Structural studies on members of the CNF1-like family of glutamine deamidase toxins

Aspects of the mechanism, specificity and folding

Amirul Adli bin Abd Aziz

Department of Molecular Biology and Biotechnology

University of Sheffield

This thesis is submitted for the degree of

Doctor of Philosophy

September 2018

I would like to dedicate this thesis to my beloved wife, and our unborn child. I am so delighted and cannot wait to see you. I love you both . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and the relevant sections in author contributions.

Amirul Adli bin Abd Aziz

September 2018

Acknowledgements

I would like to offer the biggest token of appreciation to Prof. David Rice for his unrelenting support, advice and all the guidance that I could not have needed more in completing this piece of work. I am deeply honoured to have been given this opportunity to work with such an inspiring and enthusiastic supervisor. Thank you to my PhD advisors, Prof. Per Bullough and Dr. John Rafferty, and also to Dr. Patrick Baker for all your help and directions throughout this study. Thanks also to Dr. George Mobbs, Dr. Claudine Bisson, Dr. Jason Wilson and Dr. Hayley Owen for helping me settle down and find my feet at the start of this journey.

I have been so grateful to be married to such an awesome, smart and beautiful wife. I want to thank you from the deepest core of my heart for always being there for me with your love and affection that transcend beyond space and time. Your presence, despite the physical distance that we have had to endure, is essentially what keeps me going. So much love and thanks to my family who have been really supportive all these years. So much quality times had been missed and I am looking forward to mending it.

A big shout out to my comrades in arms, Sam Dix and Alicia Churchill-Angus, for all the late night snacks and chat during long-hours of data collection all for the sake of science, and for making this journey a lot more fun and much less difficult than it could well have been. Also thanks to Kak Azura Mohd. Noor for your support and prayers. Special thanks to Fiona Rodgers for maintaining the laboratory and my sanity by preparing fresh coffee every day and Dr. Svetlana Sedelnikova for your wisdom in protein purification. Special

mention also goes to colleagues in D08, Dr. Sami Melebari, Dr. Naer Al-Kaabawi for the meaningless Arabic conversations in between stressful experiments and Dr. Roxanne Lau for the boardgames adventure.

Needless to say, I am deeply honoured to be surrounded by a friendly and supportive bunch of people, including my capoeira 'familia de Ouro' for all the music, dance, and friendships, friends and relatives for all your prayers and motivations and all the people I have encountered in the UK.

Last but definitely not least, I would like to thank Majlis Amanah Rakyat (MARA) Malaysia for providing me with financial support for this study, without which, none of these would have been remotely possible. Thanks very much, everybody.

Abstract

This thesis describes structural studies on the CNF1-like family of glutamine deamidase toxins. The first manuscript provides a detailed account of the structure determination and analysis of two crystal forms of the BLF1 C94S:eIF4A complex, providing the first insights into the specificity and mechanism of any CNF1-like family of toxins. Analysis of the interactions formed between BLF1 and eIF4A and comparison with other deamidases and proteases suggest that the deamidation of Gln339 in eIF4A by BLF1 follows a similar mechanism to that of the cysteine protease, papain.

The second manuscript describes the identification of another member of the CNF1-like family of toxins, homologous to BLF1, SSR1 from *Skermanella stibiiresistens*. The manuscript provides details on the crystallisation and data collection on a monomeric and a dimeric form of SSR1 and demonstrates that SSR1 also deamidates Gln339 in eIF4A.

The third manuscript describes the structure determination of an SSR1 monomer which is clipped between residues Ser191 and Ser192, and an unclipped domain-swapped dimer. Analysis of these structures and a number of mutants suggests that there is a strain during the folding of SSR1 that determines whether domain-swapping or clipping occurs. The origin of the strain in SSR1 lies in a clash between residues surrounding the clip site, which stalls the folding of the monomer increasing chances of domain-swapping. The resolution of this strain by clipping requires the presence of either a Ser, Cys, or Thr at position 192 in a mechanism mimicking that of self-splicing inteins. The outcomes of this investigation provide fresh

insights into strain and its resolution during folding which, in general, raise many interesting questions on the evolutionary aspects of protein folding.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

3D	Three dimension
ADP	Adenosine diphosphate
ADT	Asparagine Deamidase Toxins
BLF	Burkholderia Lethal Factor
CFE	Cell-free Extract
Cif	Cycle-inhibiting factor
CNF	Cytotoxic Necrotizing Factor
CV	Column Volume
DEAE	Diethylethanolamine
DLS	Diamond Light Source
DNA	Deoxyribonucleic acids
DT	Deamidase toxins
DTT	Dithiothreitol

EDTA	Ethylenediaminetetraacetic acid
eIF4A	eukaryotic initiation factor 4A
GAP	GTPase activating protein
GDT	Glutamine Deamidase Toxins
GEF	Guanine-nucleotide exchange factor
gfl	gel-filtration load
GTP	Guanosine triphosphate
GUI	Graphical user interface
HAB	Harmful Algal blooms
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HSV	Herpes Simplex Virus
IEC	Ion Exchange Chromatography
IMAC	Immobilised Metal-ion Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny Broth
LDS	Lithium dodecyl-sulphate
LPS	Lipopolysaccharides
LSGC	Leucine-serine-glycine-cysteine
MCP	Methyl-Accepting Chemotaxis Proteins

MCS	Multiple Cloning Site
MES	2-(N-morpholino)ethanesulfonic acid
MPD	2-Methyl-2,4-pentanediol
MR	Molecular Replacement
MWCO	Molecular weight cut-off
NAC	Near Attack Complex
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NMR	Nuclear Magnetic Resonance
NTA	nitrilotriacetic acid
OD	Optical Density
OMV	Outer membrane vesicles
PAGE	Poly-acrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Poly-ethylene glycol
PISA	Proteins, Interfaces, Structures and Assemblies
Pmt	Pasteurella multocida toxin
rmsd	root-mean-square deviation

RNA	Ribonucleic acids
SAD	Single Anomalous Dispersion
SDS	Sodium dodecyl sulphate
SEC	Size Exclusion Chromatography
SIB	Swiss Institute of Bioinformatics
T3SS	Type III secretion system
TAE	Tris-Acetate-EDTA
TEMED	Tetramethylethylenediamine
TNF	Tumour necrosis factor
TOF	time-of-flight
TRIS	2-Amino-2-(hydroxymethyl)propane-1,3-diol
WT	wild-type

Chapter 1

Introduction

1.1 Bacterial pathogens and diseases

Bacteria have evolved to co-exist with a variety of more complex multicellular host organisms especially mammals, plants and insects. Some inhabit the host in a symbiotic fashion whilst others, bacterial pathogens, hijack aspects of the host system for their own gain resulting in damage and ultimately disease. With advances in genetics, microbiology and molecular biology, scientists have begun to elucidate the molecular basis of bacterial pathogenesis through the studies on potential virulence factors identified in the genome of pathogenic bacteria. Some virulence factors are encoded in pathogenicity islands in the genome of the bacteria and are normally expressed selectively either when they are under stress from the immune response of the host or through other signals that prompt the opportunity to begin the phase of invasion and spread. Whereas, some are simply structural components of the bacteria that are able to trigger a host response. Virulence factors, otherwise referred to as toxins, can come in many shapes and forms, including, and not limited to, cell wall components that trigger immune response in host via host cell surface receptor activation, and a variety of secreted protein products that possess a range of functions either by aiming

for specific modification of intracellular host proteins or through non-specific activation of host immune cells.

