAAA       540
ATGGAAAAGA TCCGCCAGTT AAAGAAGAAA GAAACGAAAA ATCAGGAGGA TGATGTTGAA       600
CAGCCGTTTA ACGATAGCGG CTGCCTGCTG GTAGATAAAG ACCTCTTTGA GACTGGCCTG       660
GAAGATGAAA ACAACAGCCC TCTGGAAGAC GAAGAATCTT TAGAATCCAT TCGTGCCGCA       720
GTTAAGAATA AAGTCAAGAA ACACAAGAAA AAGGAACCAT CTCTGGAATC TGGTGTTCAC       780
TCCTTCGAAG AAGGCAGCGA ACTGAGCAAA GGCACGACCC GCAAAGAACG TAAAGCAGCA       840
CGTTTATCTA AAGAAGCTCT CAAGCAGCTG CATTCCGAGA CCCAACGCCT GATTCGCGAA       900
TCTGCTCTGA ACCTGCCGTA TCACATGCCG GAGAACAAAA CCATCCACGA CTTCTTTAAA       960
CGCAAGCCAC GTCCGACCTG TCATGGTAAC GCGATGGCCT TACTCAAAAG CAGCAAGTAC      1020
CAGTCGTCGC ACCATAAAGA AATTATCGAT ACTGCGAATA CCACGGAGAT GAATAGCGAT      1080
CATCACAGCA AGGGTAGCGA GCAGACCACC GGCGCCGAAA ATGAAGTCGA AACTAACGCA      1140
TTACCAGTCG TGTCTAAAGA GACCCAAATT ATCACGGGCA GCGATGAAAG CTGCCGCAAG      1200
GACCTGGTAA AGAACGAGGA GCTGGAAATT CAGGAAAAAC AGAAACAGTC CGACATTCGT      1260
CCGAGCCCAG GCGATAGCAG CGTTTTACAG CAGGAGTCTA ACTTTCTCGG TAACAATCAC      1320
AGCGAGGAAT GCCAGGTAGG TGGCCTGGTG GCATTTGAAC CGCACGCATT AGAGGGTGAA      1380
GGCCCTCAAA ACCCGGAAGA AACCGATGAA AAGGTTGAGG AACCTGAGCA GCAAAATAAA      1440
TCTAGCGCAG TTGGTCCACC GGAAAAAGTG CGTCGCTTCA CGCTGGATCG TCTGAAACAG      1500
CTGGGCGTGG ACGTATCAAT CAAACCGCGT CTGGGTGCGG ATGAAGATAG CTTCGTTATT      1560
CTCGAACCGG AAACCAACCG CGAACTGGAG GCGTTAAAGC AACGTTTTTG GAAGCACGCA      1620
AACCCTGCAG CCAAACCGCG TGCTGGTCAA ACCGTCAACG TAAACGTCAT TGTTAAAGAC      1680
ATGGGTACCG ATGGCAAGGA AGAACTGAAA GCGGATGTAG TCCCTGTGAC CTTAGCTCCG      1740
AAAAAGCTCG ACGGTGCATC TCATACGAAA CCGGGTGAAA AACTGCAGGT GCTCAAAGCA      1800
AAACTGCAGG AGGCCATGAA ACTGCGTCGT TTTGAAGAAC GTCAGAAGCG TCAGGCGCTG      1860
TTCAAATTAG ATAACGAGGA CGGCTTCGAG GAAGAAGAAG AAGAGGAAGA GGAAATGACC      1920
GATGAGTCTG AAGAAGACGG TGAAGAGAAA GTTGAAAAAG AAGAAAAGGA AGAAGAATTA      1980
GAGGAGGAGG AAGAAAAGGA GGAGGAGGAA GAAGAAGAGG GCAATCAAGA AACCGCTGAG      2040
TTCCTGCTCA GCTCCGAAGA AATCGAGACC AAAGATGAAA AAGAGATGGA TAAAGAAAAC      2100
AACGACGGTA GCTCAGAAAT TGGTAAAGCA GTAGGTTTTC TGAGCGTTCC GAAGTCTCTG      2160
AGCTCGGACT CCACGCTGCT GCTCTTCAAG GACTCTAGCA GCAAAATGGG TTACTTTCCG      2220
ACGGAAGAAA AATCTGAGAC TGATGAAAAT TCAGGTAAAC AGCCGTCTAA ACTGGACGAA      2280
GACGACTCTT GCTCCCTGCT GACGAAAGAA TCCAGCCACA ATTCCAGCTT CGAGCTGATC      2340
GGCTCCACCA TCCCTAGCTA CCAGCCGTGT AACCGTCAAA CGGGTCGTGG CACGTCATTC      2400
TTCCCAACTG CGGGCGGCTT CCGCAGCCCG TCTCCTGGCC TGTTTCGCGC TTCCCTGGTT      2460
AGCAGCGCAA GCAAATCCTC CGGTAAACTG AGCGAACCGT CCCTGCCGAT CGAAGATTCT      2520
CAGGATTTAT ATAACGCGAG CCCAGAACCT AAAACCCTGT TCCTCGGTGC GGGTGACTTC      2580
CAGTTTTGCT TAGAAGATGA CACCCAGAGC CAACTGCTGG ACGCAGATGG CTTCCTCAAT      2640
GTGCGCAATC ATCGCAACCA ATATCAGGCT TTAAAACCGC GTCTGCCGCT GGCTAGCATG      2700
GACGAGAACG CTATGGATGC GAACATGGAC GAACTGCTGG ATCTGTGTAC GGGCAAATTC      2760
ACCTCTCAGG CAGAGAAACA CCTGCCTCGT AAATCCGATA AGAAAGAGAA CATGGAAGAG      2820
TTACTGAATC TGTGCTCTGG CAAGTTCACT TCTCAGGATG CTTCTACGCC AGCTTCGTCT      2880
GAGCTGAATA AACAAGAAAA AGAAAGCTCT ATGGGCGACC CGATGGAAGA AGCTCTGGCA      2940
TTATGTAGCG GCTCGTTCCC TACCGACAAG GAAGAGGAAG ATGAGGAAGA AGAGTTCGGT      3000
GATTTTCGCC TGGTAAGCAA TGATAACGAG TTCGATTCTG ACGAAGACGA ACATTCTGAC      3060
TCCGGTAACG ACCTGGCCTT AGAAGATCAC GAGGACGATG ATGAAGAGGA ACTGCTTAAA      3120
CGTTCTGAAA AGCTGAAACG TCAGATGCGC CTGCGTAAAT ATCTGGAAGA CGAAGCGGAA      3180
GTCAGCGGTT CTGATGTGGG CTCAGAAGAT GAATATGACG GTGAGGAAAT TGACGAATAC      3240
GAAGAAGACG TAATTGATGA AGTCCTGCCG TCTGATGAGG AGCTGCAATC TCAGATTAAG      3300
AAAATTCACA TGAAAACCAT GCTGGATGAC GATAAGCGCC AGCTGCGCTT ATATCAAGAG      3360
CGTTATCTGG CGGATGGTGA CCTCCACTCC GACGGTCCAG GCCGTATGCG TAAATTTCGC      3420
TGGAAAAACA TCGATGACGC TAGCCAGATG GATCTGTTCC ATCGCGACTC CGATGATGAT      3480
CAGACGGAAG AACAATTAGA TGAAAGCGAG GCACGTTGGC GTAAGGAACG TATTGAGCGC      3540
GAACAGTGGC TGCGTGATAT GGCTCAGCAA GGTAAGATTA CGGCTGAAGA GGAGGAAGAG      3600
ATCGGTGAAG ATTCCCAGTT CATGATTTTA GCGAAGAAAG TGACCGCGAA GGCGCTGCAG      3660
AAAAATGCTT CGCGTCCAAT GGTGATTCAA GAATCCAAAT CTCTGCTGCG TAACCCTTTC      3720
GAAGCGATCC GTCCGGGTAG CGCTCAACAG GTCAAGACCG GTAGCCTGTT AAATCAACCG      3780
AAAGCTGTTC TCCAGAAACT GGCTGCACTG TCGGACCACA ACCCATCGGC TCCACGCAAC      3840
TCCCGTAACT TTGTATTCCA CACGCTGAGC CCAGTCAAAG CAGAGGCTGC GAAAGAATCC      3900
TCAAAATCTC AAGTCAAAAA GCGTGGTCCG TCGTTCATGA CCAGCCCGTC CCCAAAACAC      3960
CTGAAAACCG ATGACTCTAC CTCCGGTCTG ACCCGTAGCA TTTTCAAGTA CCTCGAATCC      4020
CACCATCACC ATCACCATTA ATAAAAGCTT                                       4050
```

Appendix 1.3 DNA sequence of the synthetic human Claspin gene.

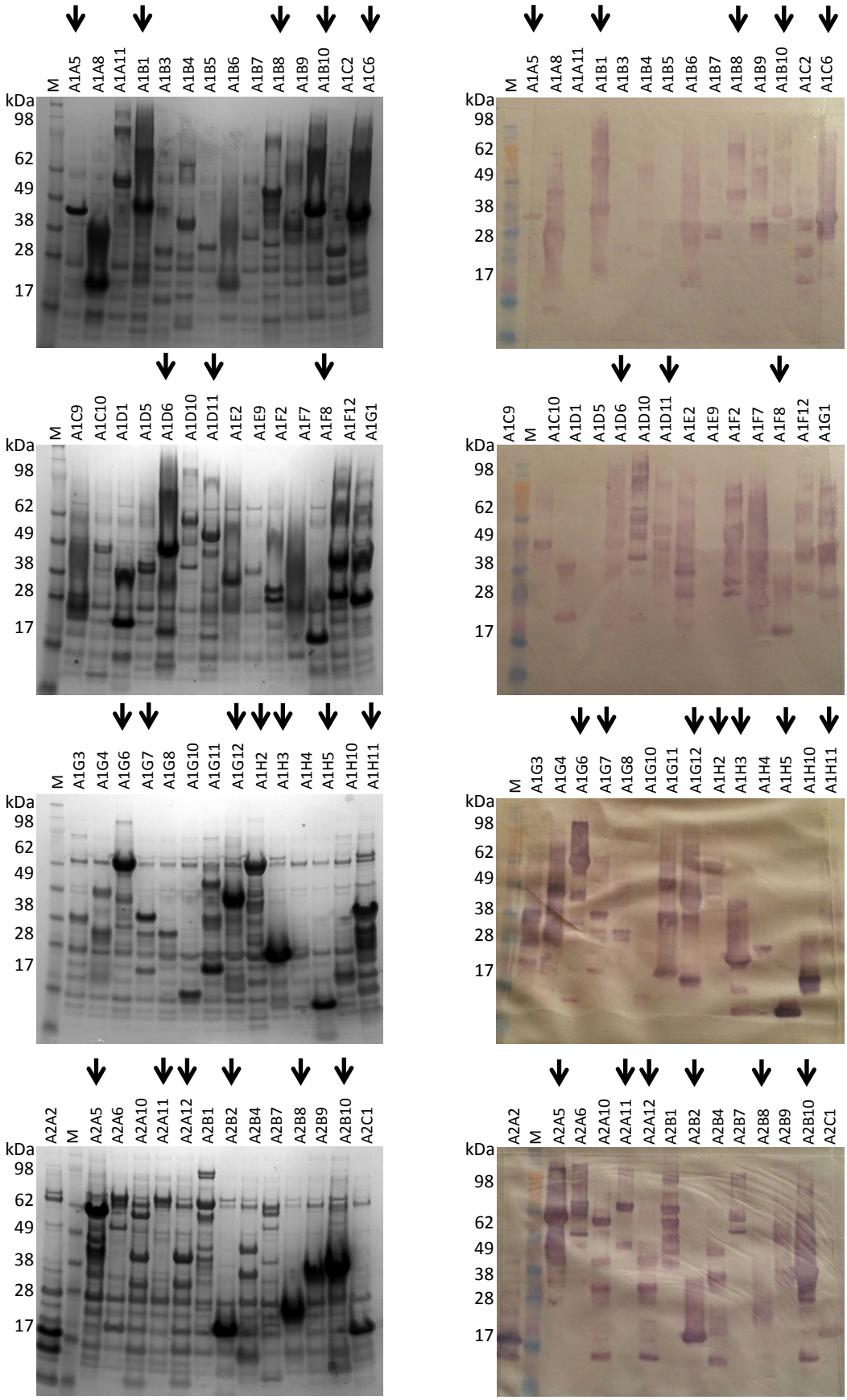

A1A
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1E
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1B
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1F
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1C
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1G
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1D
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A1H
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17

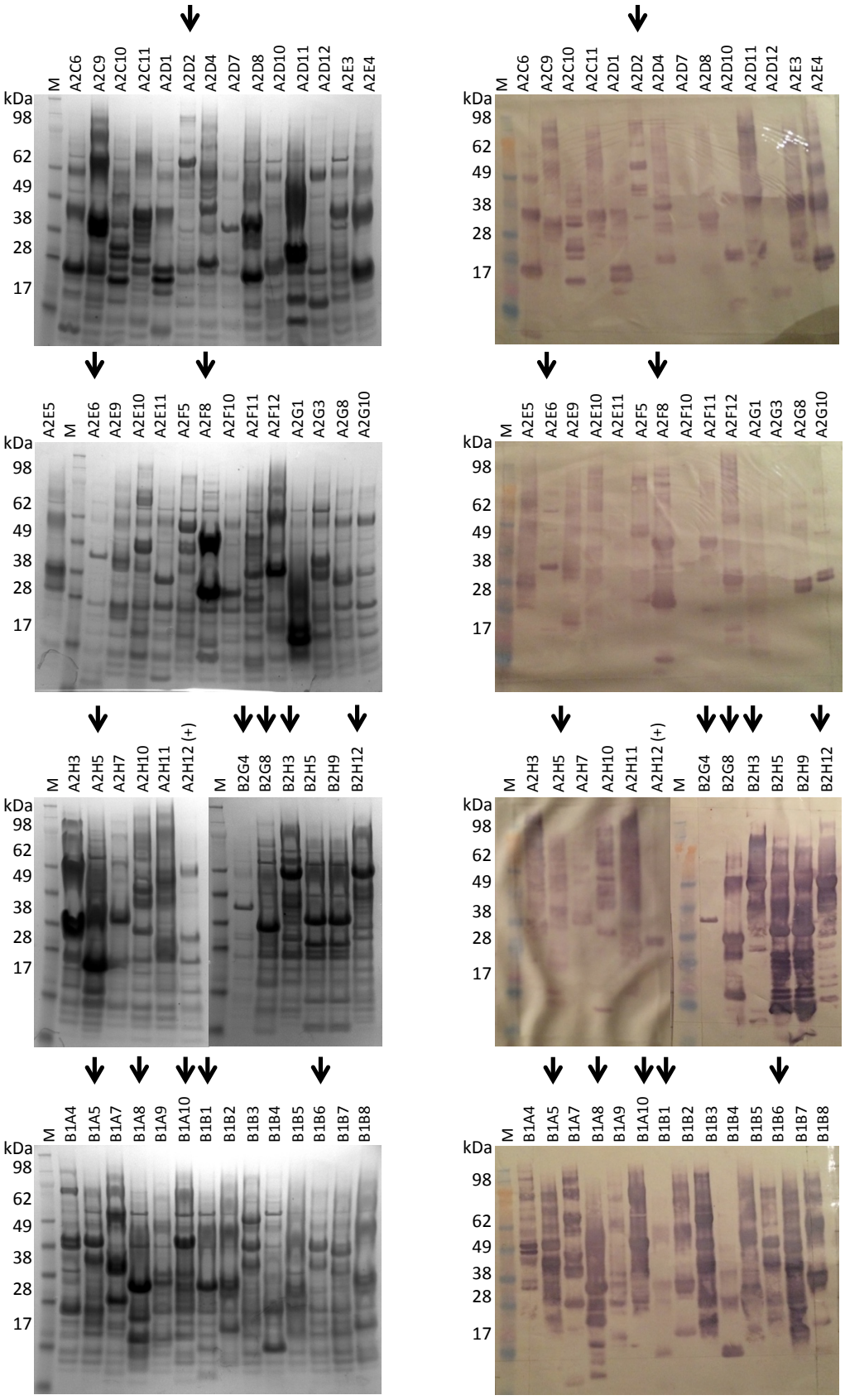

A2A
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2E
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2B
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2F
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2C
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2G
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2D
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
A2H
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17

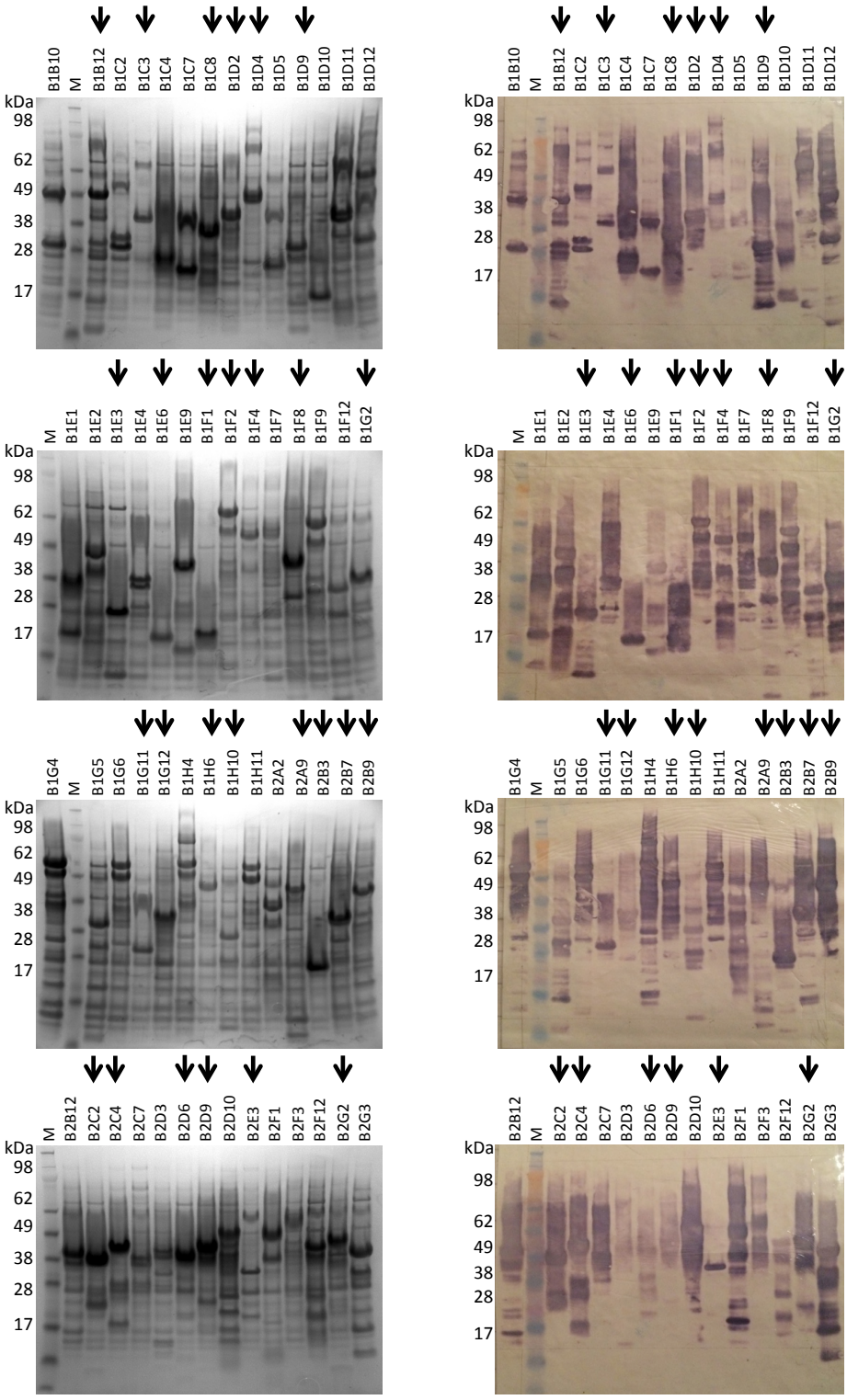

B1A
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1E
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1B
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1F
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1C
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1G
M 1 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1D
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17
B1H
1 M 2 3 4 5 6 7 8 9 10 11 12
kDa
98
62
49
38
28
17



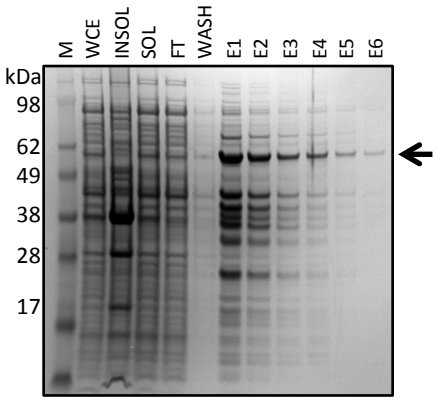
Appendix 1.4 SDS-PAGE analysis of IMAC eluates.
The eluates from the one-step IMAC purification step of CDH were analysed by SDS-PAGE (Instant Blue-stained, 4-12% Bis-Tris SDS-PAGE gels). Gels are labelled in the top left corner with respect to plate ID. Labelling of 1 to 12 indicates the clone ID; forming the unique four digit code. Samples taken forward to the solubility screen step are indicated by an arrow. M=molecular mass marker.


M A1A5 A1A8 A1A11 A1B1 A1B3 A1B4 A1B5 A1B6 A1B7 A1B8 A1B9 A1B10 A1C2 A1C6
kDa 98 62 49 38 28 17
M A1C9 A1C10 A1D1 A1D5 A1D6 A1D10 A1D11 A1E2 A1E9 A1F2 A1F7 A1F8 A1F12 A1G1
kDa 98 62 49 38 28 17
M A1G3 A1G4 A1G6 A1G7 A1G8 A1G10 A1G11 A1G12 A1H2 A1H3 A1H4 A1H5 A1H10 A1H11
kDa 98 62 49 38 28 17
A2A2 M A2A5 A2A6 A2A10 A2A11 A2A12 A2B1 A2B2 A2B4 A2B7 A2B8 A2B9 A2B10 A2C1
kDa 98 62 49 38 28 17

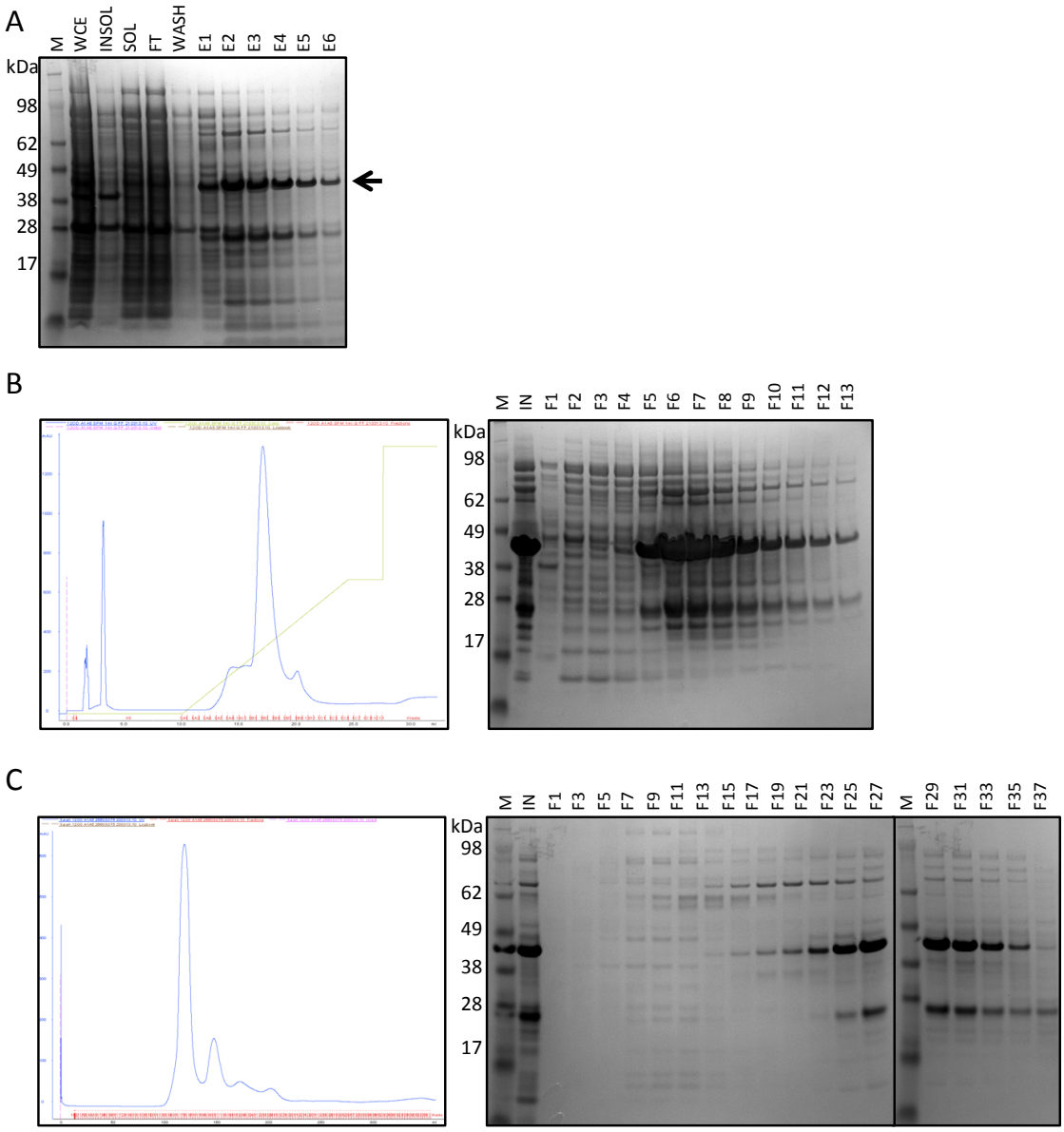

M A2C6 A2C9 A2C10 A2C11 A2D1 A2D2 A2D4 A2D7 A2D8 A2D10 A2D11 A2D12 A2E3 A2E4
kDa 98 62 49 38 28 17
A2E5 M A2E6 A2E9 A2E10 A2E11 A2F5 A2F8 A2F10 A2F11 A2F12 A2G1 A2G3 A2G8 A2G10
M A2H3 A2H5 A2H7 A2H10 A2H11 A2H12 (+) M B2G4 B2G8 B2H3 B2H5 B2H9 B2H12
M B1A4 B1A5 B1A7 B1A8 B1A9 B1A10 B1B1 B1B2 B1B3 B1B4 B1B5 B1B6 B1B7 B1B8



Appendix 1.5 CDH solubility screening.