1.1.1 Endotoxins and Exotoxins

Bacterial toxins can be classified into two groups; Endotoxins and Exotoxins. Endotoxins are large molecules that are found at the outer membrane of gram-negative bacteria and are composed of lipids and polysaccharides, and are thus generally referred to as Lipopolysaccharides (LPS). LPS consists of three main component; the O-antigen, the core oligosaccharides, and Lipid A. The third component, Lipid A, is the crucial constituent that is responsible for the toxic activity of LPS, interacting with major host immune cells including macrophages and monocytes, inducing the release of bioactive peptides such as tumour necrosis factor alpha (TNFα) and cytokines which, if not controlled, could lead to severe toxic shock (Rietschel et al., 1994) (Figure 1.1). Endotoxins are therefore bacterial components that are only transferred to the host via a responsive mechanism either through the destruction of the bacterial cell wall or through bacterial outer membrane vesicles (OMV) (Kulp and Kuehn, 2010). Exotoxins however, are products of the invading bacterial pathogens that are secreted into the surrounding (Figure 1.1), which are targeted to modify crucial components of host cells or tissues by means of different modes of actions.

Pathogenic bacteria have evolved sophisticated multi-complex macromolecules to ensure transport of these toxins to the correct destination, which involves crossing the bacterial cell membranes into the surrounding environment, from which internalisation of virulence factors into targeted host cell cytoplasm would occur. In Gram-negative bacteria, these complex nanomachines exist as membrane associated proteins that can span either one of the bacterial membranes or form double-membrane-embedded secretion systems (Green and Mecsas, 2016). SecYEG translocon and Tat system are examples of the inner-membrane associated transporters where substrates, in unfolded or folded form, respectively, are exported through

the inner membrane into the periplasm and through a dedicated outer membrane transporters into the extracellular matrix (Nivaskumar and Francetic, 2014). Consistent with their involvement in transferring effectors that can be associated with virulence, gene knockouts of the Tat system in *E. coli* O157 not only block the secretion of the Shiga toxin but also result in reduced toxicity and poor cell motility (Pradel et al., 2003). Double-membrane-spanning secretion systems have been classified into Type 1 Secretion System (T1SS), T2SS, T3SS, T4SS and T6SS (Costa et al., 2015). Although the complete mechanism of these systems is yet to be fully understood, advances in structural biology techniques have contributed major insights on how each secretion system functions. Prominent among these, and relevant to transport of virulence factors, the T2SS and T3SS will be briefly discussed.

The T2SS comprises four components including a cytoplasmic ATPase, an inner membrane platform, a periplasmic pilus-like structure, and an outer membrane complex (Korotkov et al., 2012). Effectors are first transported into the periplasm by the SecYEG translocon or the Tat transporter (Nivaskumar and Francetic, 2014). These substrates are then believed to be pushed by the periplasmic pseudopilus assembly in an ATP-dependent manner through the outer membrane complex into the extracellular milieu (Nivaskumar et al., 2014). One well-characterised T2SS is the Eps system in *Vibrio cholera* that exports the cholera toxin (Nivaskumar and Francetic, 2014). Complementation studies on mutant *V. cholera* strain defective in the secretion of cholera toxin identified genes belonging to the T2SS gene clusters, implicating their pivotal role in export of bacterial effectors (Overbye et al., 1993; Sandkvist et al., 1993).

The T3SS, or generally termed injectisome due to the syringe-like architecture, is a multiprotein megastructure embedded in the entire cell envelope that consists of two sets of inner and outer rings forming the base and a needle-like filament that protrudes into the extracellular matrix and acts as a tube that channels the bacterial outer membrane to the cytoplasm of the host cell (Kubori et al., 2000; Schraidt and Marlovits, 2011). Conformational

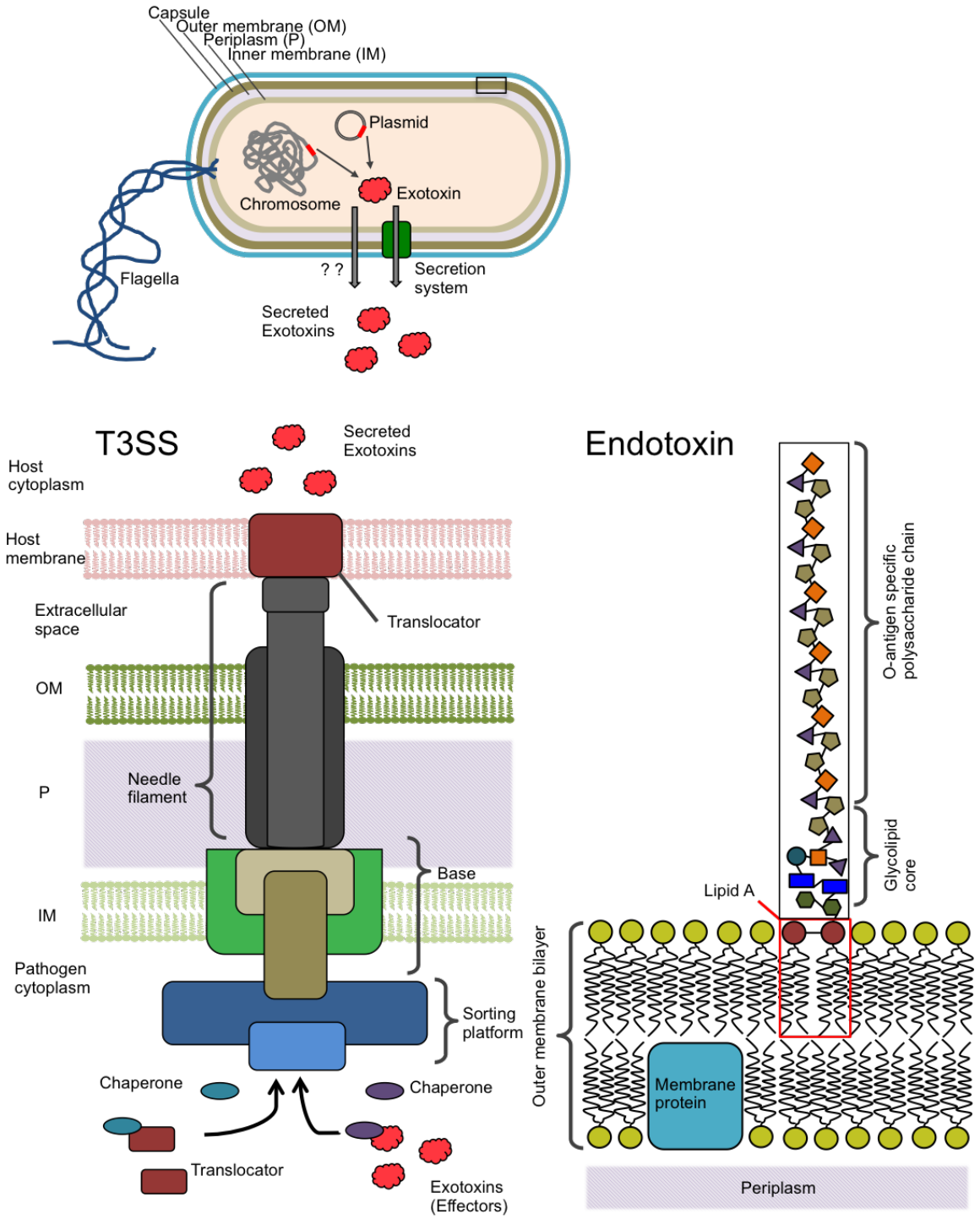


Fig. 1.1 **A schematic of a bacterial cell showing different virulence factors and the T3SS.** Endotoxins (LPS) lie in the boxed region (top middle) which is amplified (bottom right) showing the components of an LPS molecule. Lipid A in red box is the toxic component of LPS. In contrast, exotoxins are produced and secreted, either by a known secretion system, an example being the T3SS (bottom left) (adapted from (Costa et al., 2015)) or an unknown mechanism into the surrounding to be transported to host cells.

change of the filament upon host cell contact is believed to trigger a signalling mechanism to prompt secretions of translocators that form a pore in the target host membrane and designated virulence factors in the form of unfolded polypeptides (Blocker et al., 2008). Precise control of the secretion of the substrates is governed by their N-terminal secretion signals, which are recognised by dedicated chaperones to coordinate translocation through the base and the needle filament via a sorting platform in an ATP-dependent manner (Figure 1.1) (Büttner, 2012; Lara-Tejero et al., 2011). The T3SS is utilised by a variety of pathogenic bacteria including the *Salmonella*, *Yersinia*, *Pseudomonas*, and the enteropathogenic *Escherichia* genera, whereby bacterial effectors are delivered directly into the cytoplasm of the respected host cells, where they disrupt key host cellular functions to promote bacterial invasion, survival and colonisation (Cornelis, 2006; Galán and Wolf-Watz, 2006). Owing to the tightly-controlled nature of the T3SS secretion stages, mutation or deletion of its components, often leads to defective secretion phenotype, for instance, knockout of *Shigella* InvJ, a protein that controls the needle length, results in abnormally long faulty needles (Deane et al., 2010).

Exotoxins can be classified into three major modes of actions; 1) Intracellular-targeting toxins 2) Membrane-disrupting toxins and 3) Superantigens. Intracellular-targeting toxins are generally known as the A-B toxins, whereby the B component is important to mediate initial binding to specific host cell receptors, upon which internalisation of the toxins occur normally via endocytosis, followed by the release of the active A component into the host cytoplasm where it interferes with the specific cellular function by modifying its target (Williams and Tsai, 2016). Membrane-disrupting toxins work by either forming a cytolytic pore on host cell membrane or by disrupting the phospholid bilayer structure with phospholipases toxins (Istivan and Coloe, 2006; Peraro and van der Goot, 2016). Finally, superantigens evoke non-specific immune response of T-cells triggering a massive release of cytokines which ultimately lead to toxic shock syndrome (Fraser, 2011) (Figure 1.2).

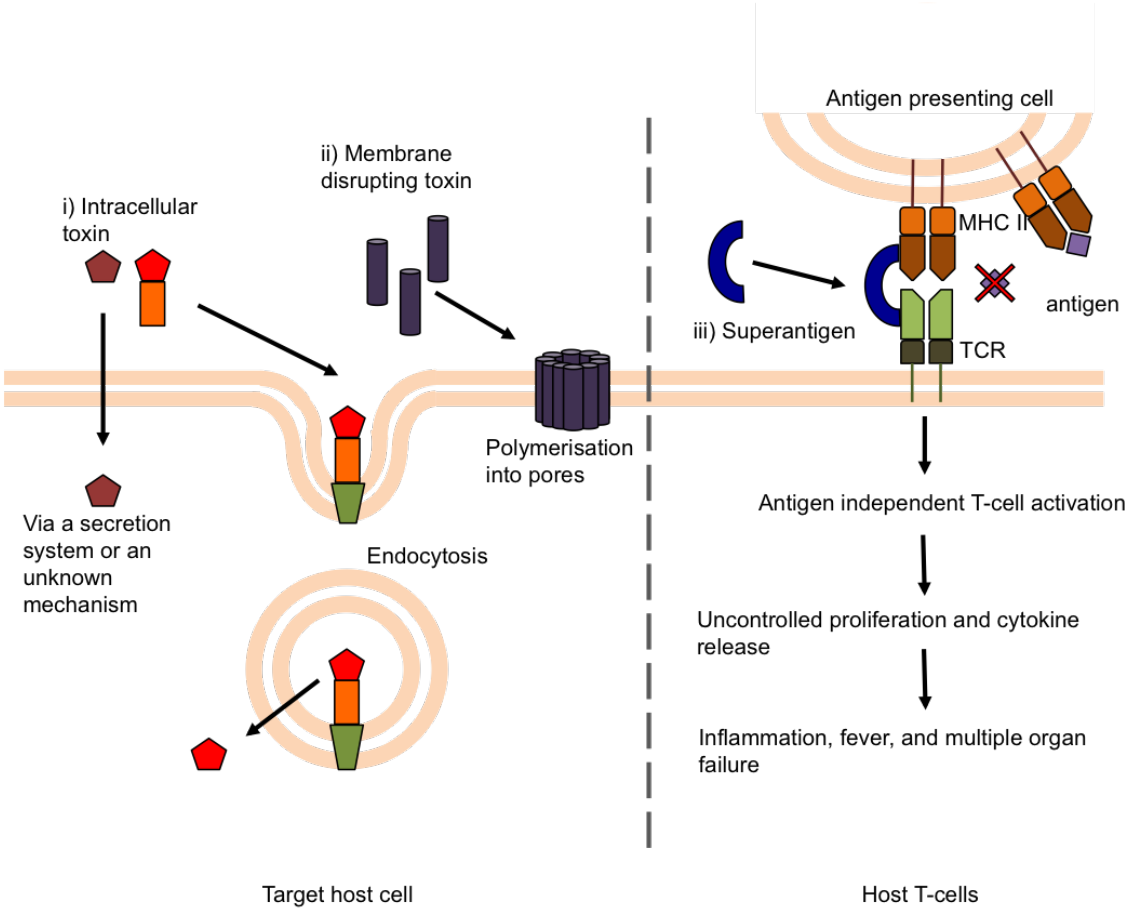


Fig. 1.2 **Different types of exotoxins and their mode of actions.** A schematic showing three major types of exotoxins. Intracellular toxins are transported, via endocytosis, a secretion system or an unknown mechanism, into the host cytosol where they target modification of host proteins (Williams and Tsai, 2016). Membrane disrupting toxins form a pore-like structures on the host cell membranes causing loss of osmotic balance (Istivan and Coloe, 2006). Superantigens bypass the normal presentation of antigens leading to severe allergic reactions (Fraser, 2011).