Selected eluates from the one-step IMAC purification step of CDH were centrifuged and the soluble supernatant was re-analysed by (left) SDS-PAGE (Instant Blue-stained, 4-12% Bis-Tris SDS-PAGE gels) and by (right) anti-His western blot (using an anti-His primary antibody, and an AP-conjugated secondary antibody with colourmetric detection). Samples are labelled above their respective gel lane. Samples taken forward for DNA sequencing are indicated by an arrow. M=molecular mass marker.

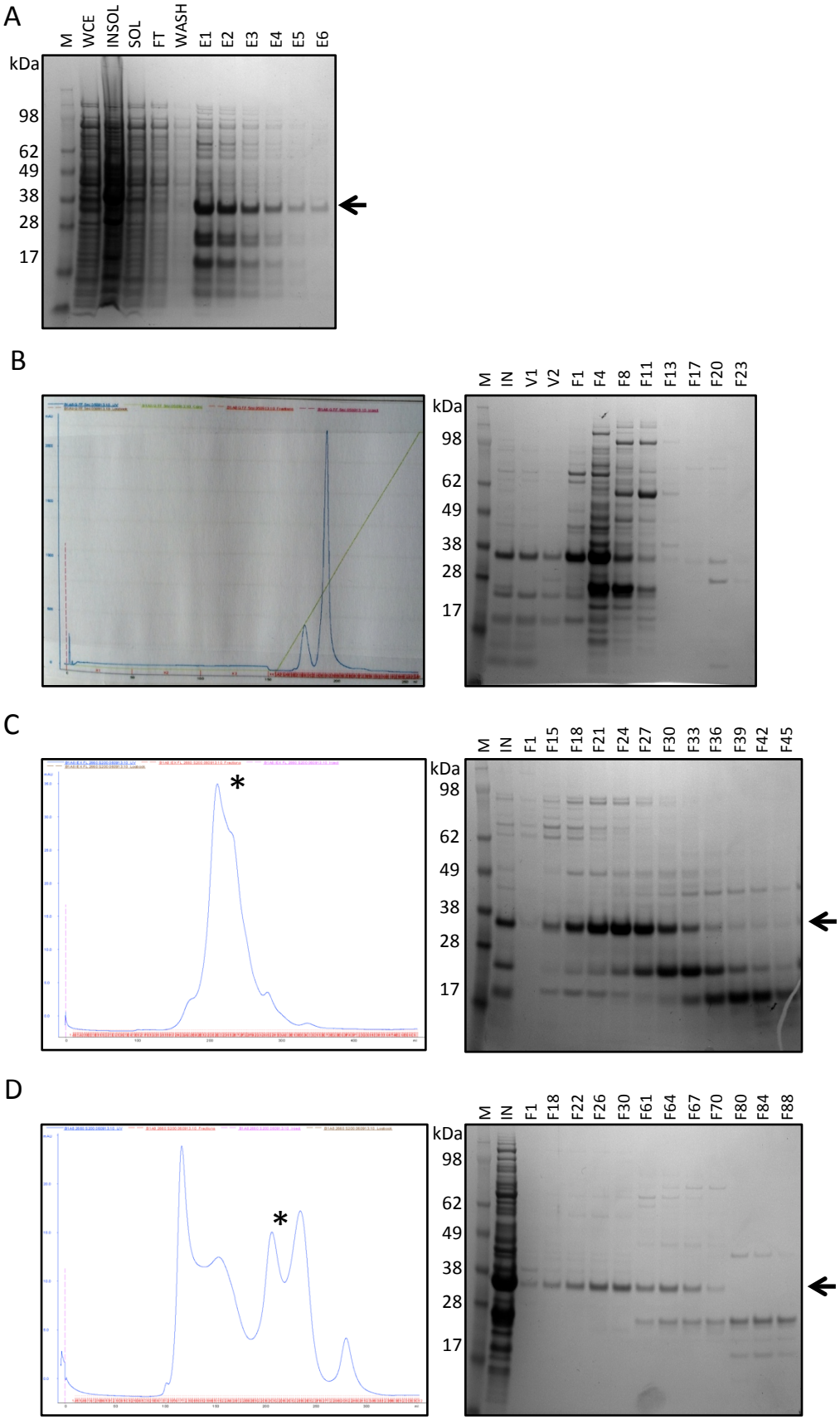


Appendix 1.6 Purification of expression construct A2A5: IMAC step.
SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps for A2A5 (aa 688-1114). M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. The arrow indicates the migration position of the full-length recombinant protein. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 1.7 Purification of expression construct A1A5: IMAC, Q-Sepharose and SEC steps.

(**A**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps for A1A5 (aa 1002-1305). M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. The arrow indicates the migration position of the full-length recombinant protein. (**B**) Q-Sepharose chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 100% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, F=indicated fraction. (**C**) SEC using a 26/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. The asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

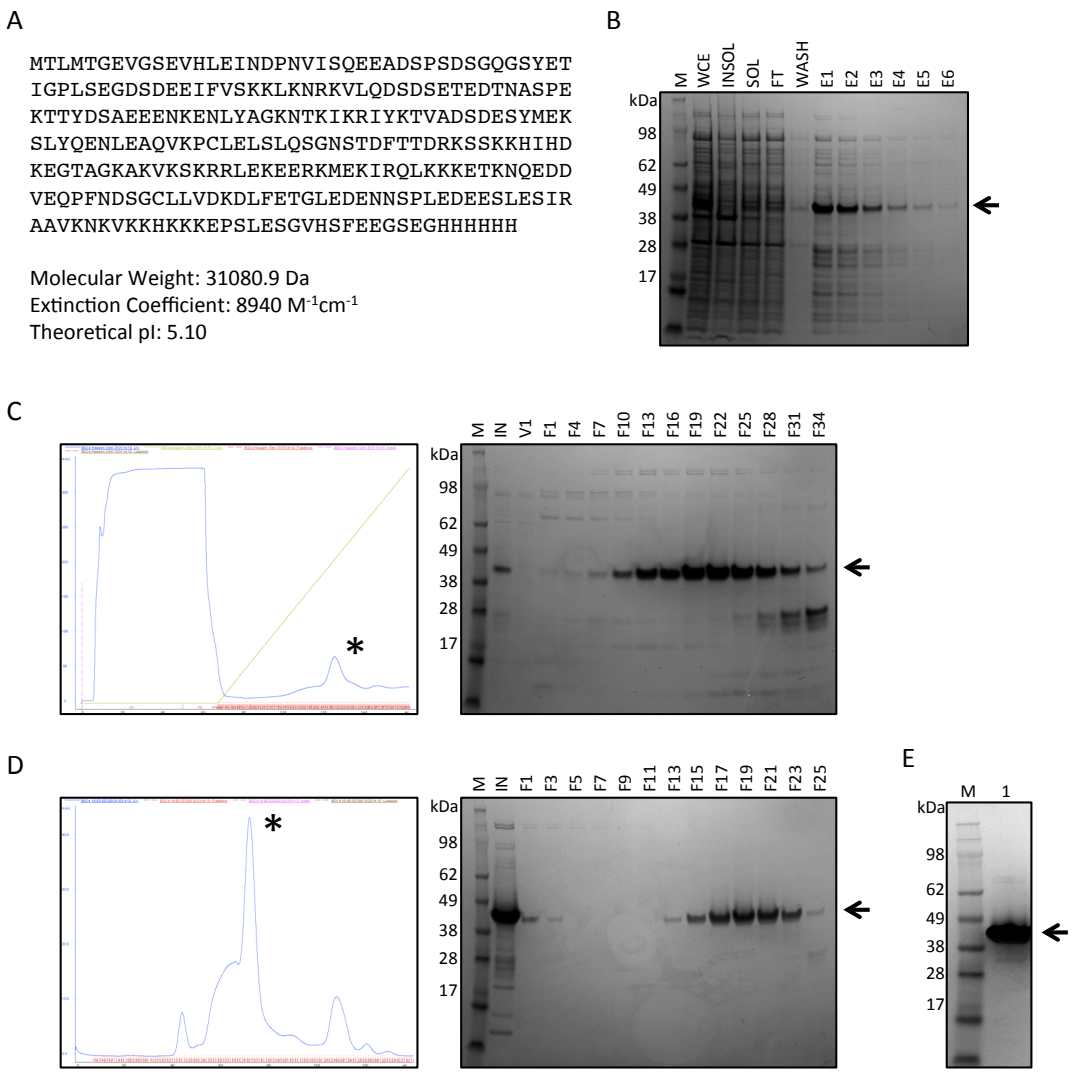

A
M
WCE
INSOL
SOL
FT
WASH
E1
E2
E3
E4
E5
E6
kDa
98
62
49
38
28
17
B
M
IN
V1
V2
F1
F4
F8
F11
F13
F17
F20
F23
kDa
98
62
49
38
28
17
C
*
M
IN
F1
F15
F18
F21
F24
F27
F30
F33
F36
F39
F42
F45
kDa
98
62
49
38
28
17
D
*
M
IN
F1
F18
F22
F26
F30
F61
F64
F67
F70
F80
F84
F88
kDa
98
62
49
38
28
17

Appendix 1.8 Purification of expression construct B1A8: IMAC, Q-Sepharose and SEC steps.

(**A**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps for B1A8 (aa 2224-2926). M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**B**) Q-Sepharose chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 100% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**C**) SEC using a 26/60 SD200 column for 'bound' B1A8. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**D**) SEC using a 26/60 SD200 column for 'unbound' B1A8. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

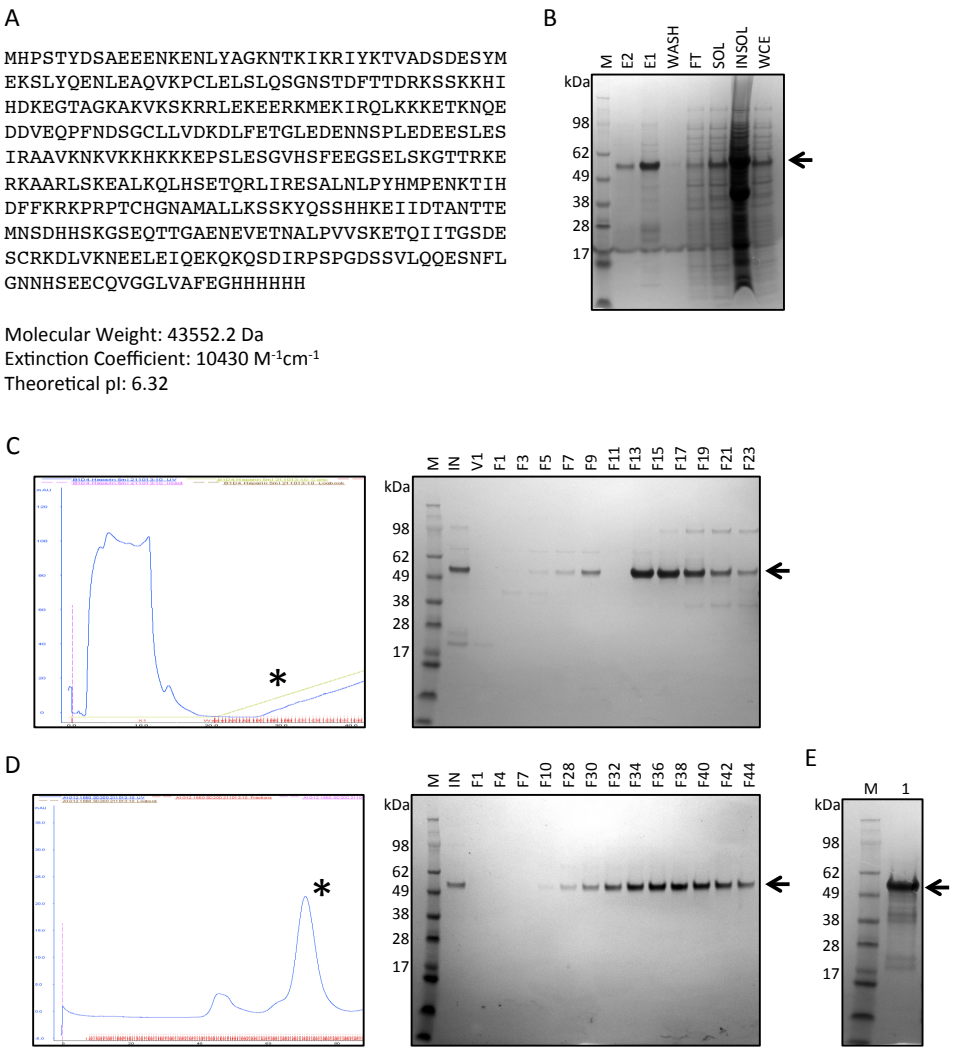
Appendix 2

Supplementary information for Chapter 5



Appendix 2.1 Purification of expression construct B2C4: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of B2C4 (aa 1-266), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.2 Purification of expression construct B1D4: IMAC, Heparin and SEC steps.

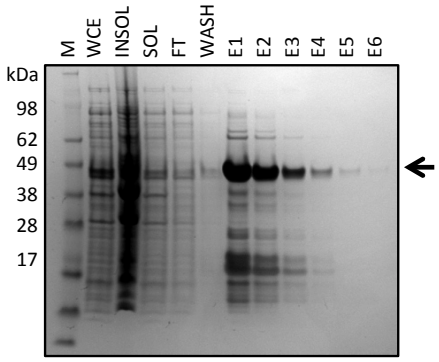
(**A**) The amino acid sequence of B1D4 (aa 80-452), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E2=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 30% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

A

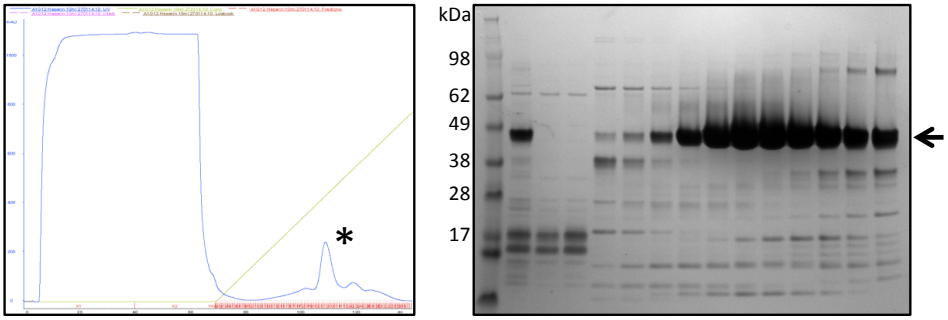
MPFYKTVADSDESYMEKSLYQENLEAQVKPCLELSLQSGN
STDFTTDRKSSKKHIHDKEGTAGKAKVKSKRRLEKEERKM
EKIRQLKKKETKNQEDDVEQPFNDSGCLLVDKDLFETGLE
DENNSPLEDEESLESIRATVKNKVKKHKKKEPSLESGVHS
FEEGSELSKGTTRKERKAARLSKEALKQLHSETQRLIRES
ALNLPYHMPENKTIHDFFKRKPRPTCHGNAMALLKSSKYQ
SSHHKEIIDTANTTEMNSDHHSKGSEQTTGAENEVETNAL
PVVSKETQIITGSDESCRKDLVKNEELEIQEKQGHHHHHH

Molecular Weight: 36521.6 Da
Extinction Coefficient: 7450 $M^{-1}cm^{-1}$
Theoretical pI: 6.71

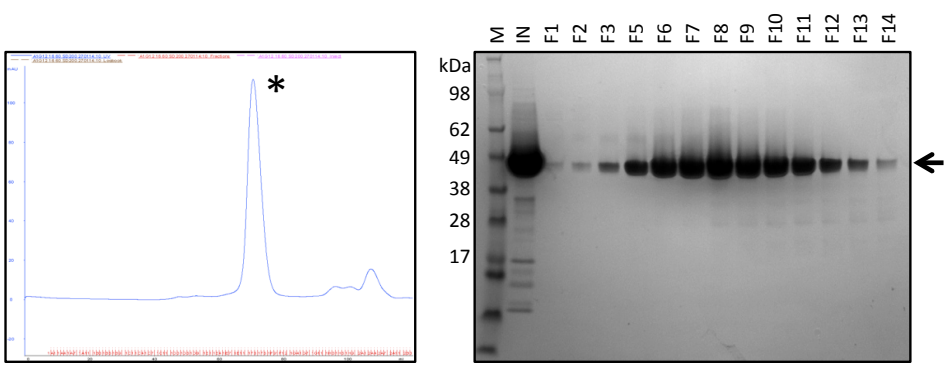
B



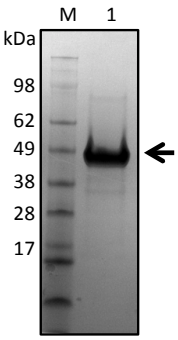
C



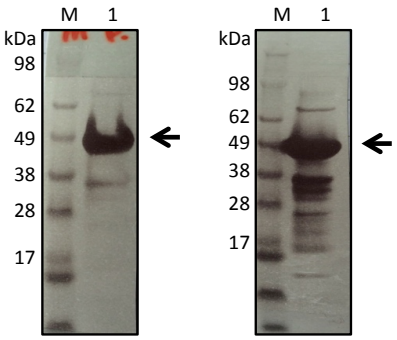
D



E

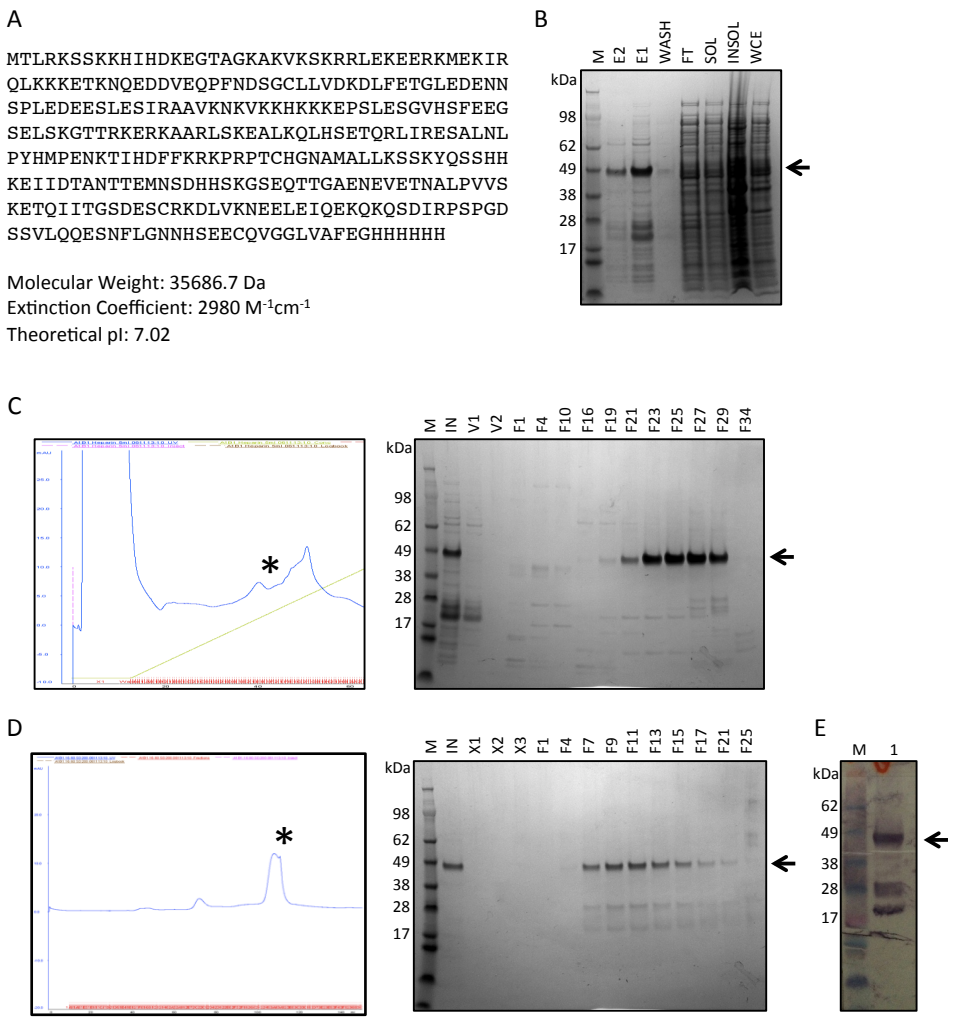


F



Appendix 2.3 Purification of expression construct A1G12: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of A1G12 (aa 104-413), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 80% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. (**F**) Assessment of A1G12 protein stability. (Left) protein after two weeks incubation at 4 °C. (Right) after four weeks incubation at 4 °C. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.4 Purification of expression construct A1B1: IMAC, Heparin and SEC steps.

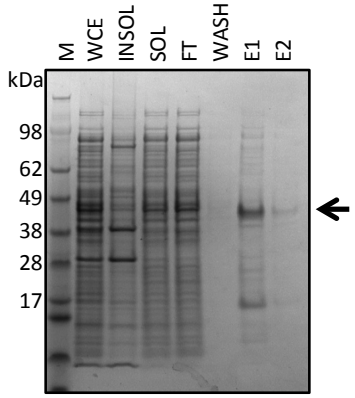
(**A**) The amino acid sequence of A1B1 (aa 148-452), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E2=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 30% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) Identification of the breakdown products of purified A1B1 protein using a colourmetric anti-His western blot (anti-His primary and AP-conjugated secondary antibodies). The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

A

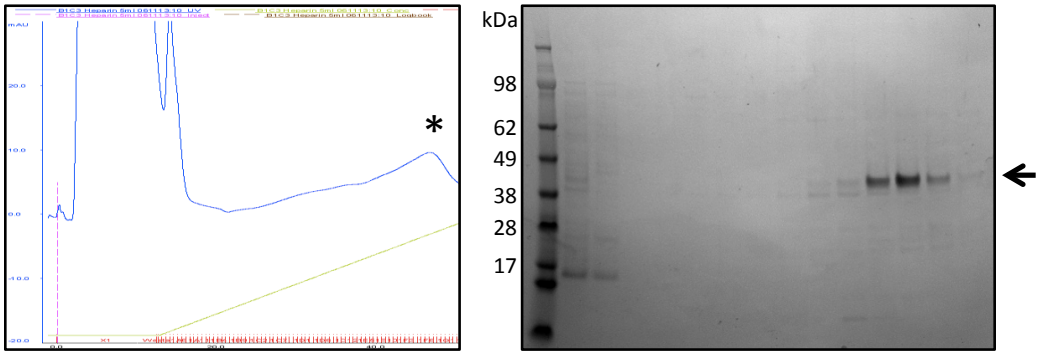
MTLIHDKEGTAGKAKVKSKRRLEKEERKMEKIRQLKKKET
KNQEDDVEQPFNDSGCLLVDKDLFETGLEDENNSPLEDEE
SLESIRAAVKNKVKKHKKKEPSLESGVHSFEEGSELSKGT
TRKERKAARLSKEALKQLHSETQRLIRESALNLPYHMPEN
KTIHDFFKRKPRPTCHGNAMALLKSSKYQSSHHKEIIDTA
NTTEMNSDHHSKGSEQTTGAENEVETNALPVVSKETQIIT
GSDESCRKDLVKNEELEIQEKQKQSDKGHHHHHH

Molecular Weight: 31234.8 Da
Extinction Coefficient: 2980 $M^{-1}cm^{-1}$
Theoretical pI: 7.28

B

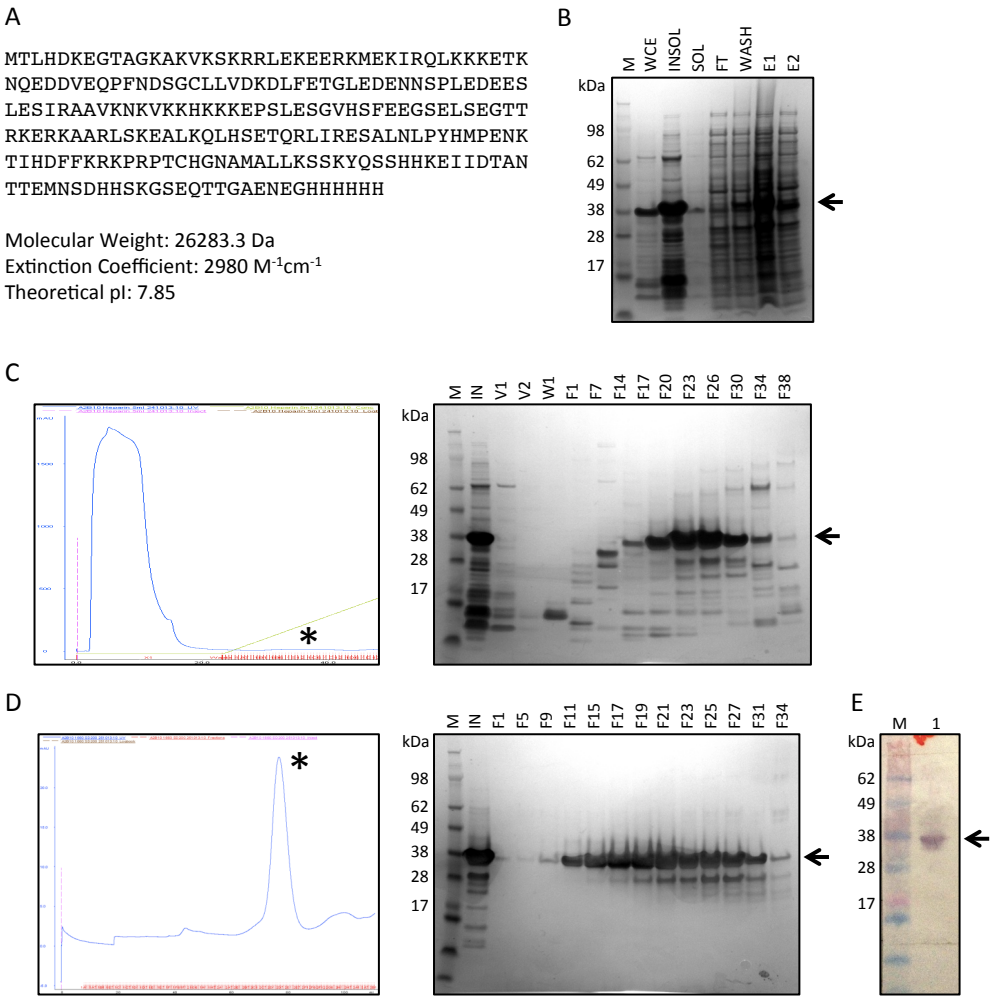


C



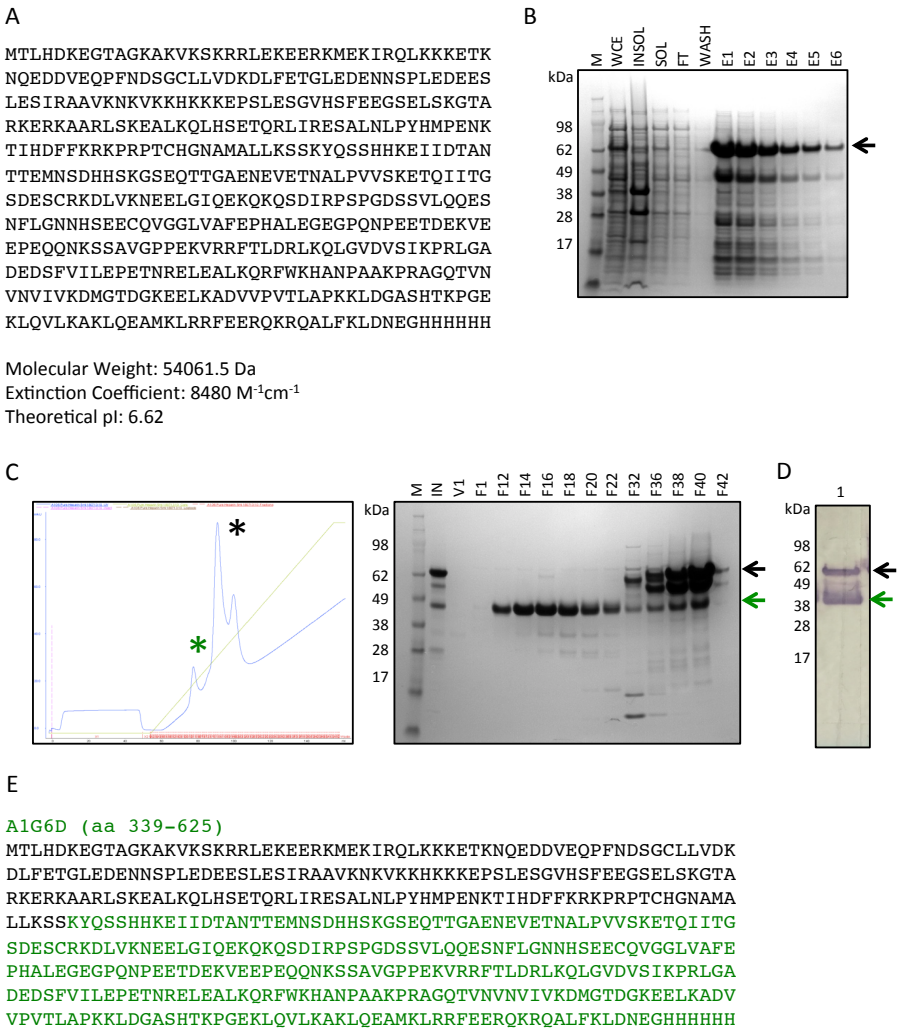
Appendix 2.5 Purification of expression construct B1C3: IMAC and Heparin steps.