1.1.2 Toxins with various enzymatic properties

The intracellular toxins secreted by bacterial pathogens can have a range of different enzymatic functions including kinases, ADP-ribosylases, phosphatases, hydrolases, lipases, and proteases. For example, Doc is a kinase toxin, from bacteriophage P1, which phosphorylates an essential threonine residue, Thr382, of the elongation factor EF-Tu, hence blocking translation elongation in *E. coli* (Cruz et al., 2014). In addition, Botulinum neurotoxins (BoNTs), one of the most lethal toxins known, from *Clostridium botulinum*, are zinc-dependant endoproteases that cleave SNARE proteins including SNAP-25, VAMP and Syntaxin, proteins that play major roles in presynaptic membrane of the neurones (Lebeda et al., 2010). Cleavage of these SNARE proteins prevents the release of neurotransmitter vesicles resulting in disruption in nerve signalling and ultimately causes paralysis, a symptom of the disease botulism (Montecucco and Molgó, 2005). One enzymatic activity that will be discussed as the main topic of this thesis is deamidation. Deamidases (enzymes that carry out a deamidation reaction) are enzymes that convert amide groups into carboxylic acids (Figure 1.3).

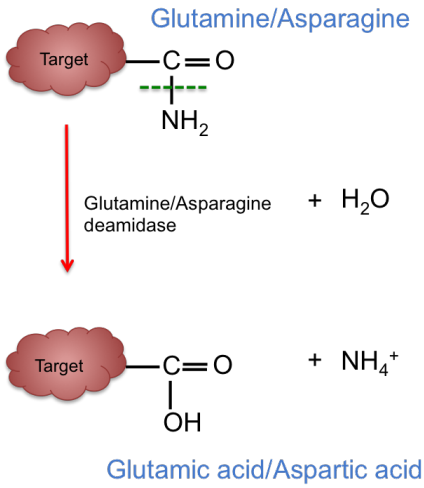


Fig. 1.3 **Deamidation reaction catalysed by deamidase toxins on a glutamine or asparagine residue.** Deamidases cleave the C-N bond on the carboxyamide group and converts it to a carboxyl group.

Whilst some bacteria utilise deamidase toxins as virulence factors, deamidation can also be a component of a normal bacterial cellular processes ensuring bacterial cell motility and survival. One example of the latter is CheD in *Bacillus subtilis*, which is a chemoreceptor modification enzyme important in chemotaxis (Kristich and Ordal, 2002). CheD was found to have a glutamine deamidase activity responsible for deamidating specific glutamine residues in Methyl-Accepting Chemotaxis Proteins (MCPs) (Chao et al., 2006). MCPs are bacterial receptor proteins that sense various chemical signals in the environment through ligand-receptor binding and relay the information towards the flagellar assembly of the bacteria to allow proper adaptation by flagellar motility (Falke and Hazelbauer, 2001).

Not surprisingly, bacterial pathogens have evolved toxins that employ deamidation activity to modify proteins in host cells by means of deamidation often on a side chain of residues critical for certain cellular functions mostly those that are involved in maintaining host cell survival and structural integrity. These bacterial effectors will be referred to as deamidase toxins.

1.2 Deamidase Toxins

Deamidase Toxins (DT) are toxins secreted by pathogens that modify host target proteins via deamidation of a critical amino acid which result in either loss or change in activity of the host protein and eventually lead to cell damage and apoptosis. A rising number of bacteria have evolved to have DTs as weapons in their arsenal to cause harm to host cells to maintain bacterial survival, invasion and spread during infection. The first DT discovered in bacterial pathogen was the Cytotoxic Necrotizing Factor-1 (CNF1) from *Escherechia coli*, which causes deamidation of critical glutamine residue in a number of small GTPases (Flatau et al., 1997). Following this, more deamidase toxins were discovered in a number of other bacterial pathogens which deamidate critical glutamine residues in a variety of target host proteins. This family of DTs will therefore be called Glutamine Deamidase Toxins (GDT).

Very recently, an asparagine deamidase toxin (ADT), TecA, was identified in an opportunistic pathogen *Burkholderia cenocepacia* (Aubert et al., 2016). In addition, a protein UL37 in Herpes Simplex Virus 1 (HSV-1) deamidates two critical asparagine residues in RIG-1, a host receptor that senses invading viral dsRNA, and thus blocks the RIG-1 dependent innate immune response towards viral infection (Zhao et al., 2016).

1.2.1 Cytotoxic Necrotizing Factor 1 (CNF1) from Uropathogenic *Escherichia coli* sets the scene

Cytotoxic Necrotizing Factor 1 (CNF1) is the first protein toxin characterised in uropathogenic *E. coli* (Caprioli et al., 1983). The gene for CNF1 encodes 1014 amino acids with the final protein product had been shown to be divided into 3 parts; 1) a cell-binding domain 2) a central region which include a number of membrane spanning helices and 3) an active catalytic toxin domain (Boquet, 2001) (Figure 1.4).

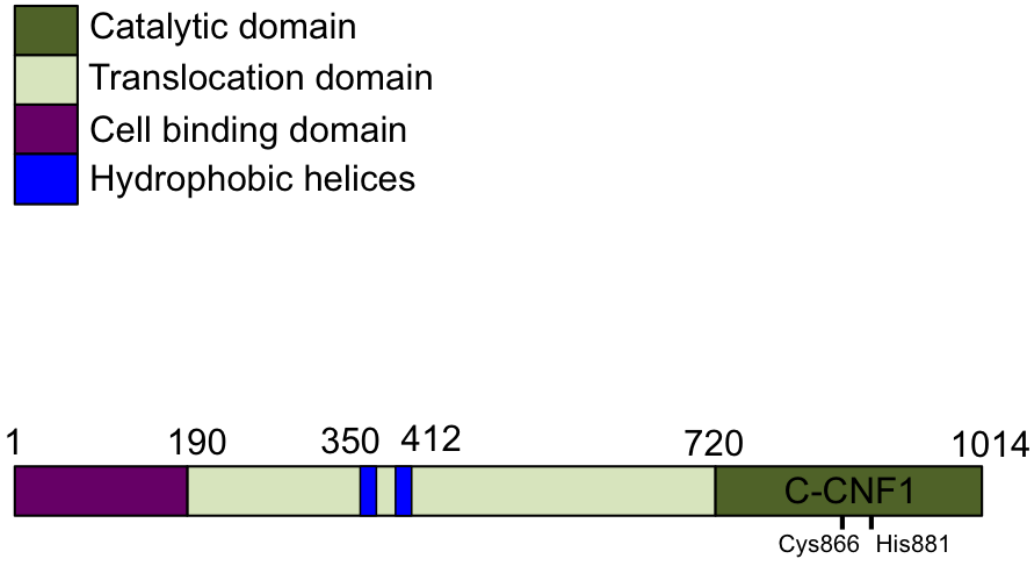


Fig. 1.4 **Domain organisation of CNF1.** CNF1 consists of three domains including a host cell-binding domain, a membrane translocation domain and a deamidase domain.

Due to the extracellular lifestyle of *E. coli*, a delivery system is thus required to make sure the catalytic toxin domain is transported across the membrane of target host cells. Once the toxin has been transported across the membrane, the deamidation domain gets access into the cytoplasm of the host cell where the target host protein resides. The catalytic domain of CNF1 (C-CNF1) was found to deamidate a critical glutamine residue, Gln63, in the switch II region of the small GTP binding protein RhoA (Flatau et al., 2000) and Gln61 in Rac and cdc42 (Lerm et al., 1999). The deamidation of these critical glutamine residues blocks the hydrolysis of GTP either from their intrinsic GTPase activity or from the induction of GTPase activating protein (GAP). This modification renders the GTPase in a constitutively active state, stimulating the activation of downstream effectors which results in the formation of multinucleated cells, actin cytoskeleton remodelling and stress fibre formation, which lead to membrane ruffling and filopedia and ultimately cell death (Figure 1.5) (Falzano et al., 1993).

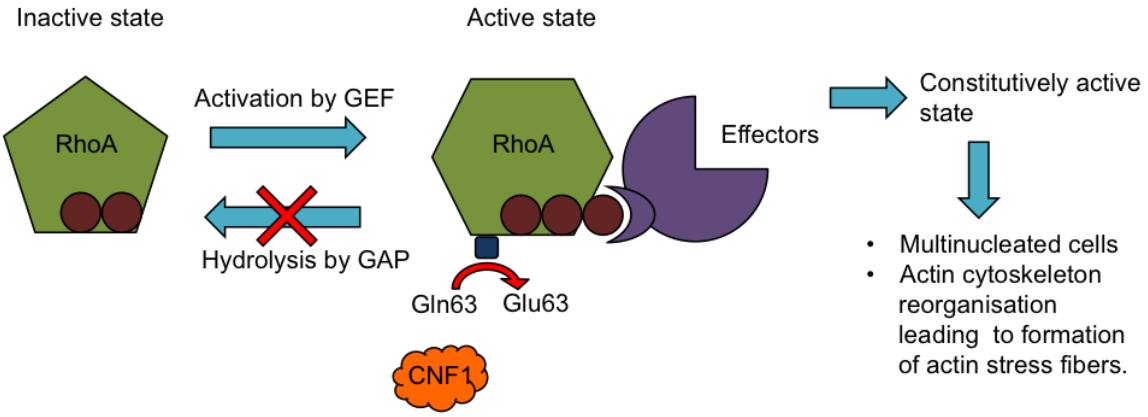


Fig. 1.5 **Mode of action of the catalytic domain of CNF1.** CNF1 constitutively activates RhoA GTPase leading to the disruption of normal cellular activity causing damage to host cells. Figure modified from (Boquet, 2001).

The discovery of this glutamine deamidase activity in a bacterial protein toxin was the first of its kind (Flatau et al., 1997; Schmidt et al., 1997). Although the exact mechanism of glutamine deamidase catalysis has yet to be determined, the reaction is thought to mimic the hydrolysis of a peptide bond by cysteine peptidases like papain as a result of cysteine-histidine (Cys-His) pairs in their respective active sites (Buetow et al., 2001) (Figure 1.6). The only difference between these reactions is the target of modification; glutamine deamidases target glutamine side chain amide while cysteine peptidases break the peptide bond in the protein backbone (Lowe, 1976).

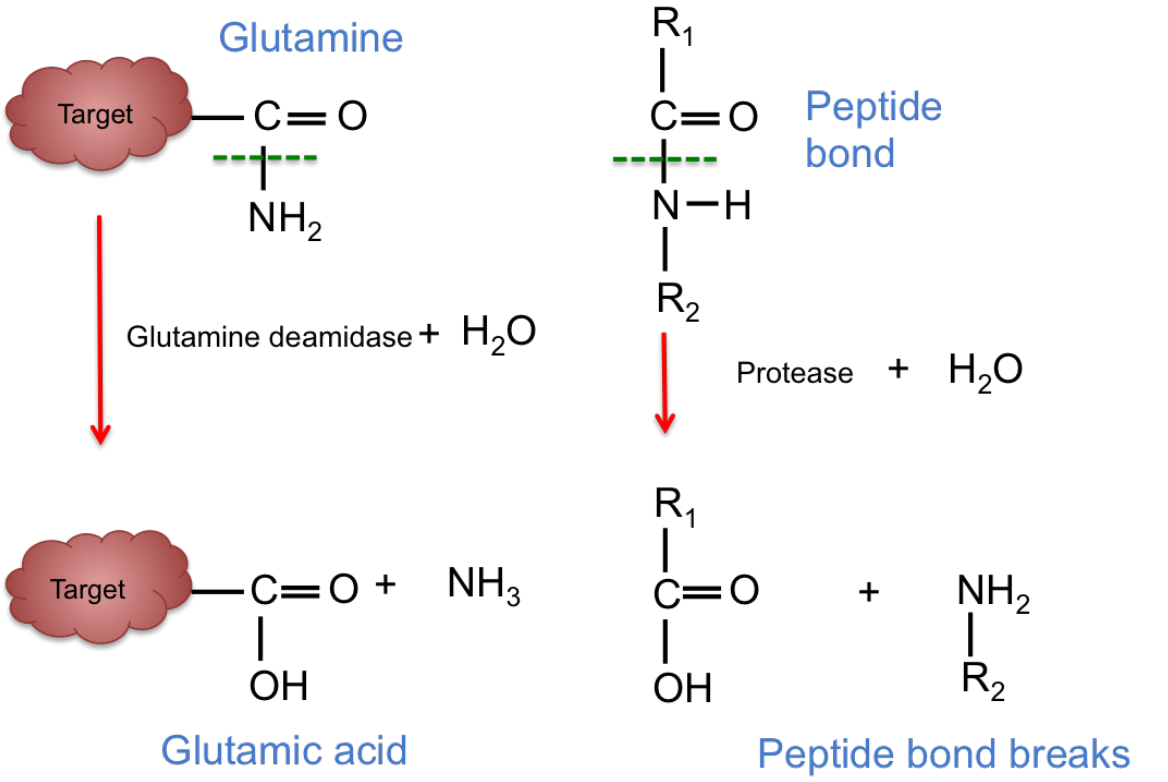


Fig. 1.6 **Similar reaction chemistry between deamidation and peptide bond cleavage.** Deamidase toxins cleave a similar C-N bond to that cleaved in proteases and given the functional similarities of residues around their respective active sites, this suggests that they share a related mechanism.

Following the discovery of CNF1, an increasing number of glutamine deamidases have been discovered, for some of which the structures have been solved either on their own or in complex with their substrates (Washington et al., 2013). The rising number of characterised glutamine deamidase toxins has highlighted deamidation as a potent type of chemistry adopted by effectors in the arsenal of bacterial pathogenesis. Structural analysis and sequence comparisons have grouped the glutamine deamidases into two families, each with a different fold but both utilising a Cys-His pair in the active site. The Cys-His pair is thought to be critical in deamidating glutamine target via nucleophilic attack by the thiolate group of the cysteine on the carboxyamide group of the glutamine side chain in target host proteins. The modification of the target then causes major disruption in downstream processes of critical cellular pathways, which could ultimately lead to cell death (Boquet, 2001; Falzano et al., 1993; Orth et al., 2007; Washington et al., 2013).