(**A**) The amino acid sequence of B1C3 (aa 155-417), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E2=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 30% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



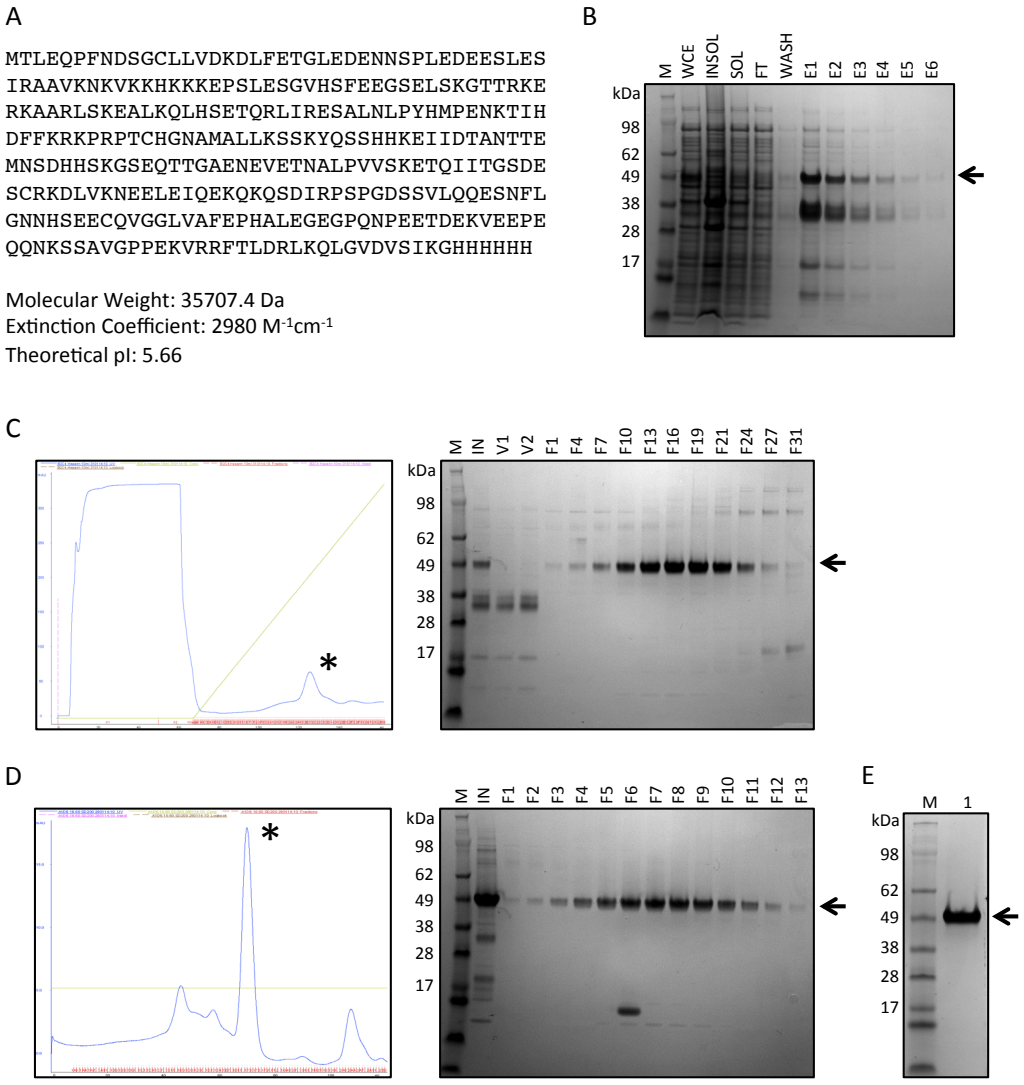
Appendix 2.6 Purification of expression construct A2B10: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of A2B10 (aa 156-347), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E2=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 30% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) Identification of the breakdown products of purified A2B10 protein fragment using a colourmetric anti-His western blot (anti-His primary and AP-conjugated secondary antibodies). The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.7 Purification of expression construct A1G6: IMAC and Heparin steps.

(**A**) The amino acid sequence of A1D6 (aa 156-625), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) Identification of the breakdown products of Heparin purified A1G6 protein fragment using a colourmetric anti-His western blot (anti-His primary and AP-conjugated secondary antibodies). 1=F36. (**E**) The amino acid sequence of A1G6, with the breakdown product A1G6D in green text. The black arrow / asterisks indicates the full-length recombinant protein, whilst the green arrow / asterisks indicates the breakdown product. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.8 Purification of expression construct A1D6: IMAC, Heparin and SEC steps.

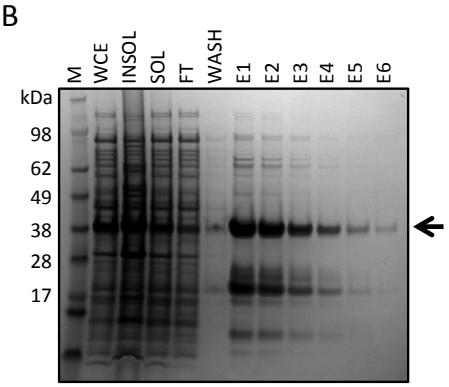
(**A**) The amino acid sequence of A1D6 (aa 199-507), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 80% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

A

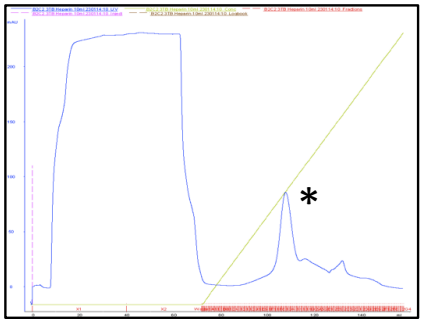
MHPSGLEDENNSPLEDEESLESIRAAVKNKVKKHKKKEPS
LESGVHSFEEGSELSKGTTRKERKAARLSKEALKQLHSET
QRLIRESALNLPYHMPENKTIHDFFKRKPRPTCHGNAMAL
LKSSKYQSSHHKEIIDTANTTEMNSDHHSKGSEQTTGAEN
EVETNALPVVSKETQIITGSDESCRKDLVKNEELEIQEKQ
KQSDIRPSPGDSSVLQQESNFLGNNHSEGHHHHHH

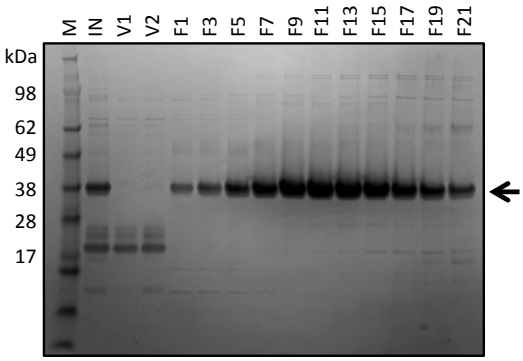
Molecular Weight: 26409.0 Da
Extinction Coefficient: 2980 $M^{-1}cm^{-1}$
Theoretical pI: 6.43

B

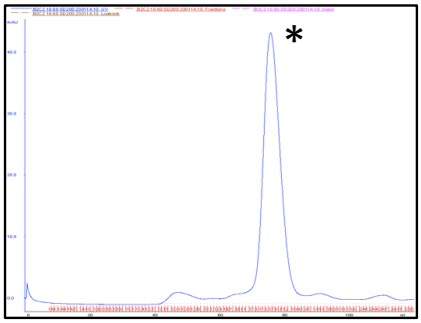


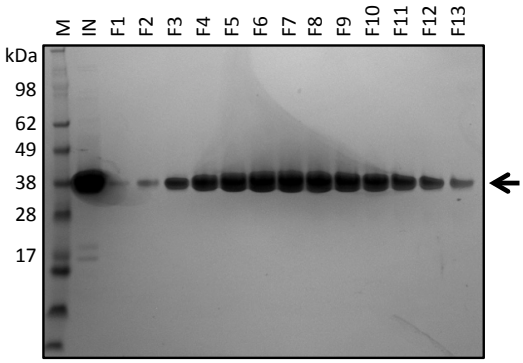
C



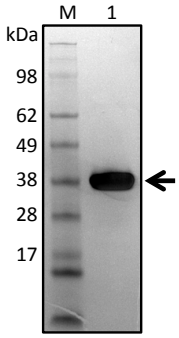


D

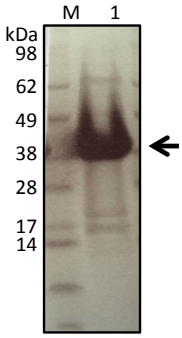




E

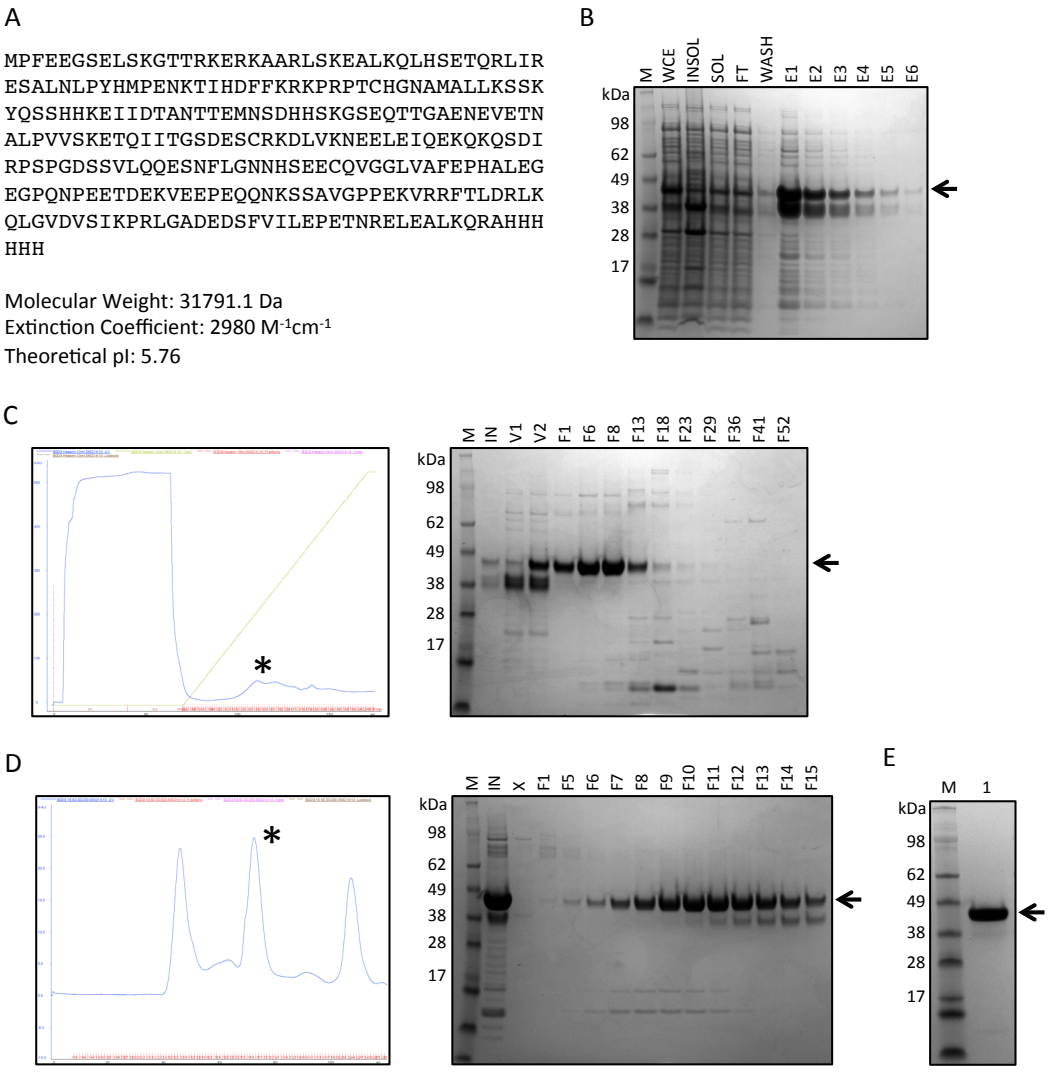


F



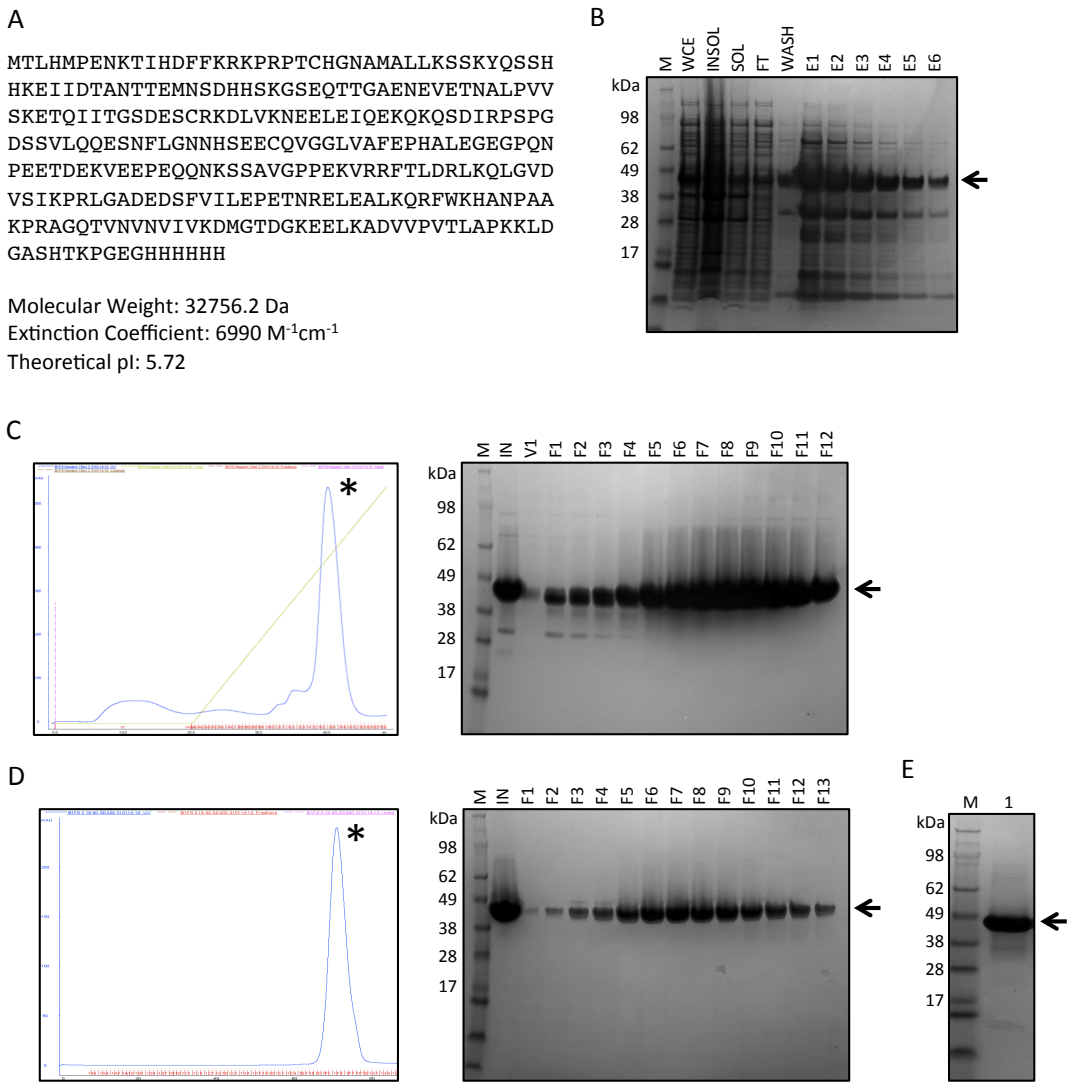
Appendix 2.9 Purification of expression construct B2C2: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of B2C2 (aa 218-441), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are detailed. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 80% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. (**F**) Assessment of B2C2 protein stability after four weeks incubation at 4 °C. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



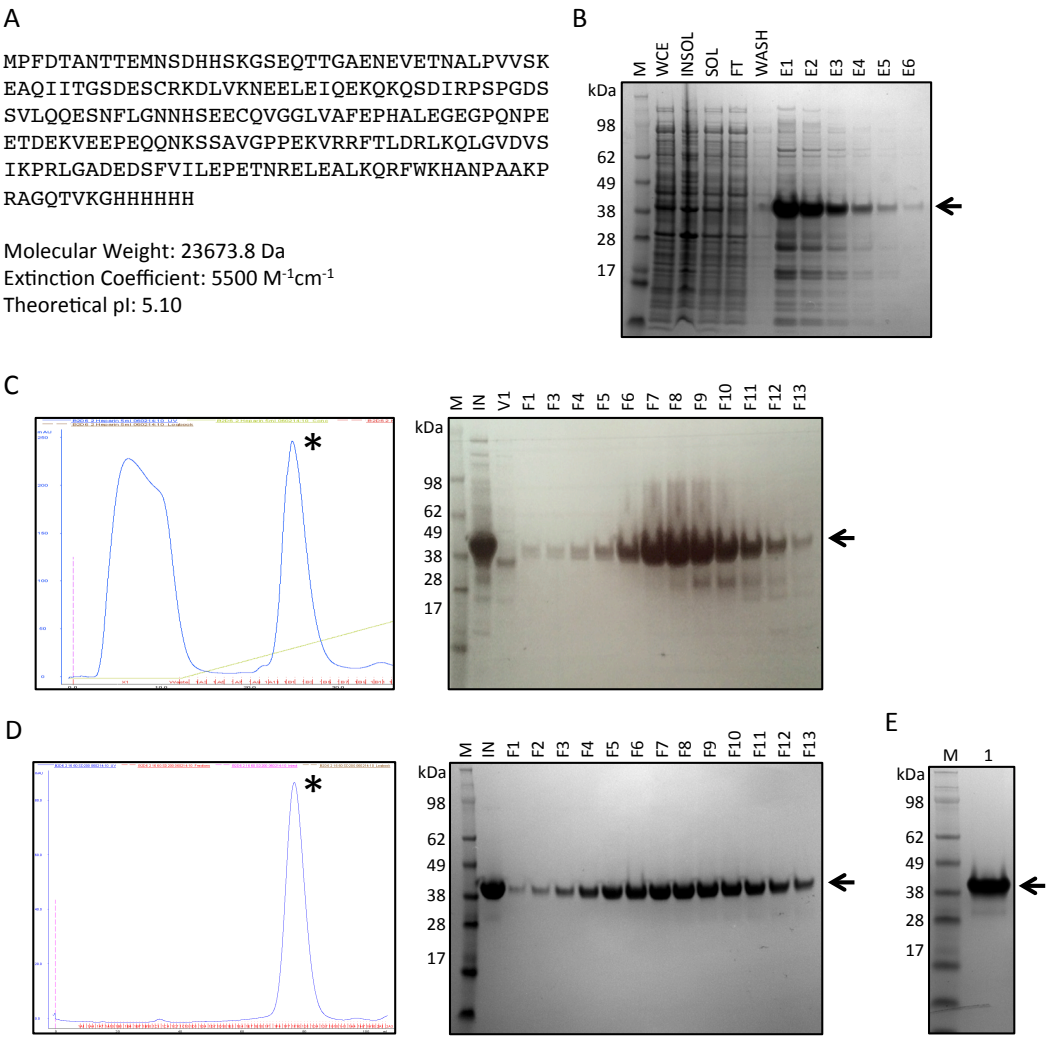
Appendix 2.10 Purification of expression construct B2D9: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of B2D9 (aa 261-534), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.11 Purification of expression construct B1F8: IMAC, Heparin and SEC steps.

(**A**) The amino acid sequence of B1F8 (aa 307-592), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 40% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.12 Purification of expression construct B2D6: IMAC, Heparin and SEC steps.