1.3 The papain-like family of GDTs

The papain-like family of deamidases consists of proteins that possess a papain-like fold comprising a single antiparallel β-sheet surrounded by α-helical moeities while retaining the spatial conservation of the Cys-His pair in the active site (Figure 1.7). The toxins in this family of deamidases include Cycle-inhibiting factors (Cifs) from a myriad of pathogenic bacteria including *Burkholderia pseudomallei*, *Yersinia pseudotuberculosis*, *Photorhabdus luminescens*, and enteropathogenic *E. coli*, Osp1 from *Shigella flexineri*, and the Pasteurella multoxida toxin (Pmt) from *Pasteurella multocida* (Washington et al., 2013).

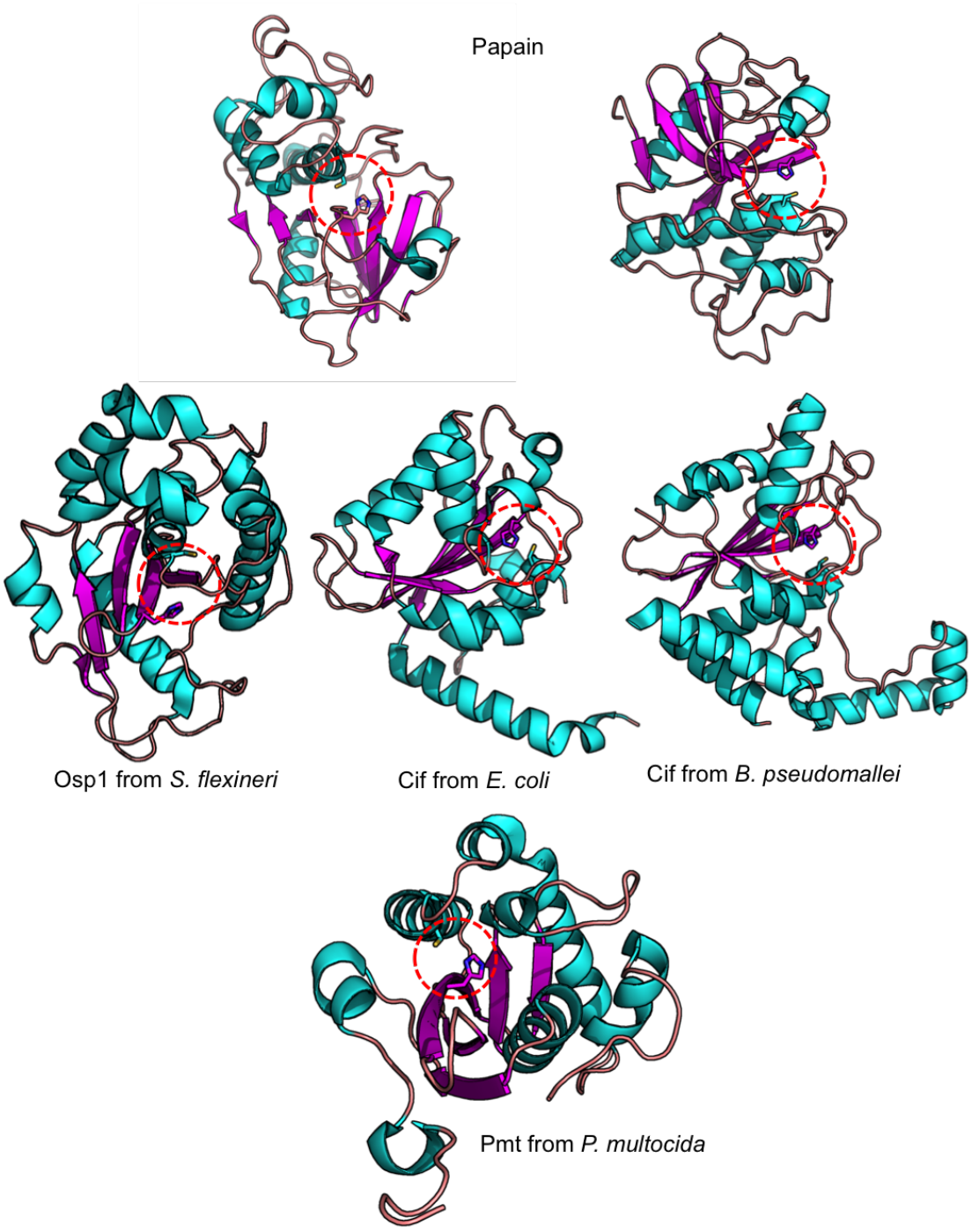


Fig. 1.7 **Papain-like family of GDTs.** The comparison between the structure of papain (PDB:9PAP) (top) with the papain-like GDTs showing a similar domain organisation consisting of a central β-sheet surrounded by α-helices (Osp1, PDB:3B21; Cif from *E.coli*, PDB:3EFY; Cif from *B. pseudomallei*, PDB:3GQM; Pmt, PDB:2EBF). The catalytic Cys-His pair in each member is circled red. Papain is rotated to show the Cys-His pair in a similar orientation to Osp1 and Pmt (top left) and the Cif deamidases (top right).

These toxins utilize deamidation to modify different host target proteins, to which they have evolved different substrate specificities. The molecular targets include E2 Ubiquitin conjugating enzyme (Osp1) (Sanada et al., 2012), heterotrimeric G proteins (Pmt) (Orth et al., 2009) and Ubiquitin and ubiquitin-like proteins (Cifs) (Crow et al., 2012; Yao et al., 2012) all of which are involved in a range of cellular processes ranging from phospholipase activation, adenyl cyclase inhibition, disruption of ubiquitin-proteosome degradation, and also an assortment of signalling pathways. These processes, when disrupted following deamidation of key glutamine residues, impose serious damage to host cells. Crystal structures of many toxins in this family in their apo form and in complex with their target proteins have been determined to shed light on the molecular basis of their specificity (Figure 1.8) (Washington et al., 2013). The conservation of the papain-like fold as well as the Cys-His dyad would imply that they have evolved by a process of divergent evolution hence suggesting that they share the same mechanism of action.

1.4 The CNF1-like family of GDTs

The second family of deamidase toxins that forms the main focus of this study, arose following the structural determination and the characterisation of the catalytic domain of *E. coli* CNF1 (C-CNF1). The structure of C-CNF1 can be seen to adopt a totally novel fold consisting of two β-sheets surrounded by α-helical moieties but containing a Cys-His pair in the active site whose relative orientation is reminiscent of papain (Buetow et al., 2001) (Figure 1.8). C-CNF1 shares essentially no sequence similarity to papain and the occurrence of a Cys-His pair in a similar orientation is an example of convergent evolution as the active site residues in CNF1 (Cys866 and His881) are separated by 14 residues whereas those in papain (Cys25 and His159) are separated by 133 residues (Figure 1.8).