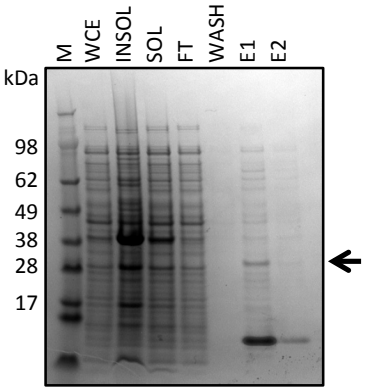
(**A**) The amino acid sequence of B2D6 (aa 349-551), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 30% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

A

MTLRPSPGDSSVLQQESNFLGNNHSEECQVGGLVAFEPHA
LEGEGPQNPEETDEKVEEPEQQNKSSAVGPPEKVRRFTLD
RLKQLGVDVSIKPRLGADEDSFVILEPETNRELEALKQRF
WKHANPAAKPRAGQTVNVNVIVKDMGTDGKEELKADVVPV
TLAPKKLDGASHTKPGEKLQVLKAKLQEAMKLRRFEGHHH
HHH

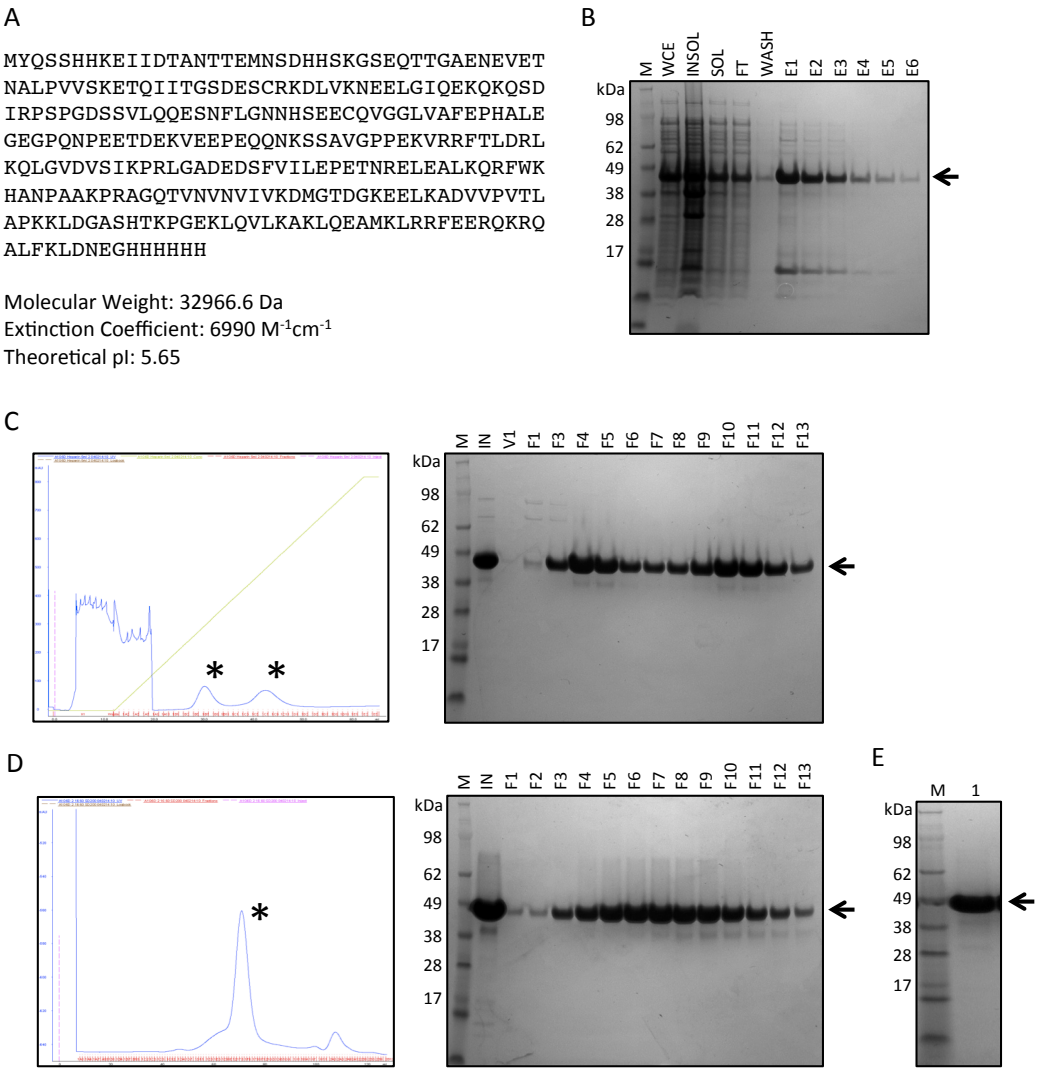
Molecular Weight: 22412.1 Da
Extinction Coefficient: 5500 $M^{-1}cm^{-1}$
Theoretical pI: 6.07

B



Appendix 2.13 Purification of expression construct B1G11: IMAC, Heparin and SEC steps.

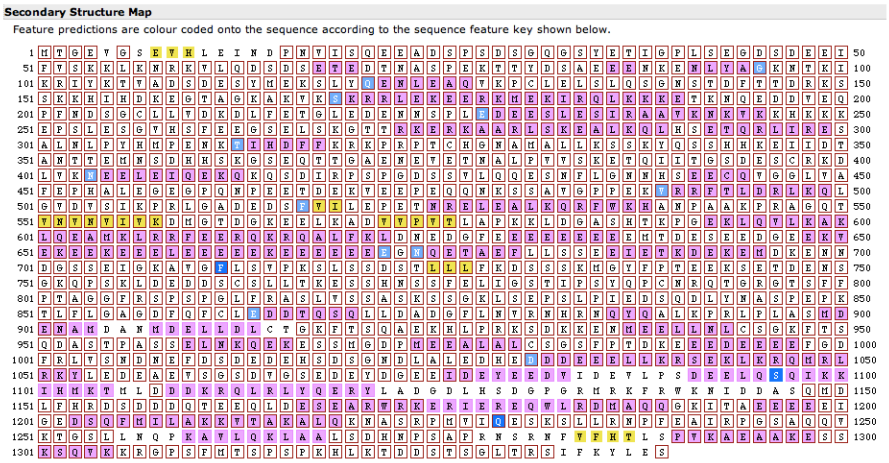
(**A**) The amino acid sequence of B1G11 (aa 419-611), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E2=successive elution fractions. The arrow indicates the migration position of the full-length recombinant protein. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 2.14 Purification of expression construct A1G6D: IMAC, Heparin and SEC steps.

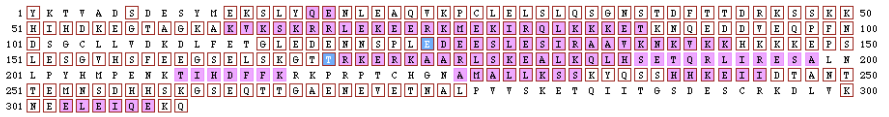
(**A**) The amino acid sequence of A1G6D (aa 339-625), as expressed from the pDMX4-TOPO vector with a C-terminal His_6 affinity tag. The predicted molecular weight, extinction coefficient and pI for the protein fragment are shown. (**B**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps. M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E6=successive elution fractions. (**C**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, V=fractions from the void volume, F=indicated fraction. (**D**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. (**E**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

Claspin

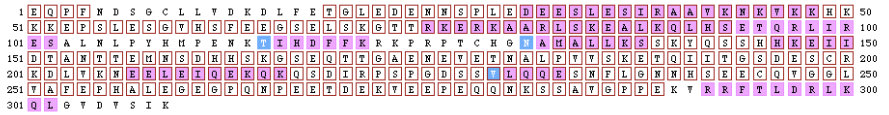


B2C4

A1G12



A1D6

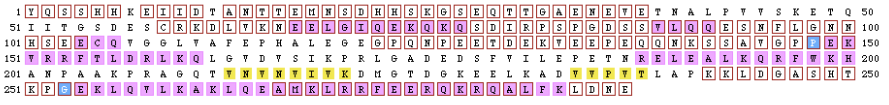


B2C2

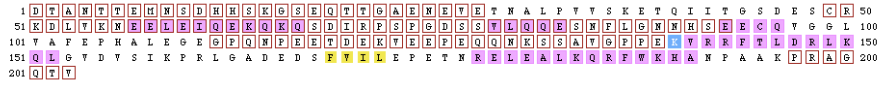
B2D9

B1F8

A1G6D



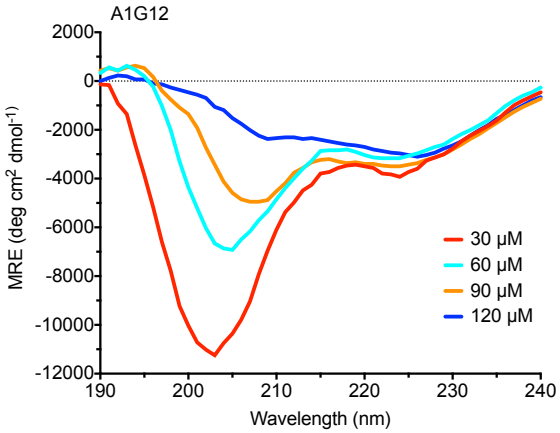
B2D6

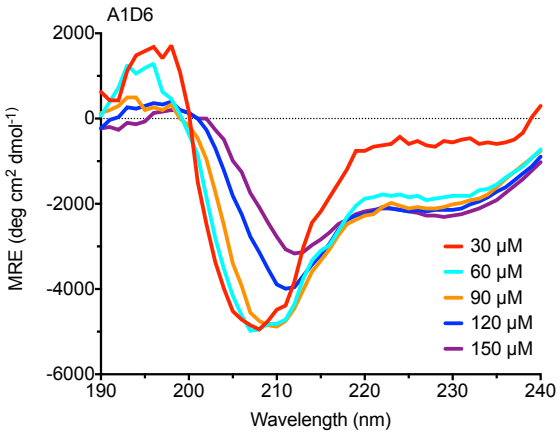


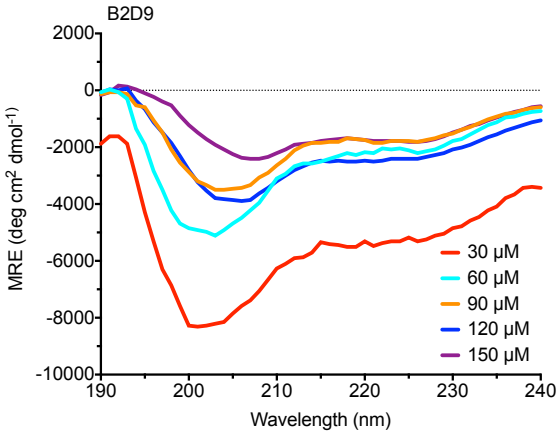
Appendix 2.15 PsiPred secondary structure predictions for selected Claspin protein fragments.

Secondary structure analysis of the full-length amino acid sequence of human Claspin in comparison to the eight N-terminal fragments. Regions predicted to form α-helical elements are represented by the pink coloured boxes (see key), whilst yellow coloured boxes represent those regions predicted to form β-strands. Regions of low amino acid sequence complexity are indicated by the red outline (Jones, 1999, Buchan et al., 2013).

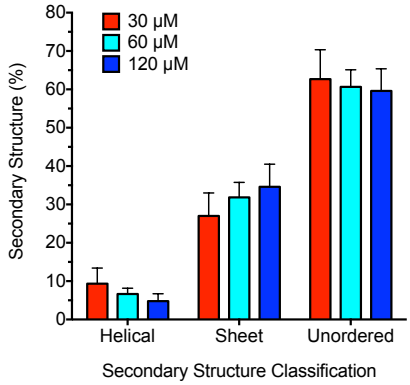
A

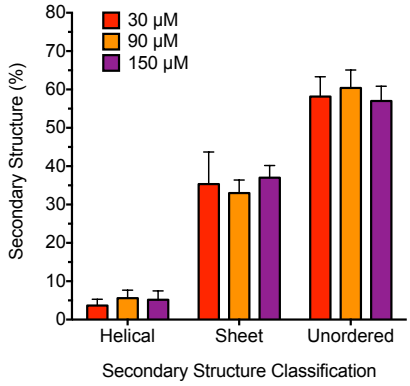

A1G12
MRE (deg cm2 dmol-1)
Wavelength (nm)
30 µM
60 µM
90 µM
120 µM

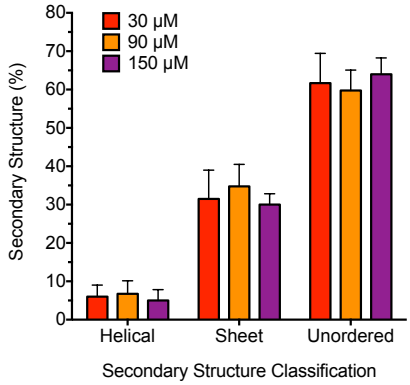

A1D6
MRE (deg cm2 dmol-1)
Wavelength (nm)
30 µM
60 µM
90 µM
120 µM
150 µM

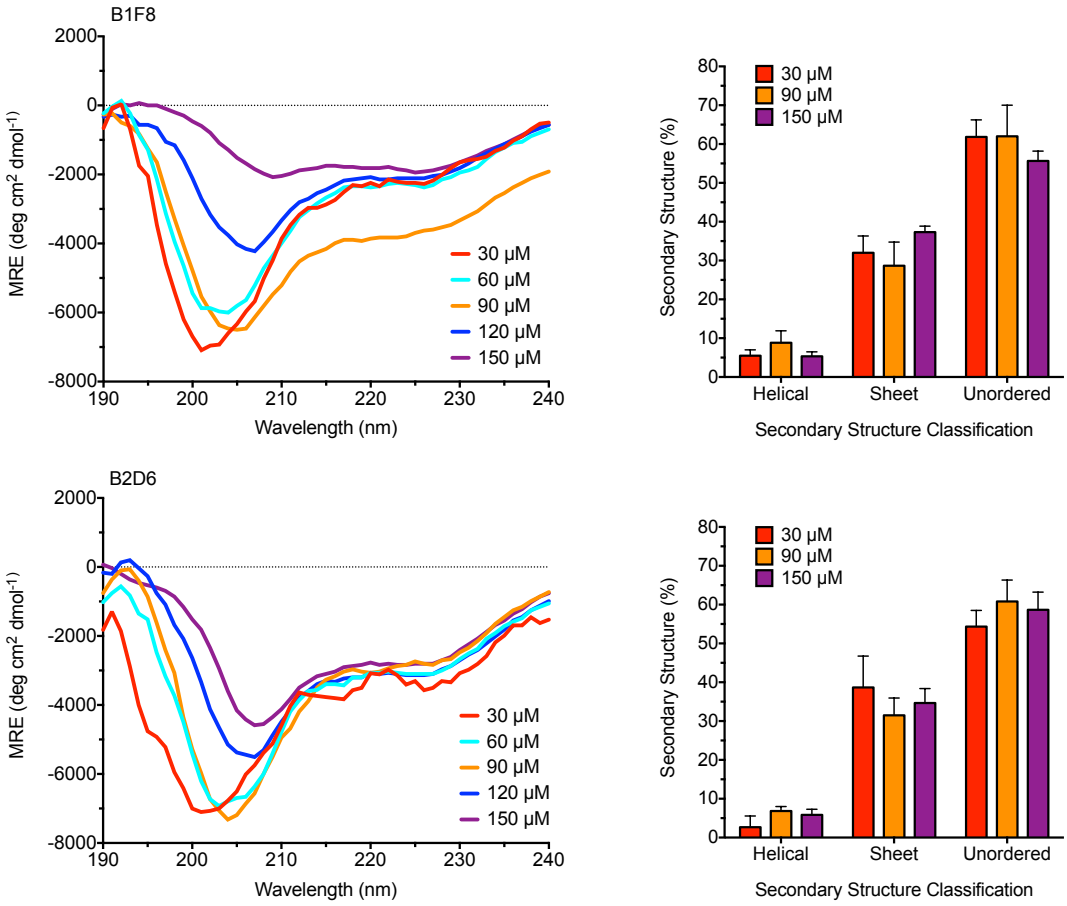

B2D9
MRE (deg cm2 dmol-1)
Wavelength (nm)
30 µM
60 µM
90 µM
120 µM
150 µM

B


Secondary Structure (%)
30 µM
60 µM
120 µM
Helical
Sheet
Unordered
Secondary Structure Classification

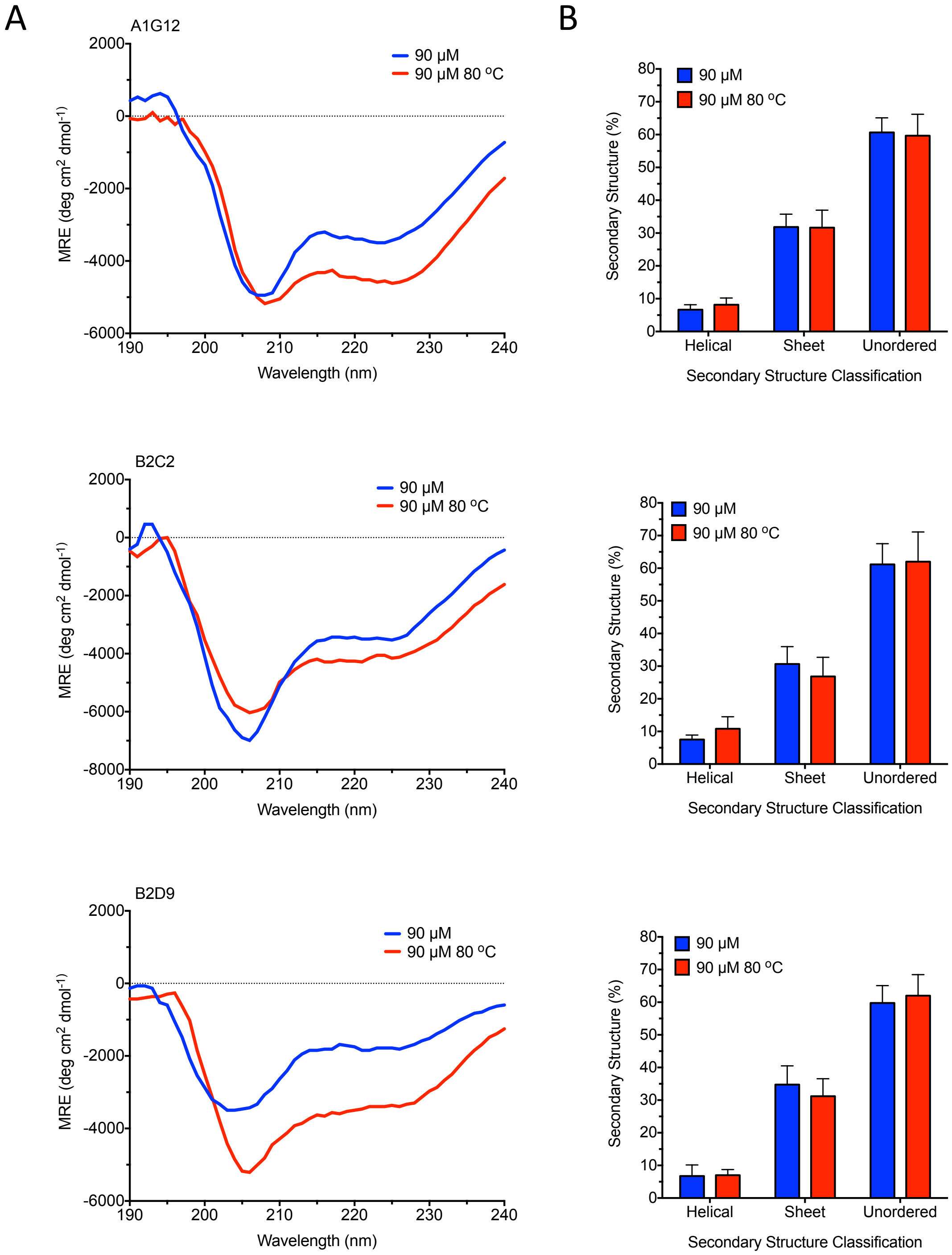

Secondary Structure (%)
30 µM
90 µM
150 µM
Helical
Sheet
Unordered
Secondary Structure Classification

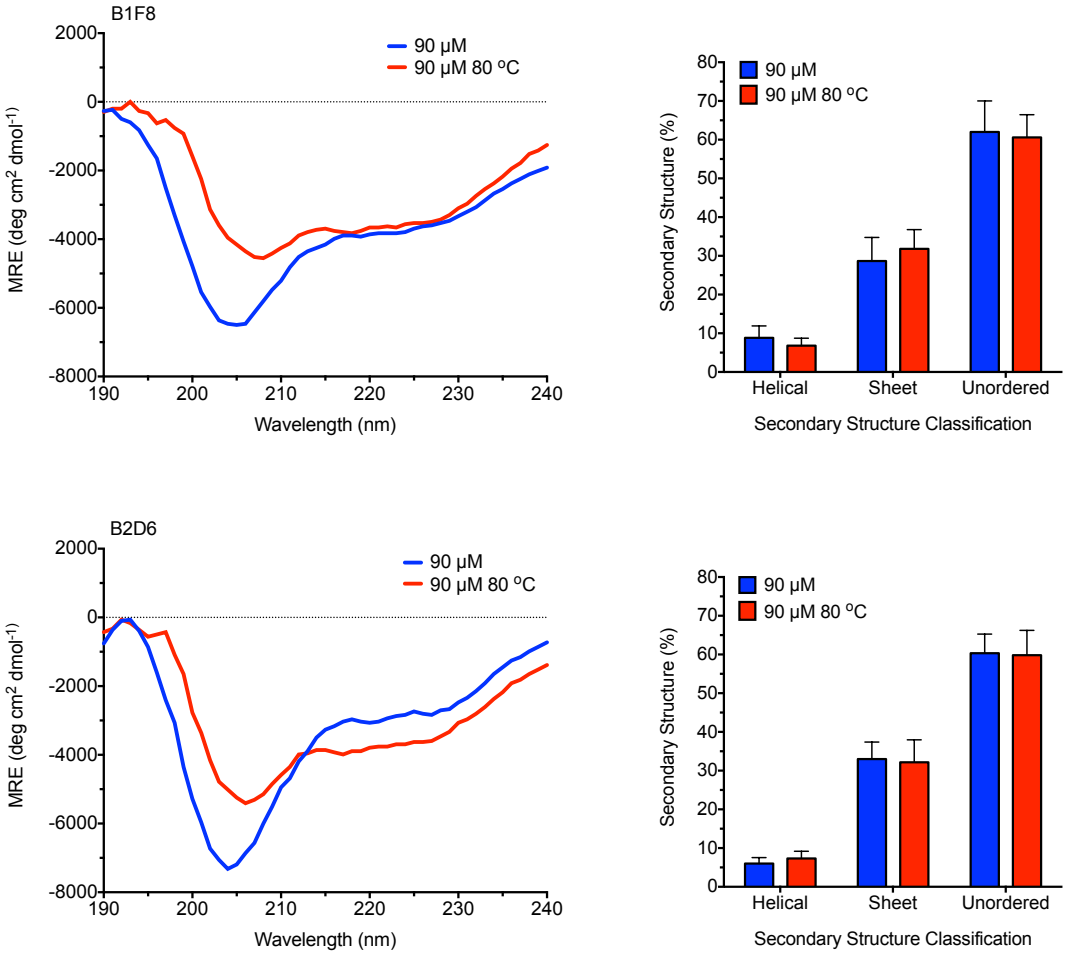

Secondary Structure (%)
30 µM
90 µM
150 µM
Helical
Sheet
Unordered
Secondary Structure Classification



Appendix 2.16 CD spectra and Dichroweb deconvolution for the N-terminal Claspin protein fragments.

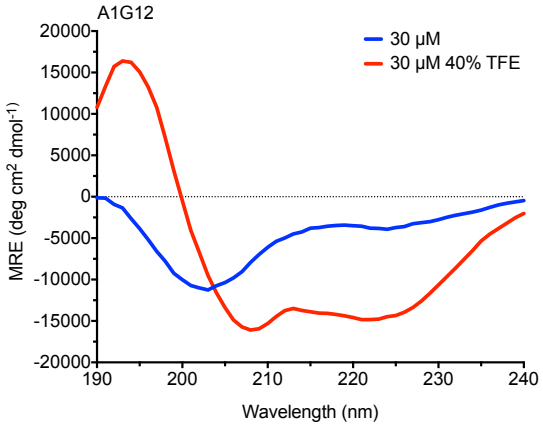
(**A**) CD spectra of recombinant Claspin protein fragments from top to bottom: A1G12, A1D6, B2D9, B1F8 and B2D6 (labelled top left). Protein samples were placed in a 0.1 mm quartz cuvette and the spectra measured between a wavelength of 190 and 240 nm using a JASCO J-715 spectropolarimeter, at 20 °C controlled by a JASCO PTC-384W peltier temperature control system. Data represents the average of three scans, from which the spectrum of buffer alone has been subtracted. The protein concentration is as indicated in the associated key. (**B**) Percentage of predicted secondary structure. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm, using six individual data sets, and then averaged. Error bars show one standard deviation. Protein concentration is indicated in the associated key (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).

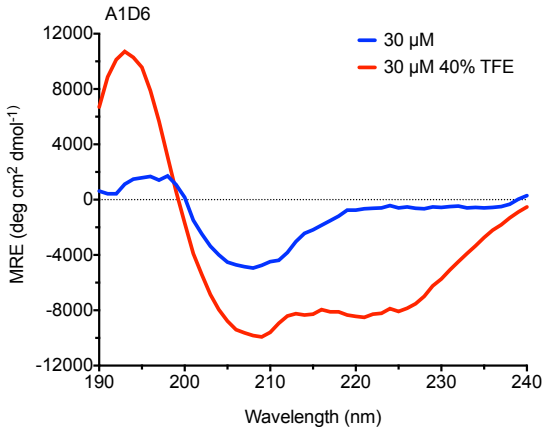

A
B
A1G12
B2C2
B2D9
90 µM
90 µM 80 °C
MRE (deg cm2 dmol-1)
Wavelength (nm)
Secondary Structure (%)
Secondary Structure Classification
Helical
Sheet
Unordered
2000
0
-2000
-4000
-6000
-8000
190
200
210
220
230
240
0
10
20
30
40
50
60
70
80

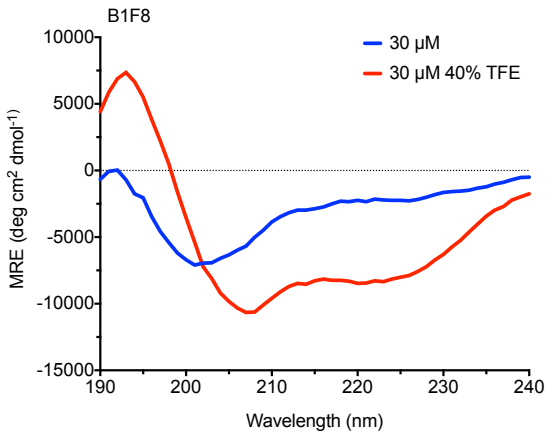


Appendix 2.17 Two temperature point CD spectra and Dichroweb deconvolution for the N-terminal Claspin protein fragments.
(**A**) CD spectra, measured at two temperature points, for selected recombinant Claspin protein fragments. From top to bottom: A1G12, A1D6, B2D9, B1F8 and B2D6. The protein concentration as well as the temperature (80 °C) is shown in the associated key. Samples were placed in a 0.1 mm quartz cuvette and CD spectra were measured at wavelengths between 190 and 240 using a JASCO J-715 spectropolarimeter. Samples in cuvettes were heated at 80 °C for 3 minutes controlled by a JASCO PTC-384W peltier temperature control system and the CD spectra were re-measured. Data represents an average of three scans, from which the spectrum of the buffer alone has been subtracted. (**B**) Percentage of predicted secondary structure for recombinant Claspin protein fragments. The protein concentration as well as the temperature (80 °C) is shown in the associated key. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm using six individual data sets, and then averaged. Error bars show one standard deviation (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace 2008).

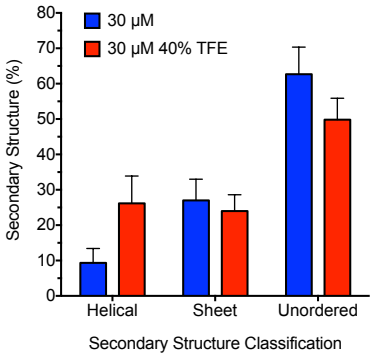
A

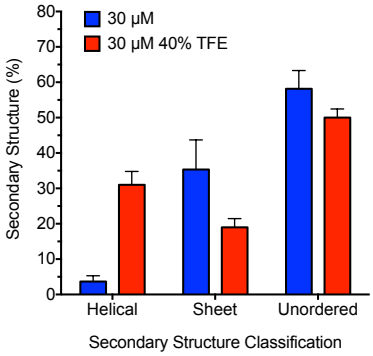

A1G12
30 μM
30 μM 40% TFE
MRE (deg cm2 dmol-1)
Wavelength (nm)

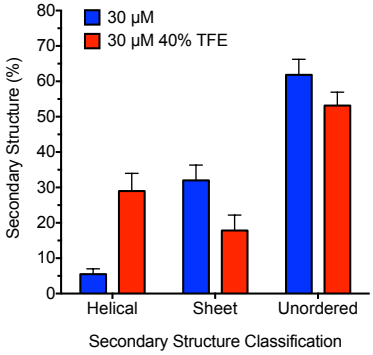

A1D6
30 μM
30 μM 40% TFE
MRE (deg cm2 dmol-1)
Wavelength (nm)

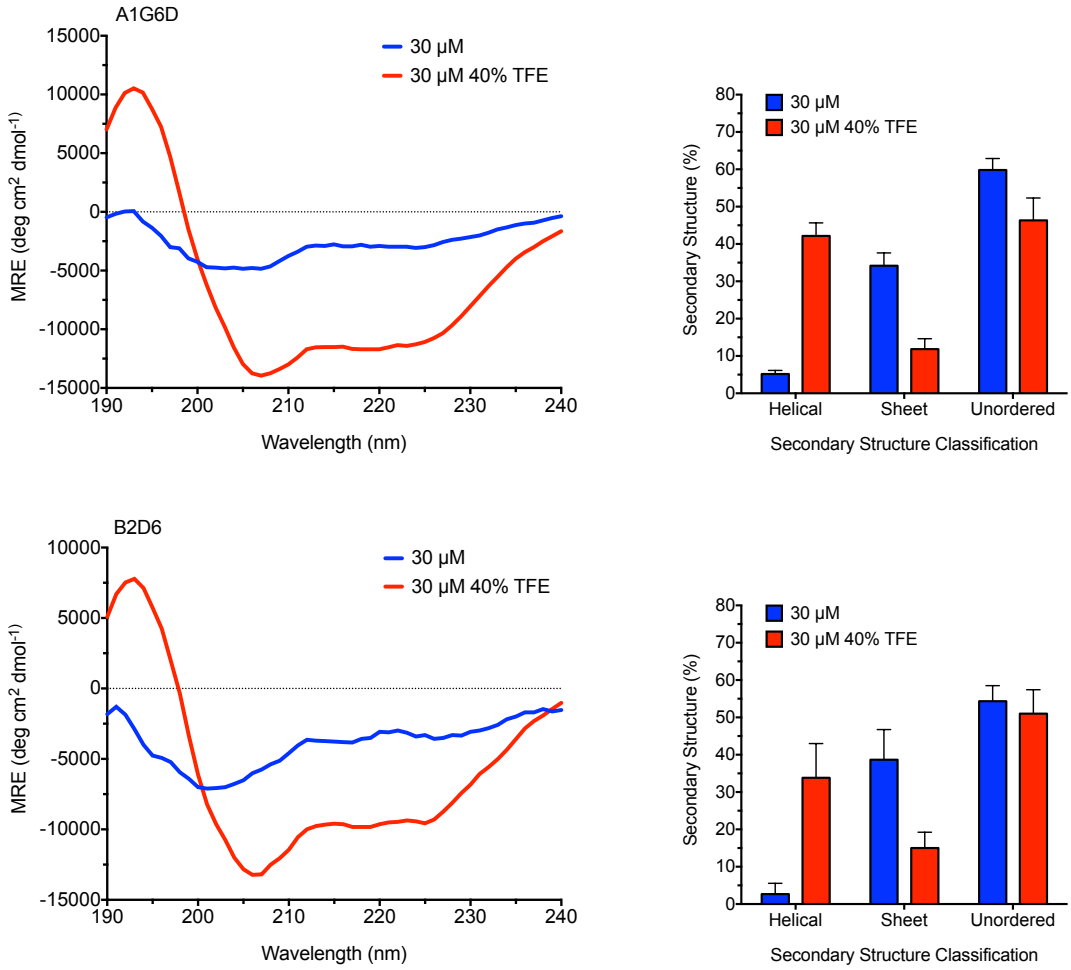

B1F8
30 μM
30 μM 40% TFE
MRE (deg cm2 dmol-1)
Wavelength (nm)

B

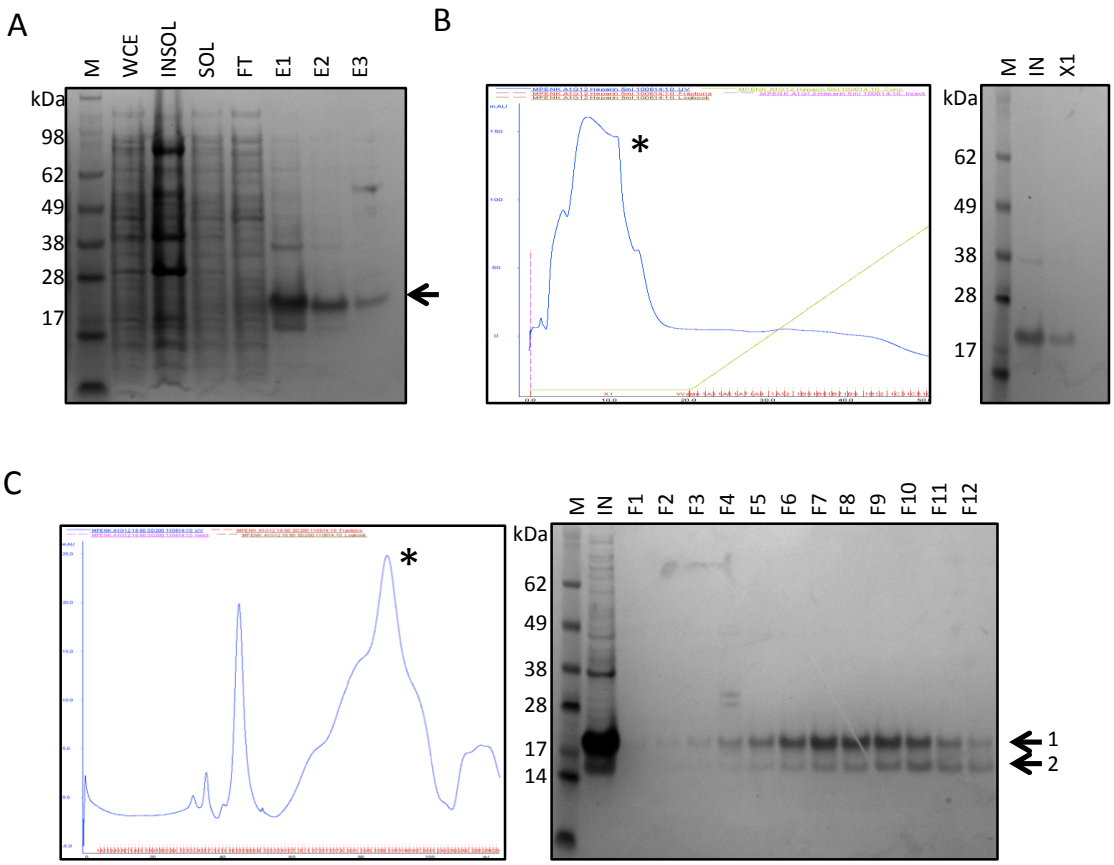

30 μM
30 μM 40% TFE
Secondary Structure (%)
Helical
Sheet
Unordered
Secondary Structure Classification


30 μM
30 μM 40% TFE
Secondary Structure (%)
Helical
Sheet
Unordered
Secondary Structure Classification


30 μM
30 μM 40% TFE
Secondary Structure (%)
Helical
Sheet
Unordered
Secondary Structure Classification

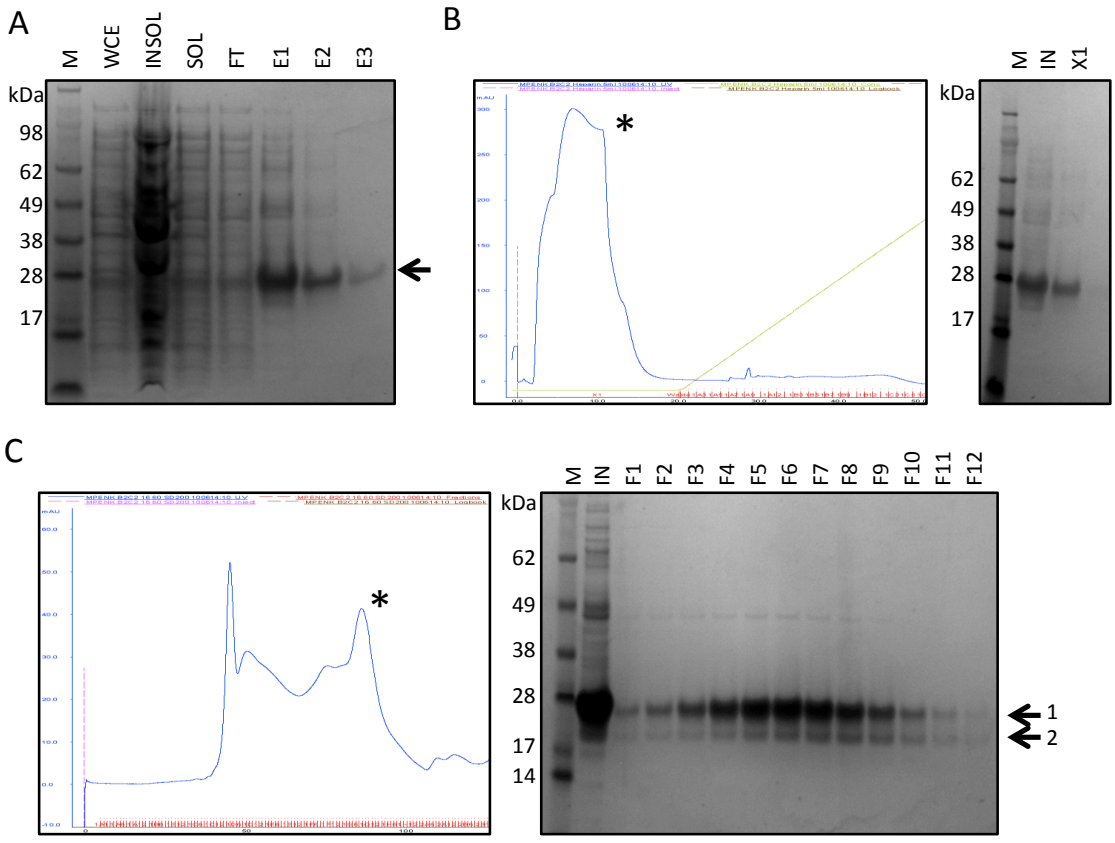


Appendix 2.18 CD spectra and DichroWeb deconvolution for the N-terminal Claspin protein fragments in the presence of TFE.
(**A**) CD spectra of recombinant Claspin protein fragments; from top to bottom: A1G12, A1D6, B1F8, A1G6D and B2D6, measured with and without 40% (v/v) TFE. The protein concentration and the TFE concentration are shown in the associated key. Samples were placed in a 0.1 mm quartz cuvette, and spectra were measured at wavelengths between 190 and 240 using a JASCO J-715 spectropolarimeter set at 20 °C controlled by a JASCO PTC-384W peltier temperature control system. Data represents the average of three scans, from which the spectra of the buffer alone has been subtracted. (**B**) Percentage of predicted secondary structure. The protein concentration as well as the TFE concentration is shown in the associated key. Spectral deconvolution was carried out using the DichroWeb server and the CDSSTR algorithm using six individual data sets, and then averaged. Error bars show one standard deviation (Sreerama et al., 2000a, Sreerama et al., 2000b, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).



Appendix 2.19 Purification of expression construct MPENK-1: IMAC, Heparin and SEC steps.

(**A**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps of MPENK-1 (aa 308-413). M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, E1 to E3=successive elution fractions. The arrow indicates the migration position of the full-length recombinant protein. (**B**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, X=fractions from the void volume. (**C**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. F=indicated fraction. Arrow 1=full-length MPENK-1, arrow 2=MPENK-1 degradation product, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

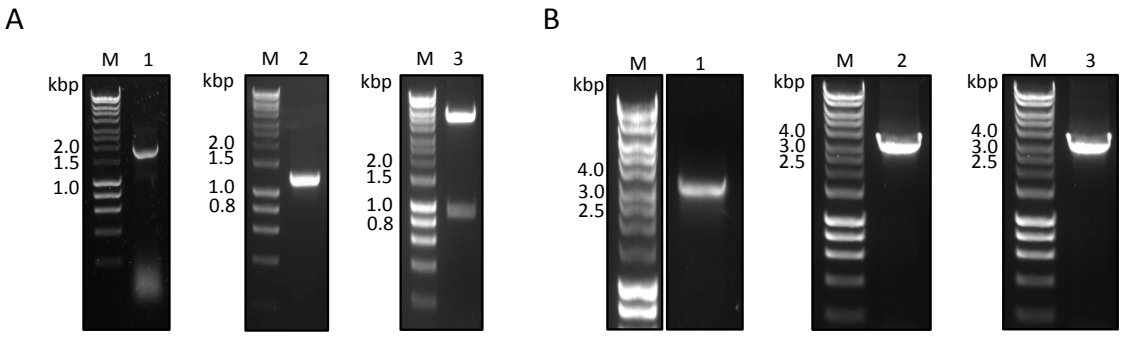


Appendix 2.20 Purification of expression construct MPENK-2: IMAC, Heparin and SEC steps.

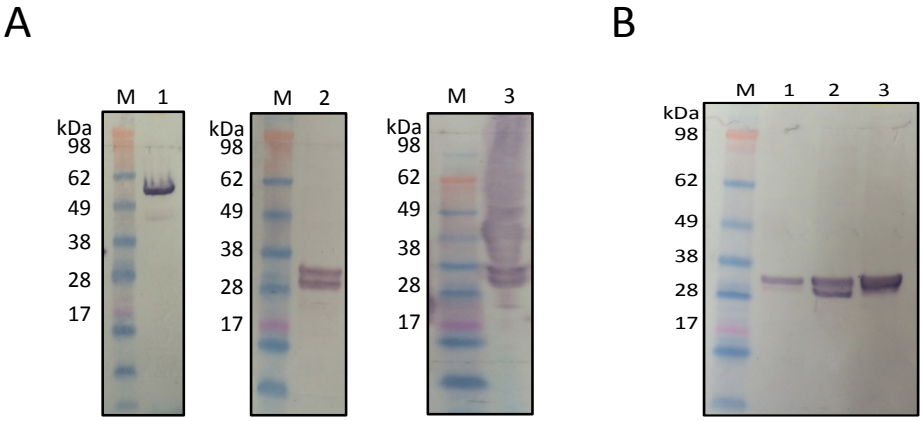
(**A**) SDS-PAGE analysis of *E. coli* cell lysate, and IMAC purification steps of MPENK-1 (aa 308-441). M=molecular mass marker, WCE=whole cell extract, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, E1 to E3=successive elution fractions. The arrow indicates the migration position of the full-length recombinant protein. (**B**) Heparin chromatography step. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line), applied NaCl gradient (to 60% IEX B buffer, green line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, X=fractions from the void volume. (**C**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. F=indicated fraction. Arrow 1=full-length MPENK-2, arrow 2=MPENK-2 degradation product, and the asterisks indicate the elution peaks. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.

Appendix 3

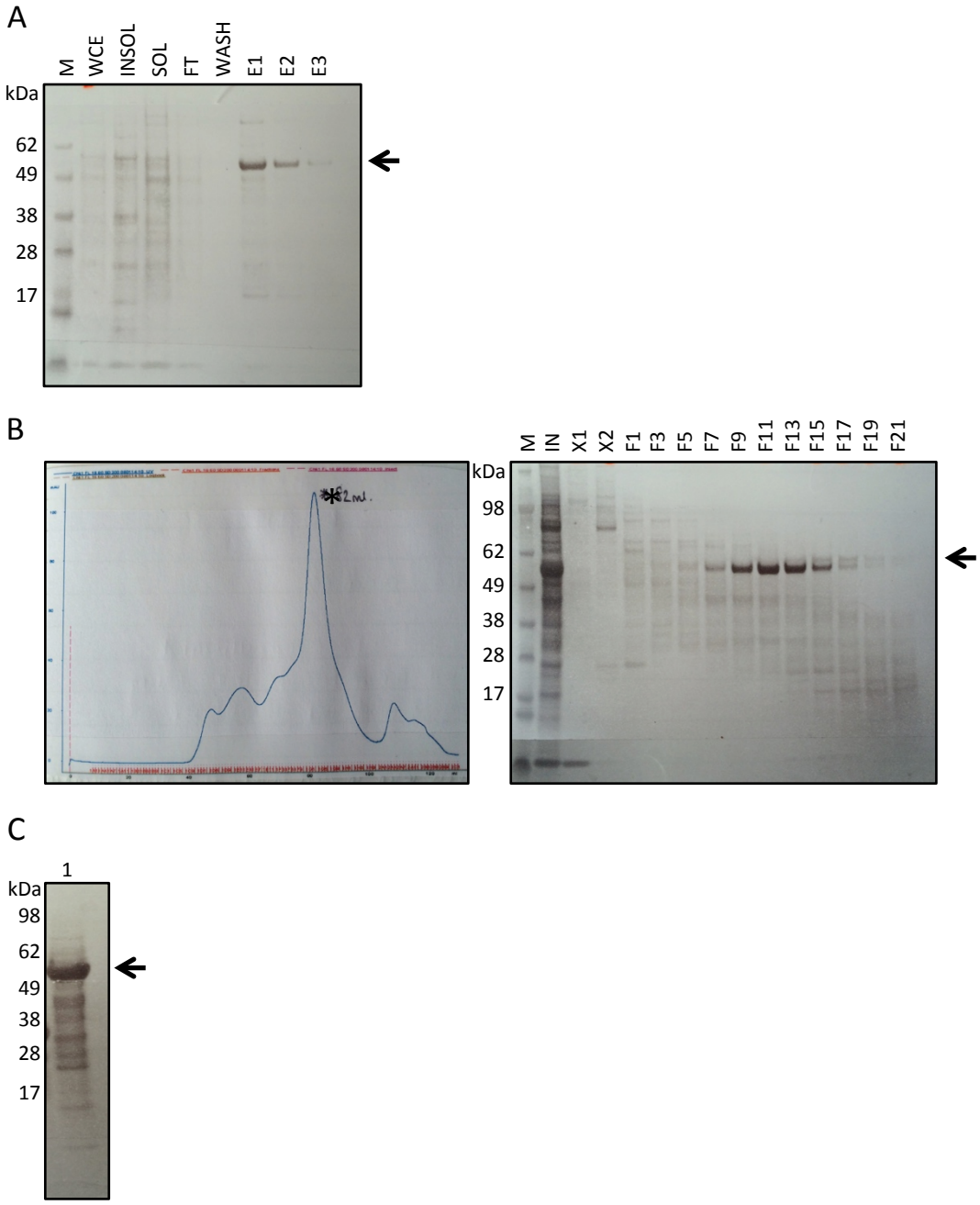
Supplementary information for Chapter 7



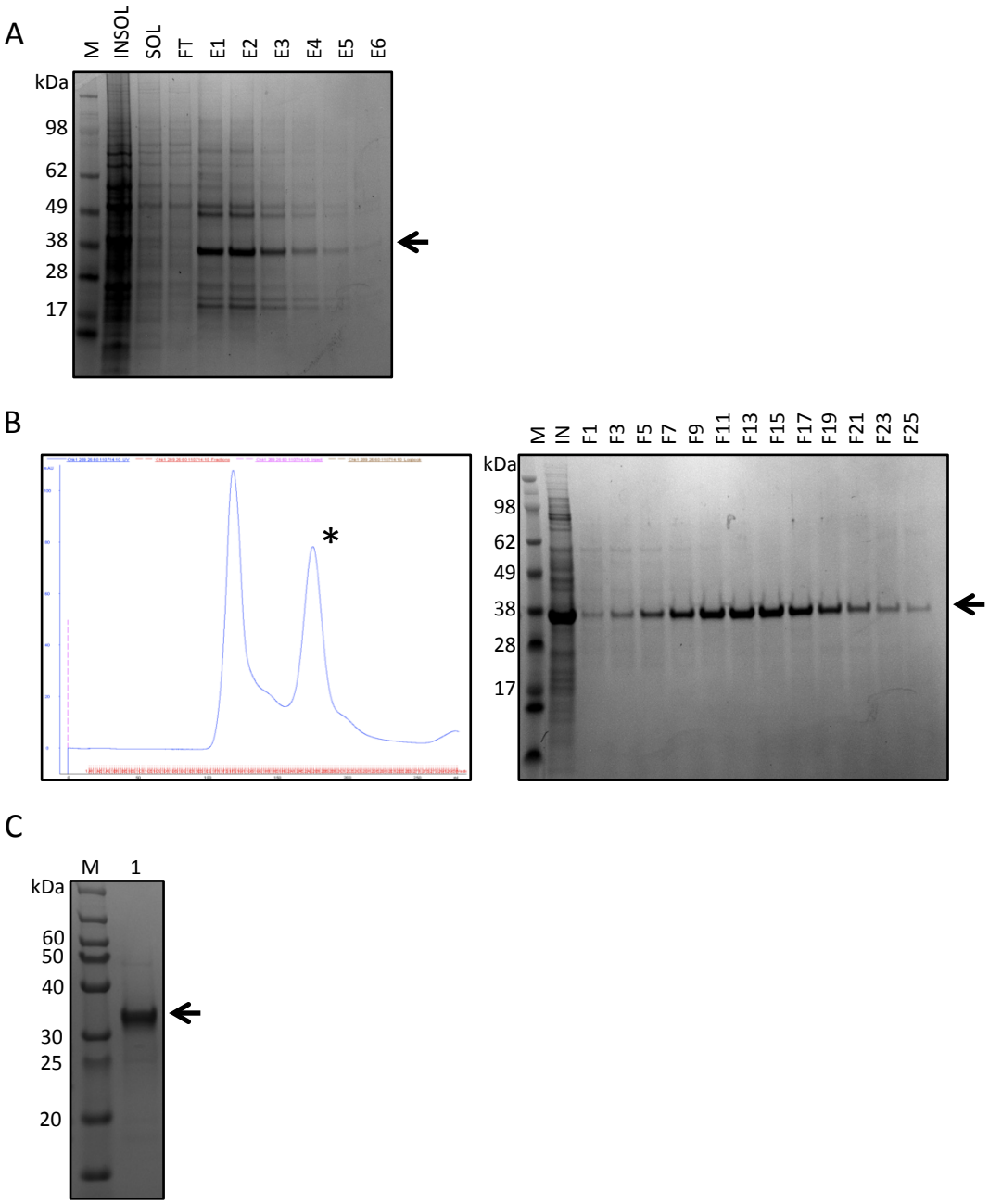
Appendix 3.1 PCR amplification, sub-cloning and confirmation of bacmid transposition for different expression constructs of human Chk1.
(**A**) Analysis of the colony PCR reactions for 1=His-Chk1 and 2=Chk1-KD$^{1-270}$-His, and a diagnostic restriction digest using BamHI and HindIII confirming sub-cloning of Chk1-KD$^{1-289}$-His into pFastBac1. (**B**) PCR analysis of recombinant bacmids, confirming successful transposition 1=His-Chk1, 2=Chk1-KD$^{1-270}$-His, 3=Chk1-KD$^{1-289}$-His. 1% (w/v) agarose gel in 1x TAE buffer. M=Hyperladder I DNA Marker.



Appendix 3.2 Chk1 *Sf*9 cell infection and expression testing.
(**A**) Resuspended P1 infected *Sf*9 cells analysed using a colourmetric anti-His western blot (anti-His primary and an AP-conjugated secondary antibodies). M=molecular mass marker, 1=His-Chk1, 2=Chk1-KD$^{1-270}$-His, 3=Chk1-KD$^{1-289}$-His. (**B**) Resuspended *Sf*9 cells from an expression culture infected with Chk1-KD$^{1-270}$-His baculovirus, analysed using colourmetric anti-His western blot (anti-His primary and an AP-conjugated secondary antibodies). M=molecular mass marker, 1=24 hours, 2=48 hours, 3=62 hours.