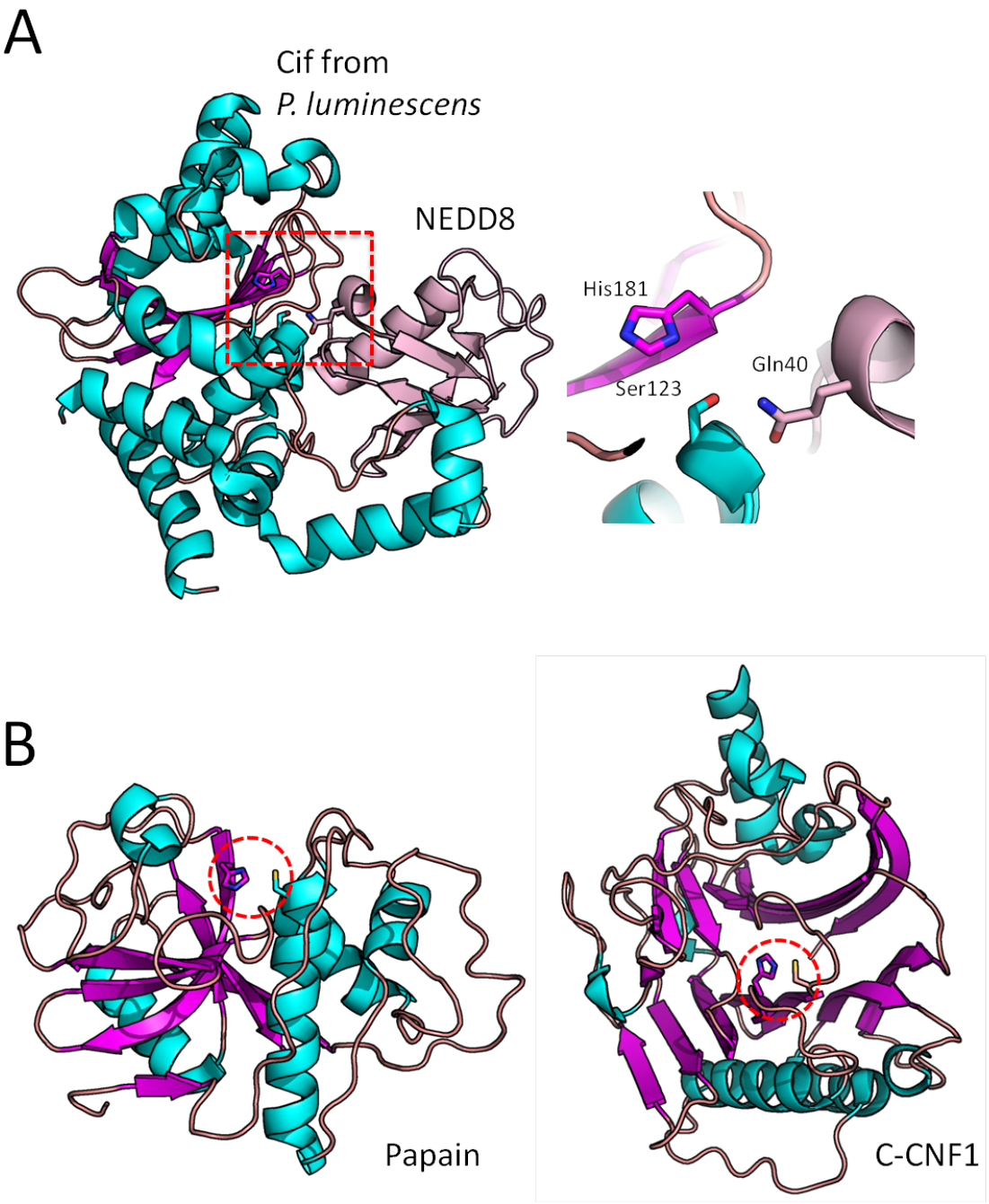


Fig. 1.8 **A papain-like deamidase in complex with its substrate and comparison of the tertiary fold of papain and C-CNF1.** A) The crystal structure of a Cys123Ser mutant of Cif deamidase from *P. luminescens* in complex with NEDD8 (PDB:4FBJ)(left) showing that the target Gln40 residue of NEDD8 is placed near the active site residues (red square, zoomed in on the right panel). B) Papain (PDB:6PAD) consists of only a single β-sheet as compared to two in C-CNF1 (PDB:1HQ0). The relative orientation of the Cys-His pair in their active sites is however maintained (red circle).

CNF1 has several homologs with high sequence homology namely CNF2 and CNF3 from other enteropathogenic *E. coli* strains and CNFY from *Yersinia pseudotuberculosis* (85, 70, and 65 % sequence identity, respectively) (Lockman et al., 2002; Orden et al., 2007; Oswald et al., 1989). However, only the crystal structure of C-CNF1, among these homologs has been determined.

1.5 The discovery of more CNF1-like family of GDTs.

Following the discovery of CNF1, it was thought that this family of deamidases was limited only to toxins closely related to CNF1 until the recent discovery and structural determination of Burkholderia Lethal Factor 1 (BLF1) from *Burkholderia pseudomallei* and HCH_03101 from *Hahella chejuensis* that share a broadly similar fold despite having very low sequence identity. These discoveries have led to the establishment of a CNF1-like family of GDTs with at least two members as described below (Cruz-Migoni et al., 2011).

1.5.1 The discovery of Burkholderia Lethal factor 1 (BLF1).

***B. pseudomallei*; an emerging global threat**

B. pseudomallei is a resilient, soil-dwelling, Gram-negative bacteria that is resistant towards a wide array of antibiotics and has the ability to withstand a range of otherwise hostile environments with more extreme values of pH and temperature (Currie, 2015). *B. pseudomallei* is a causative agent of the disease melioidosis, a potentially fatal tropical disease endemic in South East Asia regions including Vietnam, Cambodia, Thailand, Malaysia and also northern Australia (Wiersinga et al., 2006). Moreover, the ability of this organism to remain quiescent for a long time, as exemplified in a case whereby recrudescence was recorded 62 years after initial exposure, in a patient who was taken as the prisoner of war by the Japanese during World War II, represents a source of major concern (Ngauy et al., 2005). Concerns over

re-activation following latency in returning US military personnel from the Vietnam conflict had bestowed the title “The Vietnam Time-bomb” on the organism (Clayton et al., 1973; Currie, 2003). Furthermore, melioidosis is often misdiagnosed as tuberculosis or pneumonia, having almost identical symptoms and thus *B. pseudomallei* has been granted the name “The Great Mimicker” (Singh and Mahmood, 2017). Clinical manifestations of melioidosis vary from acute septicaemia, sub-acute and chronic disease and subclinical infections (Wiersinga et al., 2012, 2006). Infections could spread into multiple organs forming abscesses in lungs, spleen, liver and prostate (White, 2003). To date, there is no active vaccine that can be used to provide protection against melioidosis, and current therapy involves weeks of antibiotic treatment administered in two phases. In the first phase, intensive intravenous antibiotic treatment is provided which is followed by an oral eradication therapy in the second phase (Currie, 2015). However, recrudescence cases occurring in 10 % patients who have undergone 20 weeks of antibiotic treatment highlight the need of a new and effective treatment strategy (White, 2003).

A proteomics study comparing protein expression between *B. pseudomallei* and a related non-pathogenic strain *B. thailandensis* via a two dimensional SDS PAGE (2D-PAGE) identified a number of proteins that are only expressed or expressed in higher level in the pathogenic strain, *B. pseudomallei*, thus providing directions towards identifying virulence factors (Wongtrakoongate et al., 2007). This research has been a major reference point in carrying out studies to isolate potent factors that contribute to the pathogenesis of this lethal bacterial species. One of the proteins that was found to be over expressed in *B. pseudomallei* but not in *B. thailandensis* was the hypothetical protein BPSL1549.