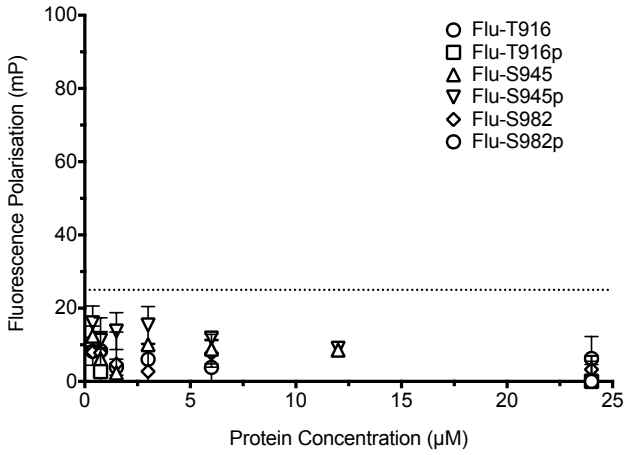


Appendix 3.3 Purification of His-Chk1: IMAC and SEC steps.
(**A**) SDS-PAGE analysis of Sf9 cell lysate, and IMAC purification steps. M=molecular mass marker, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, WASH=wash fraction, E1 to E3=successive elution fractions. (**B**) SEC using a 16/60 SD200 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, X=fractions from the void volume, F=indicated fraction. (**C**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisk indicates the elution peak. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 3.4 Purification of expression construct Chk1-KD$^{1-289}$-His: IMAC and SEC steps.

(**A**) SDS-PAGE analysis of Sf9 cell lysate, and IMAC purification steps. M=molecular mass marker, INSOL=insoluble fraction, SOL=soluble fraction, FT=column flow-through, E1 to E6=successive elution fractions. (**B**) SEC using a 26/60 SD75 column. (Left) representative chromatograph, showing UV absorbance at 280 nm (blue line) and collected fractions. (Right) SDS-PAGE analysis of selected fractions. IN=input, X=fractions from the void volume, F=indicated fraction. (**C**) SDS-PAGE analysis of purified protein. The arrow indicates the migration position of the full-length recombinant protein, and the asterisk indicates the elution peak. 4-12% Bis-Tris SDS-PAGE gel, stained with Instant Blue.



Appendix 3.5 Recombinant human Chk2-KD does not bind CKB motifs.
Binding of human Chk2-KD to 5'-fluorescein synthetic peptides (representing the CKB motifs of human Claspin) was tested by FP. In each case, the indicated peptide was at a final concentration of 100 nM. Data were measured using a POLARstar Omega multimode microplate reader (BMG Labtech). Each data point is the mean of three individual experiments and error bars represent one standard deviation.

References

ALABERT, C., BIANCO, J. N. & PASERO, P. 2009. Differential regulation of homologous recombination at DNA breaks and replication forks by the Mrc1 branch of the S-phase checkpoint. *EMBO J,* 28**,** 1131-41.

ALCASABAS, A. A., OSBORN, A. J., BACHANT, J., HU, F., WERLER, P. J., BOUSSET, K., FURUYA, K., DIFFLEY, J. F., CARR, A. M. & ELLEDGE, S. J. 2001. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nat Cell Biol,* 3**,** 958-65.

ALLERA-MOREAU, C., ROUQUETTE, I., LEPAGE, B., OUMOUHOU, N., WALSCHAERTS, M., LECONTE, E., SCHILLING, V., GORDIEN, K., BROUCHET, L., DELISLE, M. B., MAZIERES, J., HOFFMANN, J. S., PASERO, P. & CAZAUX, C. 2012. DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis,* 1**,** e30.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ALVER, B., KELLY, M. K. & KIRKPATRICK, D. T. 2013. Novel checkpoint pathway organization promotes genome stability in stationary-phase yeast cells. *Mol Cell Biol,* 33**,** 457-72.

AN, Y., MERESSE, P., MAS, P. J. & HART, D. J. 2011. CoESPRIT: a library-based construct screening method for identification and expression of soluble protein complexes. *PLoS One,* 6**,** e16261.

ANGELINI, A., TOSI, T., MAS, P., ACAJJAOUI, S., ZANOTTI, G., TERRADOT, L. & HART, D. J. 2009. Expression of Helicobacter pylori CagA domains by library-based construct screening. *FEBS J,* 276**,** 816-24.

ARIA, V., DE FELICE, M., DI PERNA, R., UNO, S., MASAI, H., SYVAOJA, J. E., VAN LOON, B., HUBSCHER, U. & PISANI, F. M. 2013. The human Tim-Tipin complex interacts directly with DNA polymerase epsilon and stimulates its synthetic activity. *J Biol Chem,* 288**,** 12742-52.

ARLANDER, S. J., FELTS, S. J., WAGNER, J. M., STENSGARD, B., TOFT, D. O. & KARNITZ, L. M. 2006. Chaperoning checkpoint kinase 1 (Chk1), an Hsp90 client, with purified chaperones. *J Biol Chem,* 281**,** 2989-98.

ASIAL, I., CHENG, Y. X., ENGMAN, H., DOLLHOPF, M., WU, B., NORDLUND, P. & CORNVIK, T. 2013. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat Commun,* 4**,** 2901.

AZE, A., ZHOU, J. C., COSTA, A. & COSTANZO, V. 2013. DNA replication and homologous recombination factors: acting together to maintain genome stability. *Chromosoma,* 122**,** 401-13.

BABU, M. M., VAN DER LEE, R., DE GROOT, N. S. & GSPONER, J. 2011. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol,* 21**,** 432-40.

BANDO, M., KATOU, Y., KOMATA, M., TANAKA, H., ITOH, T., SUTANI, T. & SHIRAHIGE, K. 2009. Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. *J Biol Chem,* 284**,** 34355-65.

BARTEK, J. & LUKAS, J. 2007. DNA damage checkpoints: from initiation to recovery or adaptation. *Curr Opin Cell Biol,* 19**,** 238-45.

BASSERMANN, F., FRESCAS, D., GUARDAVACCARO, D., BUSINO, L., PESCHIAROLI, A. & PAGANO, M. 2008. The Cdc14B-Cdh1-Plk1 axis controls the G2 DNA-damage-response checkpoint. *Cell,* 134**,** 256-67.

BECKETT, D., KOVALEVA, E. & SCHATZ, P. J. 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci,* 8**,** 921-9.

BELL, S. P. & DUTTA, A. 2002. DNA replication in eukaryotic cells. *Annu Rev Biochem,* 71**,** 333-74.

BENEVOLO, M., MUSIO, A., VOCATURO, A., DONA, M. G., ROLLO, F., TERRENATO, I., CAROSI, M., PESCARMONA, E., VOCATURO, G. & MOTTOLESE, M. 2012. Claspin as a biomarker of human papillomavirus-related high grade lesions of uterine cervix. *J Transl Med,* 10**,** 132.

BENNETT, L. N. & CLARKE, P. R. 2006. Regulation of Claspin degradation by the ubiquitin-proteosome pathway during the cell cycle and in response to ATR-dependent checkpoint activation. *FEBS Lett,* 580**,** 4176-81.

BENNETT, L. N., LARKIN, C., GILLESPIE, D. A. & CLARKE, P. R. 2008. Claspin is phosphorylated in the Chk1-binding domain by a kinase distinct from Chk1. *Biochem Biophys Res Commun,* 369**,** 973-6.

BERENS, T. J. & TOCZYSKI, D. P. 2012. Colocalization of Mec1 and Mrc1 is sufficient for Rad53 phosphorylation in vivo. *Mol Biol Cell,* 23**,** 1058-67.

BOLANOS-GARCIA, V. M. & DAVIES, O. R. 2006. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta,* 1760**,** 1304-13.

BONNEAU, F., LENHERR, E. D., PENA, V., HART, D. J. & SCHEFFZEK, K. 2009. Solubility survey of fragments of the neurofibromatosis type 1 protein neurofibromin. *Protein Expr Purif,* 65**,** 30-7.

BRENNAN, R. G. & MATTHEWS, B. W. 1989. The helix-turn-helix DNA binding motif. *J Biol Chem,* 264**,** 1903-6.

BRODERICK, R., RAINEY, M. D., SANTOCANALE, C. & NASHEUER, H. P. 2013. Cell cycle-dependent formation of Cdc45-Claspin complexes in human cells is compromized by UV-mediated DNA damage. *FEBS J,* 280**,** 4888-902.

BRONDELLO, J. M., DUCOMMUN, B., FERNANDEZ, A. & LAMB, N. J. 2007. Linking PCNA-dependent replication and ATR by human Claspin. *Biochem Biophys Res Commun,* 354**,** 1028-33.

BUCHAN, D. W., MINNECI, F., NUGENT, T. C., BRYSON, K. & JONES, D. T. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res,* 41**,** W349-57.

BUCK, M. 1998. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys,* 31**,** 297-355.

CABANTOUS, S., NGUYEN, H. B., PEDELACQ, J. D., KORAICHI, F., CHAUDHARY, A., GANGULY, K., LOCKARD, M. A., FAVRE, G., TERWILLIGER, T. C. & WALDO, G. S. 2013. A new protein-protein interaction sensor based on tripartite split-GFP association. *Sci Rep,* 3**,** 2854.

CABANTOUS, S., PEDELACQ, J. D., MARK, B. L., NARANJO, C., TERWILLIGER, T. C. & WALDO, G. S. 2005a. Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis. *J Struct Funct Genomics,* 6**,** 113-9.

CABANTOUS, S., ROGERS, Y., TERWILLIGER, T. C. & WALDO, G. S. 2008. New molecular reporters for rapid protein folding assays. *PLoS One,* 3**,** e2387.

CABANTOUS, S., TERWILLIGER, T. C. & WALDO, G. S. 2005b. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol,* 23**,** 102-7.

CABANTOUS, S. & WALDO, G. S. 2006. In vivo and in vitro protein solubility assays using split GFP. *Nat Methods,* 3**,** 845-54.

CALZADA, A., HODGSON, B., KANEMAKI, M., BUENO, A. & LABIB, K. 2005. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev,* 19**,** 1905-19.

CAPARELLI, M. L. & O'CONNELL, M. J. 2013. Regulatory motifs in Chk1. *Cell Cycle,* 12**,** 916-22.

CAPASSO, H., PALERMO, C., WAN, S., RAO, H., JOHN, U. P., O'CONNELL, M. J. & WALWORTH, N. C. 2002. Phosphorylation activates Chk1 and is required for checkpoint-mediated cell cycle arrest. *J Cell Sci,* 115**,** 4555-64.

CARR, A. M. 2002. DNA structure dependent checkpoints as regulators of DNA repair. *DNA Repair (Amst),* 1**,** 983-94.

CHEN, E. S., HOCH, N. C., WANG, S. C., PELLICIOLI, A., HEIERHORST, J. & TSAI, M. D. 2014. Use of quantitative mass spectrometric analysis to elucidate the mechanisms of phospho-priming and auto-activation of the checkpoint kinase Rad53 in vivo. *Mol Cell Proteomics,* 13**,** 551-65.

CHEN, M. J., SHIMADA, T., MOULTON, A. D., CLINE, A., HUMPHRIES, R. K., MAIZEL, J. & NIENHUIS, A. W. 1984. The functional human dihydrofolate reductase gene. *J Biol Chem,* 259**,** 3933-43.

CHEN, P., LUO, C., DENG, Y., RYAN, K., REGISTER, J., MARGOSIAK, S., TEMPCZYK-RUSSELL, A., NGUYEN, B., MYERS, P., LUNDGREN, K., KAN, C. C. & O'CONNOR, P. M. 2000. The 1.7 A crystal structure of human cell cycle checkpoint kinase Chk1: implications for Chk1 regulation. *Cell,* 100**,** 681-92.

CHEN, S. H. & ZHOU, H. 2009. Reconstitution of Rad53 activation by Mec1 through adaptor protein Mrc1. *J Biol Chem,* 284**,** 18593-604.

CHINI, C. C. & CHEN, J. 2003. Human claspin is required for replication checkpoint control. *J Biol Chem,* 278**,** 30057-62.

CHINI, C. C., WOOD, J. & CHEN, J. 2006. Chk1 is required to maintain claspin stability. *Oncogene,* 25**,** 4165-71.

CHO, W. H., KANG, Y. H., AN, Y. Y., TAPPIN, I., HURWITZ, J. & LEE, J. K. 2013. Human Tim-Tipin complex affects the biochemical properties of the replicative DNA helicase and DNA polymerases. *Proc Natl Acad Sci U S A,* 110**,** 2523-7.

CHOI, S. H., YANG, H., LEE, S. H., KI, J. H., NAM, D. H. & YOO, H. Y. 2014. TopBP1 and Claspin contribute to the radioresistance of lung cancer brain metastases. *Mol Cancer,* 13**,** 211.

CHOU, D. M. & ELLEDGE, S. J. 2006. Tipin and Timeless form a mutually protective complex required for genotoxic stress resistance and checkpoint function. *Proc Natl Acad Sci U S A,* 103**,** 18143-7.

CHRIST, D. & WINTER, G. 2006. Identification of protein domains by shotgun proteolysis. *J Mol Biol,* 358**,** 364-71.

CLARKE, C. A., BENNETT, L. N. & CLARKE, P. R. 2005. Cleavage of claspin by caspase-7 during apoptosis inhibits the Chk1 pathway. *J Biol Chem,* 280**,** 35337-45.

CLARKE, C. A. & CLARKE, P. R. 2005. DNA-dependent phosphorylation of Chk1 and Claspin in a human cell-free system. *Biochem J,* 388**,** 705-12.

COLE, J. L., LARY, J. W., P. MOODY, T. & LAUE, T. M. 2008. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. 84**,** 143-179.

CONCEPCION, J., WITTE, K., WARTCHOW, C., CHOO, S., YAO, D., PERSSON, H., WEI, J., LI, P., HEIDECKER, B., MA, W., VARMA, R., ZHAO, L. S., PERILLAT, D., CARRICATO, G., RECKNOR, M., DU, K., HO, H., ELLIS, T., GAMEZ, J., HOWES, M., PHI-WILSON, J., LOCKARD, S., ZUK, R. & TAN, H. 2009. Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb Chem High Throughput Screen,* 12**,** 791-800.

CONSORTIUM, T. U. 2015. UniProt: a hub for protein information. *Nucleic Acids Res,* 43**,** D204-12.

CORNVIK, T., DAHLROTH, S. L., MAGNUSDOTTIR, A., FLODIN, S., ENGVALL, B., LIEU, V., EKBERG, M. & NORDLUND, P. 2006. An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. *Proteins,* 65**,** 266-73.

CORNVIK, T., DAHLROTH, S. L., MAGNUSDOTTIR, A., HERMAN, M. D., KNAUST, R., EKBERG, M. & NORDLUND, P. 2005. Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. *Nat Methods,* 2**,** 507-9.

CORTESE, M. S., UVERSKY, V. N. & DUNKER, A. K. 2008. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol,* 98**,** 85-106.

DAHLROTH, S. L., LIEU, V., HAAS, J. & NORDLUND, P. 2009. Screening colonies of pooled ORFeomes (SCOOP): a rapid and efficient strategy for expression screening ORFeomes in Escherichia coli. *Protein Expr Purif,* 68**,** 121-7.

DAHLROTH, S. L., NORDLUND, P. & CORNVIK, T. 2006. Colony filtration blotting for screening soluble expression in Escherichia coli. *Nat Protoc,* 1**,** 253-8.

DAYNE, D. 2012. BioLayer Interferometry (BLI) - How does it work? *ForteBio Interaction,* 5.

DESHPANDE, A. M., IVANOVA, I. G., RAYKOV, V., XUE, Y. & MARINGELE, L. 2011. Polymerase epsilon is required to maintain replicative senescence. *Mol Cell Biol,* 31**,** 1637-45.

DORN, M., MB, E. S., BURIOL, L. S. & LAMB, L. C. 2014. Three-dimensional protein structure prediction: Methods and computational strategies. *Comput Biol Chem,* 53PB**,** 251-276.

DUCH, A., DE NADAL, E. & POSAS, F. 2012. The p38 and Hog1 SAPKs control cell cycle progression in response to environmental stresses. *FEBS Lett,* 586**,** 2925-31.

DUNKER, A. K., BONDOS, S. E., HUANG, F. & OLDFIELD, C. J. 2015. Intrinsically disordered proteins and multicellular organisms. *Semin Cell Dev Biol,* 37C**,** 44-55.

DYSON, H. J. 2011. Expanding the proteome: disordered and alternatively folded proteins. *Q Rev Biophys,* 44**,** 467-518.

DYSON, M. R., PERERA, R. L., SHADBOLT, S. P., BIDERMAN, L., BROMEK, K., MURZINA, N. V. & MCCAFFERTY, J. 2008. Identification of soluble protein fragments by gene fragmentation and genetic selection. *Nucleic Acids Res,* 36**,** e51.

EDMAN, P. 1960. Phenylthiohydantoins in protein analysis. *Ann N Y Acad Sci,* 88**,** 602-10.

EMBL. 2015. *Protein Expression and Purification Core Facility* [Online]. Available: https://http://www.embl.de/pepcore/pepcore_services/index.html [Accessed March 2015].

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2126-32.

ENDICOTT, J. A., NOBLE, M. E. & JOHNSON, L. N. 2012. The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem,* 81**,** 587-613.

ERKKO, H., PYLKAS, K., KARPPINEN, S. M. & WINQVIST, R. 2008. Germline alterations in the CLSPN gene in breast cancer families. *Cancer Lett,* 261**,** 93-7.

ERRICO, A., COSENTINO, C., RIVERA, T., LOSADA, A., SCHWOB, E., HUNT, T. & COSTANZO, V. 2009. Tipin/Tim1/And1 protein complex promotes Pol alpha chromatin binding and sister chromatid cohesion. *EMBO J,* 28**,** 3681-92.

ERRICO, A. & COSTANZO, V. 2012. Mechanisms of replication fork protection: a safeguard for genome stability. *Crit Rev Biochem Mol Biol,* 47**,** 222-35.

ERRICO, A., COSTANZO, V. & HUNT, T. 2007. Tipin is required for stalled replication forks to resume DNA replication after removal of aphidicolin in Xenopus egg extracts. *Proc Natl Acad Sci U S A,* 104**,** 14929-34.

FAUSTRUP, H., BEKKER-JENSEN, S., BARTEK, J., LUKAS, J. & MAILAND, N. 2009. USP7 counteracts SCFbetaTrCP- but not APCCdh1-mediated proteolysis of Claspin. *J Cell Biol,* 184**,** 13-9.

FOCARELLI, M. L., SOZA, S., MANNINI, L., PAULIS, M., MONTECUCCO, A. & MUSIO, A. 2009. Claspin inhibition leads to fragile site expression. *Genes Chromosomes Cancer,* 48**,** 1083-90.

FOLOPPE, N., FISHER, L. M., HOWES, R., KIERSTAN, P., POTTER, A., ROBERTSON, A. G. & SURGENOR, A. E. 2005. Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein-ligand affinity. *J Med Chem,* 48**,** 4332-45.

FONG, C. M., ARUMUGAM, A. & KOEPP, D. M. 2013. The Saccharomyces cerevisiae F-box protein Dia2 is a mediator of S-phase checkpoint recovery from DNA damage. *Genetics,* 193**,** 483-99.

FRANKE, D. & SVERGUN, D. I. 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst,* 42**,** 342-346.

FREUDENREICH, C. H. & LAHIRI, M. 2004. Structure-forming CAG/CTG repeat sequences are sensitive to breakage in the absence of Mrc1 checkpoint function and S-phase checkpoint signaling: implications for trinucleotide repeat expansion diseases. *Cell Cycle,* 3**,** 1370-4.

GALLINA, I., COLDING, C., HENRIKSEN, P., BELI, P., NAKAMURA, K., OFFMAN, J., MATHIASEN, D. P., SILVA, S., HOFFMANN, E., GROTH, A., CHOUDHARY, C. & LISBY, M. 2015. Cmr1/WDR76 defines a nuclear genotoxic stress body linking genome integrity and protein quality control. *Nat Commun,* 6**,** 6533.

GAMBUS, A., JONES, R. C., SANCHEZ-DIAZ, A., KANEMAKI, M., VAN DEURSEN, F., EDMONDSON, R. D. & LABIB, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol,* 8**,** 358-66.

GAMBUS, A., VAN DEURSEN, F., POLYCHRONOPOULOS, D., FOLTMAN, M., JONES, R. C., EDMONDSON, R. D., CALZADA, A. & LABIB, K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J,* 28**,** 2992-3004.

GANZINELLI, M., MARIANI, P., CATTANEO, D., FOSSATI, R., FRUSCIO, R., CORSO, S., RICCI, F., BROGGINI, M. & DAMIA, G. 2011. Expression of DNA repair genes in ovarian cancer samples: biological and clinical considerations. *Eur J Cancer,* 47**,** 1086-94.

GAO, D., INUZUKA, H., KORENJAK, M., TSENG, A., WU, T., WAN, L., KIRSCHNER, M., DYSON, N. & WEI, W. 2009. Cdh1 regulates cell cycle through modulating the claspin/Chk1 and the Rb/E2F1 pathways. *Mol Biol Cell,* 20**,** 3305-16.

GAST, K., ZIRWER, D., MULLER-FROHNE, M. & DAMASCHUN, G. 1999. Trifluoroethanol-induced conformational transitions of proteins: insights gained from the differences between alpha-lactalbumin and ribonuclease A. *Protein Sci,* 8**,** 625-34.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., R.D., A. & BAIROCH, A. 2005. John M. Walker (ed): The Proteomics Protocol Handbook. *Humana Press***,** 571-607.

GATEI, M., SLOPER, K., SORENSEN, C., SYLJUASEN, R., FALCK, J., HOBSON, K., SAVAGE, K., LUKAS, J., ZHOU, B. B., BARTEK, J. & KHANNA, K. K. 2003. Ataxia-telangiectasia-mutated (ATM) and NBS1-dependent phosphorylation of Chk1 on Ser-317 in response to ionizing radiation. *J Biol Chem,* 278**,** 14806-11.

GOLD, D. A. & DUNPHY, W. G. 2010. Drf1-dependent kinase interacts with Claspin through a conserved protein motif. *J Biol Chem,* 285**,** 12638-46.

GOTTER, A. L., SUPPA, C. & EMANUEL, B. S. 2007. Mammalian TIMELESS and Tipin are evolutionarily conserved replication fork-associated factors. *J Mol Biol,* 366**,** 36-52.

GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res,* 38**,** W695-9.

GRANT, S. K. 2009. Therapeutic protein kinase inhibitors. *Cell Mol Life Sci,* 66**,** 1163-77.

GRASLUND, S., NORDLUND, P., WEIGELT, J., HALLBERG, B. M., BRAY, J., GILEADI, O., KNAPP, S., OPPERMANN, U., ARROWSMITH, C., HUI, R., MING, J., DHE-PAGANON, S., PARK, H. W., SAVCHENKO, A., YEE, A., EDWARDS, A., VINCENTELLI, R., CAMBILLAU, C., KIM, R., KIM, S. H., RAO, Z., SHI, Y., TERWILLIGER, T. C., KIM, C. Y., HUNG, L. W., WALDO, G. S., PELEG, Y., ALBECK, S., UNGER, T., DYM, O., PRILUSKY, J., SUSSMAN, J. L., STEVENS, R. C., LESLEY, S. A., WILSON, I. A., JOACHIMIAK, A., COLLART, F., DEMENTIEVA, I., DONNELLY, M. I., ESCHENFELDT, W. H., KIM, Y., STOLS, L., WU, R., ZHOU, M., BURLEY, S. K., EMTAGE, J. S., SAUDER, J. M., THOMPSON, D., BAIN, K., LUZ, J., GHEYI, T., ZHANG, F., ATWELL, S., ALMO, S. C., BONANNO, J. B., FISER, A., SWAMINATHAN, S., STUDIER, F. W., CHANCE, M. R., SALI, A., ACTON, T. B., XIAO, R., ZHAO, L., MA, L. C., HUNT, J. F., TONG, L., CUNNINGHAM, K., INOUYE, M., ANDERSON, S., JANJUA, H., SHASTRY, R., HO, C. K., WANG, D., WANG, H., JIANG, M., MONTELIONE, G. T., STUART, D. I., OWENS, R. J., DAENKE, S., SCHUTZ, A., HEINEMANN, U., YOKOYAMA, S., BUSSOW, K. & GUNSALUS, K. C. 2008a. Protein production and purification. *Nat Methods,* 5**,** 135-46.

GRASLUND, S., SAGEMARK, J., BERGLUND, H., DAHLGREN, L. G., FLORES, A., HAMMARSTROM, M., JOHANSSON, I., KOTENYOVA, T., NILSSON, M., NORDLUND, P. & WEIGELT, J. 2008b. The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expr Purif,* 58**,** 210-21.

GREENFIELD, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc,* 1**,** 2876-90.

GROTHUES, D., CANTOR, C. R. & SMITH, C. L. 1993. PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucleic Acids Res,* 21**,** 1321-2.

GUILLIGAY, D., TARENDEAU, F., RESA-INFANTE, P., COLOMA, R., CREPIN, T., SEHR, P., LEWIS, J., RUIGROK, R. W., ORTIN, J., HART, D. J. & CUSACK, S. 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol,* 15**,** 500-6.

HANKS, S. K. & HUNTER, T. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J,* 9**,** 576-96.

HANKS, S. K. & QUINN, A. M. 1991. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol,* 200**,** 38-62.

HANSEN, J. C., LU, X., ROSS, E. D. & WOODY, R. W. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem,* 281**,** 1853-6.

HART, D. J. & WALDO, G. S. 2013. Library methods for structural biology of challenging proteins and their complexes. *Curr Opin Struct Biol,* 23**,** 403-8.

HAYANO, M., KANOH, Y., MATSUMOTO, S. & MASAI, H. 2011. Mrc1 marks early-firing origins and coordinates timing and efficiency of initiation in fission yeast. *Mol Cell Biol,* 31**,** 2380-91.

HEDDLE, C. & MAZALEYRAT, S. L. 2007. Development of a screening platform for directed evolution using the reef coral fluorescent protein ZsGreen as a solubility reporter. *Protein Eng Des Sel,* 20**,** 327-37.

HODGSON, B., CALZADA, A. & LABIB, K. 2007. Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. *Mol Biol Cell,* 18**,** 3894-902.

HOWLETT, G. J., MINTON, A. P. & RIVAS, G. 2006. Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol,* 10**,** 430-6.

HUANG, X. 2003. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J Biomol Screen,* 8**,** 34-8.

IWANAGA, R., KOMORI, H., ISHIDA, S., OKAMURA, N., NAKAYAMA, K., NAKAYAMA, K. I. & OHTANI, K. 2006. Identification of novel E2F1 target genes regulated in cell cycle-dependent and independent manners. *Oncogene,* 25**,** 1786-98.

JACKSON, J. R., GILMARTIN, A., IMBURGIA, C., WINKLER, J. D., MARSHALL, L. A. & ROSHAK, A. 2000. An indolocarbazole inhibitor of human checkpoint kinase (Chk1) abrogates cell cycle arrest caused by DNA damage. *Cancer Res,* 60**,** 566-72.

JACOBS, S. A., PODELL, E. R., WUTTKE, D. S. & CECH, T. R. 2005. Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. *Protein Sci,* 14**,** 2051-8.

JACQUES, D. A. & TREWHELLA, J. 2010. Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein Sci,* 19**,** 642-57.

JEONG, S. Y., KUMAGAI, A., LEE, J. & DUNPHY, W. G. 2003. Phosphorylated claspin interacts with a phosphate-binding site in the kinase domain of Chk1 during ATR-mediated activation. *J Biol Chem,* 278**,** 46782-8.

JOHNSON, W. C. 1999. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins,* 35**,** 307-12.

JONES, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol,* 292**,** 195-202.

KACZKA, P., WINIEWSKA, M., ZHUKOV, I., REMPOLA, B., BOLEWSKA, K., LOZINSKI, T., EJCHART, A., POZNANSKA, A., WIERZCHOWSKI, K. L. & POZNANSKI, J. 2014. The TFE-induced transient native-like structure of the intrinsically disordered sigma(4)(7)(0) domain of Escherichia coli RNA polymerase. *Eur Biophys J,* 43**,** 581-94.

KALLBERG, M., WANG, H., WANG, S., PENG, J., WANG, Z., LU, H. & XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc,* 7**,** 1511-22.

KASTAN, M. B. & BARTEK, J. 2004. Cell-cycle checkpoints and cancer. *Nature,* 432**,** 316-23.

KATOU, Y., KANOH, Y., BANDO, M., NOGUCHI, H., TANAKA, H., ASHIKARI, T., SUGIMOTO, K. & SHIRAHIGE, K. 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature,* 424**,** 1078-83.

KATSURAGI, Y. & SAGATA, N. 2004. Regulation of Chk1 kinase by autoinhibition and ATR-mediated phosphorylation. *Mol Biol Cell,* 15**,** 1680-9.

KAWAGUCHI, M., TERAI, T., UTATA, R., KATO, M., TSUGANEZAWA, K., TANAKA, A., KOJIMA, H., OKABE, T. & NAGANO, T. 2008. Development of a novel fluorescent probe for fluorescence correlation spectroscopic detection of kinase inhibitors. *Bioorg Med Chem Lett,* 18**,** 3752-5.

KAWASAKI, M. & INAGAKI, F. 2001. Random PCR-based screening for soluble domains using green fluorescent protein. *Biochem Biophys Res Commun,* 280**,** 842-4.

KAY, L. E. 2005. NMR studies of protein structure and dynamics. *J Magn Reson,* 173**,** 193-207.

KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

KELLY, S. M., JESS, T. J. & PRICE, N. C. 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta,* 1751**,** 119-39.

KEMP, M. G., AKAN, Z., YILMAZ, S., GRILLO, M., SMITH-ROE, S. L., KANG, T. H., CORDEIRO-STONE, M., KAUFMANN, W. K., ABRAHAM, R. T., SANCAR, A. & UNSAL-KACMAZ, K. 2010. Tipin-replication protein A interaction mediates Chk1 phosphorylation by ATR in response to genotoxic stress. *J Biol Chem,* 285**,** 16562-71.

KENNELLY, P. J. 2002. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol Lett,* 206**,** 1-8.

KENNETH, N. S., MUDIE, S. & ROCHA, S. 2010. IKK and NF-kappaB-mediated regulation of Claspin impacts on ATR checkpoint function. *EMBO J,* 29**,** 2966-78.

KIM, J. M., KAKUSHO, N., YAMADA, M., KANOH, Y., TAKEMOTO, N. & MASAI, H. 2008. Cdc7 kinase mediates Claspin phosphorylation in DNA replication checkpoint. *Oncogene,* 27**,** 3475-82.

KING, D. A., HALL, B. E., IWAMOTO, M. A., WIN, K. Z., CHANG, J. F. & ELLENBERGER, T. 2006. Domain structure and protein interactions of the silent information regulator Sir3 revealed by screening a nested deletion library of protein fragments. *J Biol Chem,* 281**,** 20107-19.

KLOCK, H. E., KOESEMA, E. J., KNUTH, M. W. & LESLEY, S. A. 2008. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins,* 71**,** 982-94.

KNIGHTON, D. R., ZHENG, J. H., TEN EYCK, L. F., ASHFORD, V. A., XUONG, N. H., TAYLOR, S. S. & SOWADSKI, J. M. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science,* 253**,** 407-14.

KOGANTI, S., HUI-YUEN, J., MCALLISTER, S., GARDNER, B., GRASSER, F., PALENDIRA, U., TANGYE, S. G., FREEMAN, A. F. & BHADURI-MCINTOSH, S. 2014. STAT3 interrupts ATR-Chk1 signaling to allow oncovirus-mediated cell proliferation. *Proc Natl Acad Sci U S A,* 111**,** 4946-51.

KOMATA, M., BANDO, M., ARAKI, H. & SHIRAHIGE, K. 2009. The direct binding of Mrc1, a checkpoint mediator, to Mcm6, a replication helicase, is essential for the replication checkpoint against methyl methanesulfonate-induced stress. *Mol Cell Biol,* 29**,** 5008-19.

KONAREV, P. V., PETOUKHOV, M. V., VOLKOV, V. V. & SVERGUN, D. I. 2006. ATSAS2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography,* 39**,** 277-286.

KOSOY, A. & O'CONNELL, M. J. 2008. Regulation of Chk1 by its C-terminal domain. *Mol Biol Cell,* 19**,** 4546-53.

KUMAGAI, A. & DUNPHY, W. G. 2000. Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts. *Mol Cell,* 6**,** 839-49.

KUMAGAI, A. & DUNPHY, W. G. 2003. Repeated phosphopeptide motifs in Claspin mediate the regulated binding of Chk1. *Nat Cell Biol,* 5**,** 161-5.

KUMAGAI, A., GUO, Z., EMAMI, K. H., WANG, S. X. & DUNPHY, W. G. 1998. The Xenopus Chk1 protein kinase mediates a caffeine-sensitive pathway of checkpoint control in cell-free extracts. *J Cell Biol,* 142**,** 1559-69.

KUMAGAI, A., KIM, S. M. & DUNPHY, W. G. 2004. Claspin and the activated form of ATR-ATRIP collaborate in the activation of Chk1. *J Biol Chem,* 279**,** 49599-608.

KURIHARA, T., SAKURAI, E., TOYOMOTO, M., KII, I., KAWAMOTO, D., ASADA, T., TANABE, T., YOSHIMURA, M., HAGIWARA, M. & MIYATA, A. 2014. Alleviation of behavioral hypersensitivity in mouse models of inflammatory pain with two structurally different casein kinase 1 (CK1) inhibitors. *Mol Pain,* 10**,** 17.

LAMERS, M. B., ANTSON, A. A., HUBBARD, R. E., SCOTT, R. K. & WILLIAMS, D. H. 1999. Structure of the protein tyrosine kinase domain of C-terminal Src kinase (CSK) in complex with staurosporine. *J Mol Biol,* 285**,** 713-25.

LEA, W. A. & SIMEONOV, A. 2011. Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov,* 6**,** 17-32.

LECLERE, A. R., YANG, J. K. & KIRKPATRICK, D. T. 2013. The role of CSM3, MRC1, and TOF1 in minisatellite stability and large loop DNA repair during meiosis in yeast. *Fungal Genet Biol,* 50**,** 33-43.

LEE, J., GOLD, D. A., SHEVCHENKO, A., SHEVCHENKO, A. & DUNPHY, W. G. 2005. Roles of replication fork-interacting and Chk1-activating domains from Claspin in a DNA replication checkpoint response. *Mol Biol Cell,* 16**,** 5269-82.

LEE, J., KUMAGAI, A. & DUNPHY, W. G. 2003. Claspin, a Chk1-regulatory protein, monitors DNA replication on chromatin independently of RPA, ATR, and Rad17. *Mol Cell,* 11**,** 329-40.

LEMAN, A. R., NOGUCHI, C., LEE, C. Y. & NOGUCHI, E. 2010. Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. *J Cell Sci,* 123**,** 660-70.

LEMAN, A. R. & NOGUCHI, E. 2012. Local and global functions of Timeless and Tipin in replication fork protection. *Cell Cycle,* 11**,** 3945-55.

LEMAN, A. R. & NOGUCHI, E. 2013. The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication. *Genes (Basel),* 4**,** 1-32.

LEWIS, T. E., SILLITOE, I., ANDREEVA, A., BLUNDELL, T. L., BUCHAN, D. W., CHOTHIA, C., CUFF, A., DANA, J. M., FILIPPIS, I., GOUGH, J., HUNTER, S., JONES, D. T., KELLEY, L. A., KLEYWEGT, G. J., MINNECI, F., MITCHELL, A., MURZIN, A. G., OCHOA-MONTANO, B., RACKHAM, O. J., SMITH, J., STERNBERG, M. J., VELANKAR, S., YEATS, C. & ORENGO, C. 2013. Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. *Nucleic Acids Res,* 41**,** D499-507.

LIN, S. & CRONAN, J. E. 2011. Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst,* 7**,** 1811-21.

LIN, S. Y., LI, K., STEWART, G. S. & ELLEDGE, S. J. 2004. Human Claspin works with BRCA1 to both positively and negatively regulate cell proliferation. *Proc Natl Acad Sci U S A,* 101**,** 6484-9.

LIN, Y. F., SHIH, H. Y., SHANG, Z., MATSUNAGA, S. & CHEN, B. P. 2014. DNA-PKcs is required to maintain stability of Chk1 and Claspin for optimal replication stress response. *Nucleic Acids Res,* 42**,** 4463-73.

LINDAHL, T., PRIGENT, C., BARNES, D. E., LEHMANN, A. R., SATOH, M. S., ROBERTS, E., NASH, R. A., ROBINS, P. & DALY, G. 1993. DNA joining in mammalian cells. *Cold Spring Harb Symp Quant Biol,* 58**,** 619-24.

LINDSEY-BOLTZ, L. A., KEMP, M. G., CAPP, C. & SANCAR, A. 2015. RHINO forms a stoichiometric complex with the 9-1-1 checkpoint clamp and mediates ATR-Chk1 signaling. *Cell Cycle,* 14**,** 99-108.

LINDSEY-BOLTZ, L. A., SERCIN, O., CHOI, J. H. & SANCAR, A. 2009. Reconstitution of human claspin-mediated phosphorylation of Chk1 by the ATR (ataxia telangiectasia-mutated and rad3-related) checkpoint kinase. *J Biol Chem,* 284**,** 33107-14.

LISTWAN, P., TERWILLIGER, T. C. & WALDO, G. S. 2009. Automated, high-throughput platform for protein solubility screening using a split-GFP system. *J Struct Funct Genomics,* 10**,** 47-55.

LITTLER, E. 2010. Combinatorial Domain Hunting: solving problems in protein expression. *Drug Discov Today,* 15**,** 461-7.

LIU, G., CHEN, X., GAO, Y., LEWIS, T., BARTHELEMY, J. & LEFFAK, M. 2012a. Altered replication in human cells promotes DMPK (CTG)(n) . (CAG)(n) repeat instability. *Mol Cell Biol,* 32**,** 1618-32.

LIU, J. W., BOUCHER, Y., STOKES, H. W. & OLLIS, D. L. 2006a. Improving protein solubility: the use of the Escherichia coli dihydrofolate reductase gene as a fusion reporter. *Protein Expr Purif,* 47**,** 258-63.

LIU, S., BEKKER-JENSEN, S., MAILAND, N., LUKAS, C., BARTEK, J. & LUKAS, J. 2006b. Claspin operates downstream of TopBP1 to direct ATR signaling towards Chk1 activation. *Mol Cell Biol,* 26**,** 6056-64.

LIU, S., SONG, N. & ZOU, L. 2012b. The conserved C terminus of Claspin interacts with Rad9 and promotes rapid activation of Chk1. *Cell Cycle,* 11**,** 2711-6.

LIU, Z. & HUANG, Y. 2014. Advantages of proteins being disordered. *Protein Sci,* 23**,** 539-50.

LOCKARD, M. A., LISTWAN, P., PEDELACQ, J. D., CABANTOUS, S., NGUYEN, H. B., TERWILLIGER, T. C. & WALDO, G. S. 2011. A high-throughput immobilized bead screen for stable proteins and multi-protein complexes. *Protein Eng Des Sel,* 24**,** 565-78.

LOPEZ-GIRONA, A., TANAKA, K., CHEN, X. B., BABER, B. A., MCGOWAN, C. H. & RUSSELL, P. 2001. Serine-345 is required for Rad3-dependent phosphorylation and function of checkpoint kinase Chk1 in fission yeast. *Proc Natl Acad Sci U S A,* 98**,** 11289-94.

LOU, H., KOMATA, M., KATOU, Y., GUAN, Z., REIS, C. C., BUDD, M., SHIRAHIGE, K. & CAMPBELL, J. L. 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Mol Cell,* 32**,** 106-17.

MACEK, B., GNAD, F., SOUFI, B., KUMAR, C., OLSEN, J. V., MIJAKOVIC, I. & MANN, M. 2008. Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol Cell Proteomics,* 7**,** 299-307.

MACLAGAN, K., TOMMASI, R., LAURINE, E., PRODROMOU, C., DRISCOLL, P. C., PEARL, L. H., REICH, S. & SAVVA, R. 2011. A combinatorial method to enable detailed investigation of protein-protein interactions. *Future Med Chem,* 3**,** 271-82.

MAESTRO, B., GALAN, B., ALFONSO, C., RIVAS, G., PRIETO, M. A. & SANZ, J. M. 2013. A new family of intrinsically disordered proteins: structural characterization of the major phasin PhaF from Pseudomonas putida KT2440. *PLoS One,* 8**,** e56904.

MAILAND, N., BEKKER-JENSEN, S., BARTEK, J. & LUKAS, J. 2006. Destruction of Claspin by SCFbetaTrCP restrains Chk1 activation and facilitates recovery from genotoxic stress. *Mol Cell,* 23**,** 307-18.

MAMELY, I., VAN VUGT, M. A., SMITS, V. A., SEMPLE, J. I., LEMMENS, B., PERRAKIS, A., MEDEMA, R. H. & FREIRE, R. 2006. Polo-like kinase-1 controls proteasome-dependent degradation of Claspin during checkpoint recovery. *Curr Biol,* 16**,** 1950-5.

MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T. & SUDARSANAM, S. 2002. The protein kinase complement of the human genome. *Science,* 298**,** 1912-34.

MANOLARIDIS, I., WOJDYLA, J. A., PANJIKAR, S., SNIJDER, E. J., GORBALENYA, A. E., BERGLIND, H., NORDLUND, P., COUTARD, B. & TUCKER, P. A. 2009. Structure of the C-terminal domain of nsp4 from feline coronavirus. *Acta Crystallogr D Biol Crystallogr,* 65**,** 839-46.

MANUAL, I. 2002. Guide to Baculovirus Expression Vector Systems (BEVS) and Insect Cell Culture. *Invitrogen, Life Technologies*.

MAO, D., WACHTER, E. & WALLACE, B. A. 1982. Folding of the mitochondrial proton adenosinetriphosphatase proteolipid channel in phospholipid vesicles. *Biochemistry,* 21**,** 4960-8.

MARION, D. 2013. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics,* 12**,** 3006-25.

MARTI-RENOM, M. A., STUART, A. C., FISER, A., SANCHEZ, R., MELO, F. & SALI, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct,* 29**,** 291-325.

MARTIN, Y., CABRERA, E., AMOEDO, H., HERNANDEZ-PEREZ, S., DOMINGUEZ-KELLY, R. & FREIRE, R. 2015. USP29 controls the stability of checkpoint adaptor Claspin by deubiquitination. *Oncogene,* 34**,** 1058-63.

MARTINEZ MOLINA, D., CORNVIK, T., ESHAGHI, S., HAEGGSTROM, J. Z., NORDLUND, P. & SABET, M. I. 2008. Engineering membrane protein overproduction in Escherichia coli. *Protein Sci,* 17**,** 673-80.

MATSUMOTO, S., SHIMMOTO, M., KAKUSHO, N., YOKOYAMA, M., KANOH, Y., HAYANO, M., RUSSELL, P. & MASAI, H. 2010. Hsk1 kinase and Cdc45 regulate replication stress-induced checkpoint responses in fission yeast. *Cell Cycle,* 9**,** 4627-4637.

MAXWELL, K. L., MITTERMAIER, A. K., FORMAN-KAY, J. D. & DAVIDSON, A. R. 1999. A simple in vivo assay for increased protein solubility. *Protein Sci,* 8**,** 1908-11.

MECHALI, M. 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol,* 11**,** 728-38.

MEIER, C., BROOKINGS, D. C., CESKA, T. A., DOYLE, C., GONG, H., MCMILLAN, D., SAVILLE, G. P., MUSHTAQ, A., KNIGHT, D., REICH, S., PEARL, L. H., POWELL, K. A., SAVVA, R. & ALLEN, R. A. 2012. Engineering human MEK-1 for structural studies: A case study of combinatorial domain hunting. *J Struct Biol,* 177**,** 329-34.

MENG, Z., CAPALBO, L., GLOVER, D. M. & DUNPHY, W. G. 2011. Role for casein kinase 1 in the phosphorylation of Claspin on critical residues necessary for the activation of Chk1. *Mol Biol Cell,* 22**,** 2834-47.

MERTENS, H. D. & SVERGUN, D. I. 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol,* 172**,** 128-41.

MIMURA, S., KOMATA, M., KISHI, T., SHIRAHIGE, K. & KAMURA, T. 2009. SCF(Dia2) regulates DNA replication forks during S-phase in budding yeast. *EMBO J,* 28**,** 3693-705.

MOROHASHI, H., MACULINS, T. & LABIB, K. 2009. The amino-terminal TPR domain of Dia2 tethers SCF(Dia2) to the replisome progression complex. *Curr Biol,* 19**,** 1943-9.

MOYER, S. E., LEWIS, P. W. & BOTCHAN, M. R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A,* 103**,** 10236-41.

MYLONAS, E. & SVERGUN, D. 2007. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography,* 40**,** 245-249.

NADAL, M., MAS, P. J., BLANCO, A. G., ARNAN, C., SOLA, M., HART, D. J. & COLL, M. 2010. Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc Natl Acad Sci U S A,* 107**,** 16078-83.

NAKAYA, R., TAKAYA, J., ONUKI, T., MORITANI, M., NOZAKI, N. & ISHIMI, Y. 2010. Identification of proteins that may directly interact with human RPA. *J Biochem,* 148**,** 539-47.

NASCHBERGER, A., FURNROHR, B. G., DUNZENDORFER-MATT, T., BONAGURA, C. A., WRIGHT, D., SCHEFFZEK, K. & RUPP, B. 2015. Cleaning protocols for crystallization robots: preventing protease contamination. *Acta Crystallogr F Struct Biol Commun,* 71**,** 100-2.

NAYLOR, M. L., LI, J. M., OSBORN, A. J. & ELLEDGE, S. J. 2009. Mrc1 phosphorylation in response to DNA replication stress is required for Mec1 accumulation at the stalled fork. *Proc Natl Acad Sci U S A,* 106**,** 12765-70.

NEDELCHEVA, M. N., ROGUEV, A., DOLAPCHIEV, L. B., SHEVCHENKO, A., TASKOV, H. B., SHEVCHENKO, A., STEWART, A. F. & STOYNOV, S. S. 2005. Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. *J Mol Biol,* 347**,** 509-21.

NG, C. P., LEE, H. C., HO, C. W., AROOZ, T., SIU, W. Y., LAU, A. & POON, R. Y. 2004. Differential mode of regulation of the checkpoint kinases CHK1 and CHK2 by their regulatory domains. *J Biol Chem,* 279**,** 8808-19.

NOGUCHI, E., NOGUCHI, C., DU, L. L. & RUSSELL, P. 2003. Swi1 prevents replication fork collapse and controls checkpoint kinase Cds1. *Mol Cell Biol,* 23**,** 7861-74.

NOGUCHI, E., NOGUCHI, C., MCDONALD, W. H., YATES, J. R., 3RD & RUSSELL, P. 2004. Swi1 and Swi3 are components of a replication fork protection complex in fission yeast. *Mol Cell Biol,* 24**,** 8342-55.

NURSE, P. 1997. Regulation of the eukaryotic cell cycle. *Eur J Cancer,* 33**,** 1002-4.

NYBERG, K. A., MICHELSON, R. J., PUTNAM, C. W. & WEINERT, T. A. 2002. Toward maintaining the genome: DNA damage and replication checkpoints. *Annu Rev Genet,* 36**,** 617-56.

O'CONNELL, M. J., RALEIGH, J. M., VERKADE, H. M. & NURSE, P. 1997. Chk1 is a wee1 kinase in the G2 DNA damage checkpoint inhibiting cdc2 by Y15 phosphorylation. *EMBO J,* 16**,** 545-54.

O'REILLY, D. R. & MILLER, L. K. 1992. Baculovirus expression vectors, a laboratory manual. *WH Freeman and Co., New York*.

OAKES, V., WANG, W., HARRINGTON, B., LEE, W. J., BEAMISH, H., CHIA, K. M., PINDER, A., GOTO, H., INAGAKI, M., PAVEY, S. & GABRIELLI, B. 2014. Cyclin A/Cdk2 regulates Cdh1 and claspin during late S/G2 phase of the cell cycle. *Cell Cycle,* 13**,** 3302-11.

OHREN, J. F., CHEN, H., PAVLOVSKY, A., WHITEHEAD, C., ZHANG, E., KUFFA, P., YAN, C., MCCONNELL, P., SPESSARD, C., BANOTAI, C., MUELLER, W. T., DELANEY, A., OMER, C., SEBOLT-LEOPOLD, J., DUDLEY, D. T., LEUNG, I. K., FLAMME, C., WARMUS, J., KAUFMAN, M., BARRETT, S., TECLE, H. & HASEMANN, C. A. 2004. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat Struct Mol Biol,* 11**,** 1192-7.

OKITA, N., MINATO, S., OHMI, E., TANUMA, S. & HIGAMI, Y. 2012. DNA damage-induced CHK1 autophosphorylation at Ser296 is regulated by an intramolecular mechanism. *FEBS Lett,* 586**,** 3974-9.

OLIVER, A. W., PAUL, A., BOXALL, K. J., BARRIE, S. E., AHERNE, G. W., GARRETT, M. D., MITTNACHT, S. & PEARL, L. H. 2006. Trans-activation of the DNA-damage signalling protein kinase Chk2 by T-loop exchange. *EMBO J,* 25**,** 3179-90.

OMURA, S., IWAI, Y., HIRANO, A., NAKAGAWA, A., AWAYA, J., TSUCHYA, H., TAKAHASHI, Y. & MASUMA, R. 1977. A new alkaloid AM-2282 OF Streptomyces origin. Taxonomy, fermentation, isolation and preliminary characterization. *J Antibiot (Tokyo),* 30**,** 275-82.

OSBORN, A. J. & ELLEDGE, S. J. 2003. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev,* 17**,** 1755-67.

OZA, V., ASHWELL, S., BRASSIL, P., BREED, J., DENG, C., EZHUTHACHAN, J., HAYE, H., HORN, C., JANETKA, J., LYNE, P., NEWCOMBE, N., OTTERBIEN, L., PASS, M., READ, J., ROSWELL, S., SU, M., TOADER, D., YU, D., YU, Y., VALENTINE, A., WEBBORN, P., WHITE, A., ZABLUDOFF, S. & ZHENG, X. 2010. Discovery of a novel class of triazolones as checkpoint kinase inhibitors--hit to lead exploration. *Bioorg Med Chem Lett,* 20**,** 5133-8.

PEARL, L. H., SCHIERZ, A. C., WARD, S. E., AL-LAZIKANI, B. & PEARL, F. M. 2015. Therapeutic opportunities within the DNA damage response. *Nat Rev Cancer,* 15**,** 166-80.

PEDELACQ, J. D., NGUYEN, H. B., CABANTOUS, S., MARK, B. L., LISTWAN, P., BELL, C., FRIEDLAND, N., LOCKARD, M., FAILLE, A., MOUREY, L., TERWILLIGER, T. C. & WALDO, G. S. 2011. Experimental mapping of soluble protein domains using a hierarchical approach. *Nucleic Acids Res,* 39**,** e125.

PEDELACQ, J. D., PILTCH, E., LIONG, E. C., BERENDZEN, J., KIM, C. Y., RHO, B. S., PARK, M. S., TERWILLIGER, T. C. & WALDO, G. S. 2002. Engineering soluble proteins for structural genomics. *Nat Biotechnol,* 20**,** 927-32.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20**,** 3551-67.

PERNOT, P., ROUND, A., BARRETT, R., DE MARIA ANTOLINOS, A., GOBBO, A., GORDON, E., HUET, J., KIEFFER, J., LENTINI, M., MATTENET, M., MORAWE, C., MUELLER-DIECKMANN, C., OHLSSON, S., SCHMID, W., SURR, J., THEVENEAU, P., ZERRAD, L. & MCSWEENEY, S. 2013. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat,* 20**,** 660-4.

PERRIN, M. 1926. Polarisation de la Lumiere de Fluorescence. Vie Moyenne des Molecule dans L′ etat excite. *J Phys Radium***,** 390-390.

PESCHIAROLI, A., DORRELLO, N. V., GUARDAVACCARO, D., VENERE, M., HALAZONETIS, T., SHERMAN, N. E. & PAGANO, M. 2006. SCFbetaTrCP-mediated degradation of Claspin regulates recovery from the DNA replication checkpoint response. *Mol Cell,* 23**,** 319-29.

PETERMANN, E., HELLEDAY, T. & CALDECOTT, K. W. 2008. Claspin promotes normal replication fork rates in human cells. *Mol Biol Cell,* 19**,** 2373-8.

PHILO, J. S. 2006. Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal Biochem,* 354**,** 238-46.

PIEPER, U., WEBB, B. M., BARKAN, D. T., SCHNEIDMAN-DUHOVNY, D., SCHLESSINGER, A., BRABERG, H., YANG, Z., MENG, E. C., PETTERSEN, E. F., HUANG, C. C., DATTA, R. S., SAMPATHKUMAR, P., MADHUSUDHAN, M. S., SJOLANDER, K., FERRIN, T. E., BURLEY, S. K. & SALI, A. 2011. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res,* 39**,** D465-74.

POULSON, F. M. 2002. A brief introduction to NMR spectroscopy of
proteins

PRADE, L., ENGH, R. A., GIROD, A., KINZEL, V., HUBER, R. & BOSSEMEYER, D. 1997. Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. *Structure,* 5**,** 1627-37.

PRADO, F. 2014. Genetic instability is prevented by Mrc1-dependent spatio-temporal separation of replicative and repair activities of homologous recombination: homologous recombination tolerates replicative stress by Mrc1-regulated replication and repair activities operating at S and G2 in distinct subnuclear compartments. *Bioessays,* 36**,** 451-62.

PRODROMOU, C., SAVVA, R. & DRISCOLL, P. C. 2007. DNA fragmentation-based combinatorial approaches to soluble protein expression Part I. Generating DNA fragment libraries. *Drug Discov Today,* 12**,** 931-8.

PROMEGA 2015. ADP-Glo Kinase Assay. *Technical manual 313*.
PROVENCHER, S. W. & GLOCKNER, J. 1981. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry,* 20**,** 33-7.
RAINEY, M. D., HARHEN, B., WANG, G. N., MURPHY, P. V. & SANTOCANALE, C. 2013. Cdc7-dependent and -independent phosphorylation of Claspin in the induction of the DNA replication checkpoint. *Cell Cycle,* 12**,** 1560-8.
RALSTON, G. 1993. Introduction to Analytical Ultracentrifugation. *Fullerton, CA: Academic Press*.
RAWLINGS, A. E., LEVDIKOV, V. M., BLAGOVA, E., COLLEDGE, V. L., MAS, P. J., TUNALEY, J., VAVROVA, L., WILSON, K. S., BARAK, I., HART, D. J. & WILKINSON, A. J. 2010. Expression of soluble, active fragments of the morphogenetic protein SpoIIE from Bacillus subtilis using a library-based construct screen. *Protein Eng Des Sel,* 23**,** 817-25.
RAZIDLO, D. F. & LAHUE, R. S. 2008. Mrc1, Tof1 and Csm3 inhibit CAG.CTG repeat instability by at least two mechanisms. *DNA Repair (Amst),* 7**,** 633-40.
RECEVEUR-BRECHOT, V. & DURAND, D. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci,* 13**,** 55-75.
REICH, S., PUCKEY, L. H., CHEETHAM, C. L., HARRIS, R., ALI, A. A., BHATTACHARYYA, U., MACLAGAN, K., POWELL, K. A., PRODROMOU, C., PEARL, L. H., DRISCOLL, P. C. & SAVVA, R. 2006. Combinatorial Domain Hunting: An effective approach for the identification of soluble protein domains adaptable to high-throughput applications. *Protein Sci,* 15**,** 2356-65.
REIERSEN, H. & REES, A. R. 2000. Trifluoroethanol may form a solvent matrix for assisted hydrophobic interactions between peptide side chains. *Protein Eng,* 13**,** 739-43.
RHIND, N. & RUSSELL, P. 2000. Chk1 and Cds1: linchpins of the DNA damage and replication checkpoint pathways. *J Cell Sci,* 113 (Pt 22)**,** 3889-96.
RHODES, A. 2006. *Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models*, Academic Press.
RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.
RUEGG, U. T. & BURGESS, G. M. 1989. Staurosporine, K-252 and UCN-01: potent but nonspecific inhibitors of protein kinases. *Trends Pharmacol Sci,* 10**,** 218-20.
SANCAR, A., LINDSEY-BOLTZ, L. A., UNSAL-KACMAZ, K. & LINN, S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem,* 73**,** 39-85.
SANCHEZ, R. & ZHOU, M. M. 2009. The role of human bromodomains in chromatin biology and gene transcription. *Curr Opin Drug Discov Devel,* 12**,** 659-65.
SANCHEZ, Y., WONG, C., THOMA, R. S., RICHMAN, R., WU, Z., PIWNICA-WORMS, H. & ELLEDGE, S. J. 1997. Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science,* 277**,** 1497-501.
SAR, F., LINDSEY-BOLTZ, L. A., SUBRAMANIAN, D., CROTEAU, D. L., HUTSELL, S. Q., GRIFFITH, J. D. & SANCAR, A. 2004. Human claspin is a ring-shaped DNA-binding protein with high affinity to branched DNA structures. *J Biol Chem,* 279**,** 39289-95.
SATO, K., SUNDARAMOORTHY, E., RAJENDRA, E., HATTORI, H., JEYASEKHARAN, A. D., AYOUB, N., SCHIESS, R., AEBERSOLD, R., NISHIKAWA, H., SEDUKHINA, A. S., WADA, H., OHTA, T. & VENKITARAMAN, A. R. 2012. A DNA-damage selective role for BRCA1 E3 ligase in claspin ubiquitylation, CHK1 activation, and DNA repair. *Curr Biol,* 22**,** 1659-66.
SAVITSKY, P., BRAY, J., COOPER, C. D., MARSDEN, B. D., MAHAJAN, P., BURGESS-BROWN, N. A. & GILEADI, O. 2010. High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol,* 172**,** 3-13.
SAVVA, R., PRODROMOU, C. & DRISCOLL, P. C. 2007. DNA fragmentation based combinatorial approaches to soluble protein expression Part II: library expression, screening and scale-up. *Drug Discov Today,* 12**,** 939-47.
SCHRODINGER, L. L. C. 2010. *PyMOL. Version 1.3* [Online]. Available: http://www.pymol.org.
SCHUTTELKOPF, A. W. & VAN AALTEN, D. M. 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr,* 60**,** 1355-63.
SCHWEDE, T. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res,* 31**,** 3381-3385.
SCORAH, J. & MCGOWAN, C. H. 2009. Claspin and Chk1 regulate replication fork stability by different mechanisms. *Cell Cycle,* 8**,** 1036-43.

SEMPLE, J. I., SMITS, V. A., FERNAUD, J. R., MAMELY, I. & FREIRE, R. 2007. Cleavage and degradation of Claspin during apoptosis by caspases and the proteasome. *Cell Death Differ,* 14**,** 1433-42.

SERÇIN, Ö. & KEMP, M. G. 2011. Characterization of functional domains in human Claspin. *Cell Cycle,* 10**,** 1599-1606.

SHIKATA, M., ISHIKAWA, F. & KANOH, J. 2007. Tel2 is required for activation of the Mrc1-mediated replication checkpoint. *J Biol Chem,* 282**,** 5346-55.

SHIMMOTO, M., MATSUMOTO, S., ODAGIRI, Y., NOGUCHI, E., RUSSELL, P. & MASAI, H. 2009. Interactions between Swi1-Swi3, Mrc1 and S phase kinase, Hsk1 may regulate cellular responses to stalled replication forks in fission yeast. *Genes Cells,* 14**,** 669-82.

SIBILLE, N. & BERNADO, P. 2012. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem Soc Trans,* 40**,** 955-62.

SIERANT, M. L., ARCHER, N. E. & DAVEY, S. K. 2010. The Rad9A checkpoint protein is required for nuclear localization of the claspin adaptor protein. *Cell Cycle,* 9**,** 548-56.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

SINGLETON, M. R., DILLINGHAM, M. S., GAUDIER, M., KOWALCZYKOWSKI, S. C. & WIGLEY, D. B. 2004. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature,* 432**,** 187-93.

SMITH, J., THO, L. M., XU, N. & GILLESPIE, D. A. 2010. The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res,* 108**,** 73-112.

SMITH, K. D., FU, M. A. & BROWN, E. J. 2009. Tim-Tipin dysfunction creates an indispensible reliance on the ATR-Chk1 pathway for continued DNA synthesis. *J Cell Biol,* 187**,** 15-23.

SMITH, V. F. & MATTHEWS, C. R. 2001. Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from E. coli: fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci,* 10**,** 116-28.

SMITH-ROE, S. L., PATEL, S. S., ZHOU, Y., SIMPSON, D. A., RAO, S., IBRAHIM, J. G., CORDEIRO-STONE, M. & KAUFMANN, W. K. 2013. Separation of intra-S checkpoint protein contributions to DNA replication fork protection and genomic stability in normal human fibroblasts. *Cell Cycle,* 12**,** 332-45.

SMITS, V. A., REAPER, P. M. & JACKSON, S. P. 2006. Rapid PIKK-dependent release of Chk1 from chromatin promotes the DNA-damage checkpoint response. *Curr Biol,* 16**,** 150-9.

SONNICHSEN, F. D., VAN EYK, J. E., HODGES, R. S. & SYKES, B. D. 1992. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry,* 31**,** 8790-8.

SREERAMA, N., VENYAMINOV, S. Y. & WOODY, R. W. 2000. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem,* 287**,** 243-51.

SREERAMA, N. & WOODY, R. W. 1993. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem,* 209**,** 32-44.

SREERAMA, N. & WOODY, R. W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem,* 287**,** 252-60.

SREERAMA, N. & WOODY, R. W. 2004. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol,* 383**,** 318-51.

STARZYK, A., BARBER-ARMSTRONG, W., SRIDHARAN, M. & DECATUR, S. M. 2005. Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: an isotope-edited FTIR study. *Biochemistry,* 44**,** 369-76.

STEMMER, W. P. 1994a. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A,* 91**,** 10747-51.

STEMMER, W. P. 1994b. Rapid evolution of a protein in vitro by DNA shuffling. *Nature,* 370**,** 389-91.

STUDACH, L., WANG, W. H., WEBER, G., TANG, J., HULLINGER, R. L., MALBRUE, R., LIU, X. & ANDRISANI, O. 2010. Polo-like kinase 1 activated by the hepatitis B virus X protein attenuates both the DNA damage checkpoint and DNA repair resulting in partial polyploidy. *J Biol Chem,* 285**,** 30282-93.

STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif,* 41**,** 207-234.

SUKACKAITE, R., JENSEN, M. R., MAS, P. J., BLACKLEDGE, M., BUONOMO, S. B. & HART, D. J. 2014. Structural and biophysical characterization of murine rif1 C terminus reveals high specificity for DNA cruciform structures. *J Biol Chem,* 289**,** 13903-11.

SVERGUN, D. I. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J,* 76**,** 2879-86.

SZYJKA, S. J., VIGGIANI, C. J. & APARICIO, O. M. 2005. Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. *Mol Cell,* 19**,** 691-7.

TANAKA, H., KATOU, Y., YAGURA, M., SAITOH, K., ITOH, T., ARAKI, H., BANDO, M. & SHIRAHIGE, K. 2009. Ctf4 coordinates the progression of helicase and DNA polymerase alpha. *Genes Cells,* 14**,** 807-20.

TANAKA, K. & RUSSELL, P. 2001. Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. *Nat Cell Biol,* 3**,** 966-72.

TANAKA, T., YOKOYAMA, M., MATSUMOTO, S., FUKATSU, R., YOU, Z. & MASAI, H. 2010. Fission yeast Swi1-Swi3 complex facilitates DNA binding of Mrc1. *J Biol Chem,* 285**,** 39609-22.

TAPIA-ALVEAL, C., CALONGE, T. M. & O'CONNELL, M. J. 2009. Regulation of chk1. *Cell Div,* 4**,** 8.

TARENDEAU, F., BOUDET, J., GUILLIGAY, D., MAS, P. J., BOUGAULT, C. M., BOULO, S., BAUDIN, F., RUIGROK, R. W., DAIGLE, N., ELLENBERG, J., CUSACK, S., SIMORRE, J. P. & HART, D. J. 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol,* 14**,** 229-33.

TARENDEAU, F., CREPIN, T., GUILLIGAY, D., RUIGROK, R. W., CUSACK, S. & HART, D. J. 2008. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog,* 4**,** e1000136.

TEWARI, R., BAILES, E., BUNTING, K. A. & COATES, J. C. 2010. Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol,* 20**,** 470-81.

TIAN-YU, J., LICHT, S., PARDEE, G., BHAT, A., CAO, Y., GAO, W., SANGALANG, E. & ZAROR, I. 2010. Binding Rate Screen - a high-throughput assay in soluble lysate for prioritizing protein expression constructs. *Anal Biochem,* 399**,** 276-83.

TOMPA, P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci,* 37**,** 509-16.

TOURRIERE, H., VERSINI, G., CORDON-PRECIADO, V., ALABERT, C. & PASERO, P. 2005. Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol Cell,* 19**,** 699-706.

TSAI, F. L., VIJAYRAGHAVAN, S., PRINZ, J., MACALPINE, H. K., MACALPINE, D. M. & SCHWACHA, A. 2015. Mcm2-7 is an Active Player in the DNA Replication Checkpoint Signaling Cascade via Proposed Modulation of its DNA Gate. *Mol Cell Biol*.

TSIMARATOU, K., KLETSAS, D., KASTRINAKIS, N. G., TSANTOULIS, P. K., EVANGELOU, K., SIDERIDOU, M., LIONTOS, M., POULIAS, I., VENERE, M., SALMAS, M., KITTAS, C., HALAZONETIS, T. D. & GORGOULIS, V. G. 2007. Evaluation of claspin as a proliferation marker in human cancer and normal tissues. *J Pathol,* 211**,** 331-9.

UNO, S. & MASAI, H. 2011. Efficient expression and purification of human replication fork-stabilizing factor, Claspin, from mammalian cells: DNA-binding activity and novel protein interactions. *Genes Cells,* 16**,** 842-56.

UNSAL-KACMAZ, K., CHASTAIN, P. D., QU, P. P., MINOO, P., CORDEIRO-STONE, M., SANCAR, A. & KAUFMANN, W. K. 2007. The human Tim/Tipin complex coordinates an Intra-S checkpoint response to UV that slows replication fork displacement. *Mol Cell Biol,* 27**,** 3131-42.

UNSAL-KACMAZ, K., MULLEN, T. E., KAUFMANN, W. K. & SANCAR, A. 2005. Coupling of human circadian and cell cycles by the timeless protein. *Mol Cell Biol,* 25**,** 3109-16.

URTISHAK, K. A., SMITH, K. D., CHANOUX, R. A., GREENBERG, R. A., JOHNSON, F. B. & BROWN, E. J. 2009. Timeless Maintains Genomic Stability and Suppresses Sister Chromatid Exchange during Unperturbed DNA Replication. *J Biol Chem,* 284**,** 8777-85.

UVERSKY, V. N. 2013. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci,* 22**,** 693-724.

UZUNOVA, S. D., ZARKOV, A. S., IVANOVA, A. M., STOYNOV, S. S. & NEDELCHEVA-VELEVA, M. N. 2014. The subunits of the S-phase checkpoint complex Mrc1/Tof1/Csm3: dynamics and interdependence. *Cell Div,* 9**,** 4.

VANDERPOOL, D., JOHNSON, T. O., PING, C., BERGQVIST, S., ALTON, G., PHONEPHALY, S., RUI, E., LUO, C., DENG, Y. L., GRANT, S., QUENZER, T., MARGOSIAK, S., REGISTER, J., BROWN, E. & ERMOLIEFF, J. 2009. Characterization of the CHK1 allosteric inhibitor binding site. *Biochemistry,* 48**,** 9823-30.

VAUGHN, J. L., GOODWIN, R. H., TOMPKINS, G. J. & MCCAWLEY, P. 1977. The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro,* 13**,** 213-7.

VERLINDEN, L., VANDEN BEMPT, I., EELEN, G., DRIJKONINGEN, M., VERLINDEN, I., MARCHAL, K., DE WOLF-PEETERS, C., CHRISTIAENS, M. R., MICHIELS, L., BOUILLON, R. & VERSTUYF, A. 2007. The E2F-regulated gene Chk1 is highly expressed in triple-negative estrogen receptor /progesterone receptor /HER-2 breast carcinomas. *Cancer Res,* 67**,** 6574-81.

VINCZE, T., POSFAI, J. & ROBERTS, R. J. 2003. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res,* 31**,** 3688-91.

VOINEAGU, I., NARAYANAN, V., LOBACHEV, K. S. & MIRKIN, S. M. 2008. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A,* 105**,** 9936-41.

VOINEAGU, I., SURKA, C. F., SHISHKIN, A. A., KRASILNIKOVA, M. M. & MIRKIN, S. M. 2009. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat Struct Mol Biol,* 16**,** 226-8.

WALDO, G. S. 2003. Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol Biol,* 230**,** 343-59.

WALDO, G. S., STANDISH, B. M., BERENDZEN, J. & TERWILLIGER, T. C. 1999. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol,* 17**,** 691-5.

WALKER, M., BLACK, E. J., OEHLER, V., GILLESPIE, D. A. & SCOTT, M. T. 2009. Chk1 C-terminal regulatory phosphorylation mediates checkpoint activation by de-repression of Chk1 catalytic activity. *Oncogene,* 28**,** 2314-23.

WALWORTH, N., DAVEY, S. & BEACH, D. 1993. Fission yeast chk1 protein kinase links the rad checkpoint pathway to cdc2. *Nature,* 363**,** 368-71.

WANG, X., ZOU, L., LU, T., BAO, S., HUROV, K. E., HITTELMAN, W. N., ELLEDGE, S. J. & LI, L. 2006. Rad17 phosphorylation is required for claspin recruitment and Chk1 activation in response to replication stress. *Mol Cell,* 23**,** 331-41.

WARTCHOW, C. A., PODLASKI, F., LI, S., ROWAN, K., ZHANG, X., MARK, D. & HUANG, K. S. 2011. Biosensor-based small molecule fragment screening with biolayer interferometry. *J Comput Aided Mol Des,* 25**,** 669-76.

WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res,* 32**,** W668-73.

WHITMORE, L. & WALLACE, B. A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers,* 89**,** 392-400.

WILLIAMSON, A. R. 2000. Creating a structural genomics consortium. *Nat Struct Biol,* 7 Suppl**,** 953.

WILLIS, J., PATEL, Y., LENTZ, B. L. & YAN, S. 2013. APE2 is required for ATR-Chk1 checkpoint activation in response to oxidative stress. *Proc Natl Acad Sci U S A,* 110**,** 10592-7.

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-42.

WINTER, G. 2009. xia2: an expert system for macromolecular crystallography data reduction *J Appl Cryst,* 43**,** 186-190.

XIANG, Z. 2006. Advances in homology protein structure modeling. *Curr Protein Pept Sci,* 7**,** 217-27.

XU, J., JIAO, F. & YU, L. 2008. Protein structure prediction using threading. *Methods Mol Biol,* 413**,** 91-121.

XU, Y. J., DAVENPORT, M. & KELLY, T. J. 2006. Two-stage mechanism for activation of the DNA replication checkpoint kinase Cds1 in fission yeast. *Genes Dev,* 20**,** 990-1003.

XU, Y. J. & KELLY, T. J. 2009. Autoinhibition and autoactivation of the DNA replication checkpoint kinase Cds1. *J Biol Chem,* 284**,** 16016-27.

YILMAZ, S., SANCAR, A. & KEMP, M. G. 2011. Multiple ATR-Chk1 pathway proteins preferentially associate with checkpoint-inducing DNA substrates. *PLoS One,* 6**,** e22986.

YOO, H. Y., JEONG, S. Y. & DUNPHY, W. G. 2006. Site-specific phosphorylation of a checkpoint mediator protein controls its responses to different DNA structures. *Genes Dev,* 20**,** 772-83.

YOO, H. Y., KUMAGAI, A., SHEVCHENKO, A., SHEVCHENKO, A. & DUNPHY, W. G. 2004. Adaptation of a DNA replication checkpoint response depends upon inactivation of Claspin by the Polo-like kinase. *Cell,* 117**,** 575-88.

YOSHIMURA, A., AKITA, M., HOSONO, Y., ABE, T., KOBAYASHI, M., YAMAMOTO, K., TADA, S., SEKI, M. & ENOMOTO, T. 2011. Functional relationship between Claspin and Rad17. *Biochem Biophys Res Commun,* 414**,** 298-303.

YOSHIYAMA, K. O., SAKAGUCHI, K. & KIMURA, S. 2013. DNA damage response in plants: conserved and variable response compared to animals. *Biology (Basel),* 2**,** 1338-56.

YOSHIZAWA-SUGATA, N. & MASAI, H. 2007. Human Tim/Timeless-interacting protein, Tipin, is required for efficient progression of S phase and DNA replication checkpoint. *J Biol Chem,* 282**,** 2729-40.

YOSHIZAWA-SUGATA, N. & MASAI, H. 2009. Roles of human AND-1 in chromosome transactions in S phase. *J Biol Chem,* 284**,** 20718-28.

YUAN, J., LUO, K., DENG, M., LI, Y., YIN, P., GAO, B., FANG, Y., WU, P., LIU, T. & LOU, Z. 2014. HERC2-USP20 axis regulates DNA damage checkpoint through Claspin. *Nucleic Acids Res,* 42**,** 13110-21.

YUE, M., SINGH, A., WANG, Z. & XU, Y. J. 2011. The phosphorylation network for efficient activation of the DNA replication checkpoint in fission yeast. *J Biol Chem,* 286**,** 22864-74.

YUMEREFENDI, H., DESRAVINES, D. C. & HART, D. J. 2011. Library-based methods for identification of soluble expression constructs. *Methods,* 55**,** 38-43.

YUMEREFENDI, H., TARENDEAU, F., MAS, P. J. & HART, D. J. 2010. ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J Struct Biol,* 172**,** 66-74.

ZEE, B. M. & GARCIA, B. A. 2013. Validation of protein acetylation by mass spectrometry. *Methods Mol Biol,* 981**,** 1-11.

ZHANG, D., ZAUGG, K., MAK, T. W. & ELLEDGE, S. J. 2006. A role for the deubiquitinating enzyme USP28 in control of the DNA-damage response. *Cell,* 126**,** 529-42.

ZHANG, J., SONG, Y. H., BRANNIGAN, B. W., WAHRER, D. C., SCHIRIPO, T. A., HARRIS, P. L., HASERLAT, S. M., ULKUS, L. E., SHANNON, K. M., GARBER, J. E., FREEDMAN, M. L., HENDERSON, B. E., ZOU, L., SGROI, D. C., HABER, D. A. & BELL, D. W. 2009a. Prevalence and functional analysis of sequence variants in the ATR checkpoint mediator Claspin. *Mol Cancer Res,* 7**,** 1510-6.

ZHANG, J., YANG, P. L. & GRAY, N. S. 2009b. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer,* 9**,** 28-39.

ZHAO, B., BOWER, M. J., MCDEVITT, P. J., ZHAO, H., DAVIS, S. T., JOHANSON, K. O., GREEN, S. M., CONCHA, N. O. & ZHOU, B. B. 2002. Structural basis for Chk1 inhibition by UCN-01. *J Biol Chem,* 277**,** 46609-15.

ZHAO, H. & PIWNICA-WORMS, H. 2001. ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol Cell Biol,* 21**,** 4129-39.

ZHAO, H. & RUSSELL, P. 2004. DNA binding domain in the replication checkpoint protein Mrc1 of Schizosaccharomyces pombe. *J Biol Chem,* 279**,** 53023-7.

ZHAO, H., TANAKA, K., NOGOCHI, E., NOGOCHI, C. & RUSSELL, P. 2003. Replication Checkpoint Protein Mrc1 Is Regulated by Rad3 and Tel1 in Fission Yeast. *Mol Cell Biol,* 23**,** 8395-8403.

ZHU, M., ZHAO, H., LIAO, J. & XU, X. 2014. HERC2/USP20 coordinates CHK1 activation by modulating CLASPIN stability. *Nucleic Acids Res,* 42**,** 13074-81.

ZHU, W., UKOMADU, C., JHA, S., SENGA, T., DHAR, S. K., WOHLSCHLEGEL, J. A., NUTT, L. K., KORNBLUTH, S. & DUTTA, A. 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes Dev,* 21**,** 2288-99.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res,* 31**,** 3406-15.

ZUO, T., LIU, D., LV, W., WANG, X., WANG, J., LV, M., HUANG, W., WU, J., ZHANG, H., JIN, H., ZHANG, L., KONG, W. & YU, X. 2012. Small-molecule inhibition of human immunodeficiency virus type 1 replication by targeting the interaction between Vif and ElonginC. *J Virol,* 86**,** 5497-507.