The BPSL1549 gene encodes a potent protein toxin

Recently, extensive biochemical and molecular biology studies on BPSL1549, identified that it encodes a potent toxin, the first of its kind from *B. pseudomallei*, later coined Burkholderia Lethal Factor 1 (BLF1) (Cruz-Migoni et al., 2011). The three-dimensional (3D) structure of BLF1 has been solved revealing a broad tertiary structure similarity to C-CNF1, consisting of two β-sheets surrounded by α-helices with the predicted active site bearing a conserved leucine-serine-glycine-cysteine (LSGC) motif containing the catalytic cysteine, Cys94, equivalent to Cys866 in CNF1 (Figure 1.9) (Cruz-Migoni et al., 2011). Sequence alignment and structural superposition between these two toxins also led to the identification of the conserved histidine, His106 in BLF1 (aligned with His881 in CNF1) that together with the Cys94, make up the critical papain-like Cys-His pair (Figure 1.9) (Cruz-Migoni et al., 2011). However, unlike CNF1 that possesses N-terminal cell-binding and translocation domains, *BLF1* gene is only limited to the C-terminal domain giving rise to questions on how BLF1 is transferred across host cell membrane (Figure 1.10).

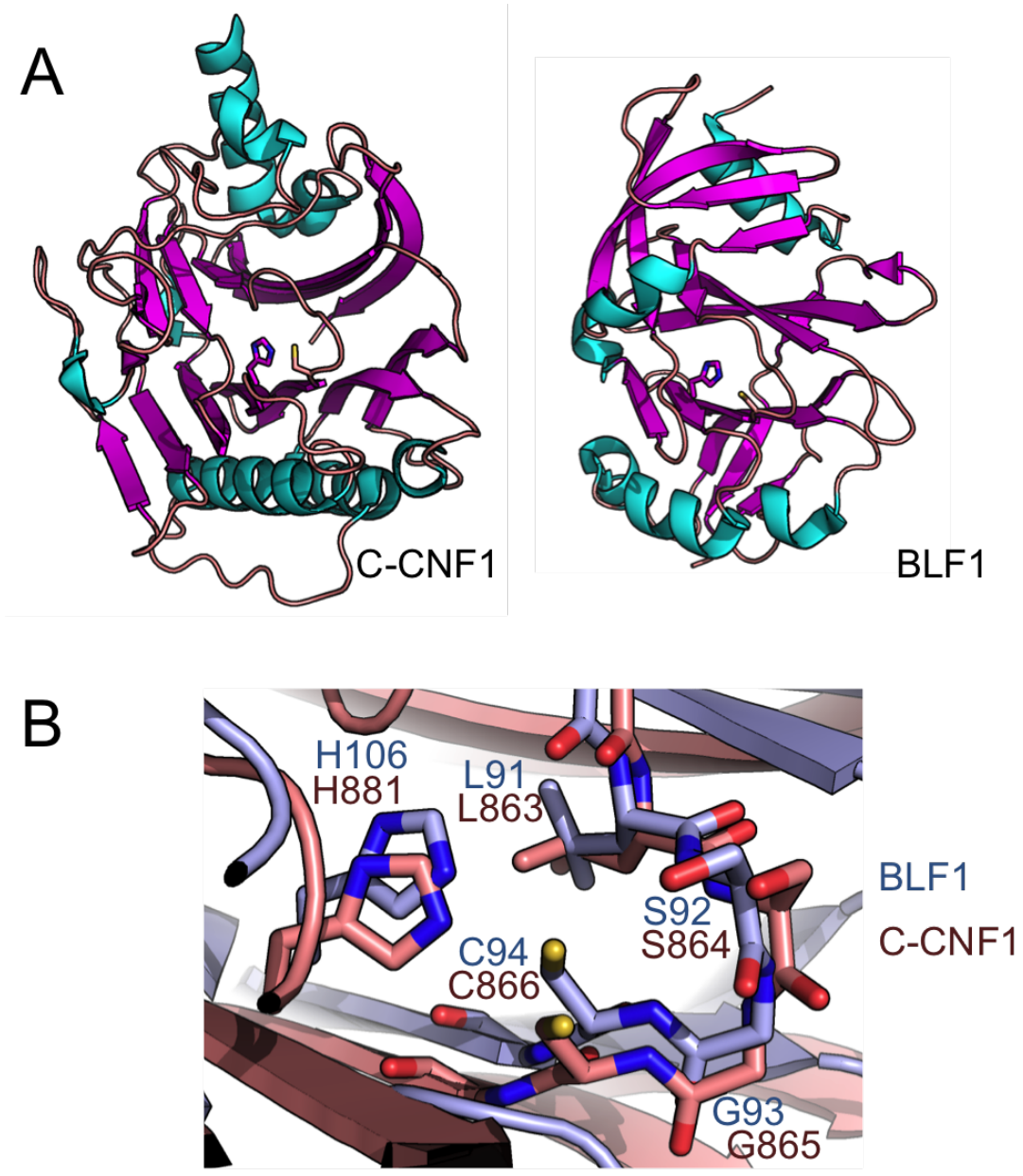


Fig. 1.9 **Comparison between C-CNF1 and BLF1.** A) C-CNF1 (PDB:1HQ0) and BLF1 (PDB:3TU8) share a broad similarity in the tertiary fold. B) Superposition of the two structures reveals the conservation of LSGC motif, containing the putative cysteine nucleophile, and a histidine residue making up the Cys-His pair.

As previously discussed, the CNF1 toxin was found to be a glutamine deamidase enzyme that acts on proteins involved in cell signalling including cdc42, Rho and Rac (Boquet, 2001). Despite having a similar overall fold and a conserved Cys-His pair in the active site, BLF1 was shown to bind to a different target protein owing to the broader and shallower molecular surface around the BLF1 active site in comparison to CNF1 (Cruz-Migoni et al., 2011). A pull-down assay on BLF1 with human cell extract revealed that BLF1 binds and forms a complex with eukaryotic Initiation Factor 4A (eIF4A), an RNA helicase, responsible for initiation-dependent protein translation suggesting that this toxin inhibits protein translation of host cells (Cruz-Migoni et al., 2011). An *in vitro* study carried out by treating J774 macrophages with BLF1 resulted in cell death within 72 hours (Cruz-Migoni et al., 2011). Balb/C mice treated with this toxin via intraperitoneal injection died by day 14, confirming the toxicity of this protein (Cruz-Migoni et al., 2011). LC MS/MS on tryptic digest of eIF4A peptides that were incubated with BLF1 toxins identified the site of deamidation as the glutamine residue at position 339 (Gln339) of eIF4A (Cruz-Migoni et al., 2011). Consistent with the abolition of deamidase activity in C-CNF1 by mutating the catalytic cysteine, Cys886, to serine, BLF1 C94S mutant was found to be non-toxic (Cruz-Migoni et al., 2011). Due to the absence of the cell binding and translocation domains, it is not currently certain how BLF1 enters the cytoplasm (Figure 1.10). However, the intracellular lifestyle of *B. pseudomallei* bypasses the need for BLF1 to have a host cell recognition and membrane translocation domains (Allwood et al., 2011).

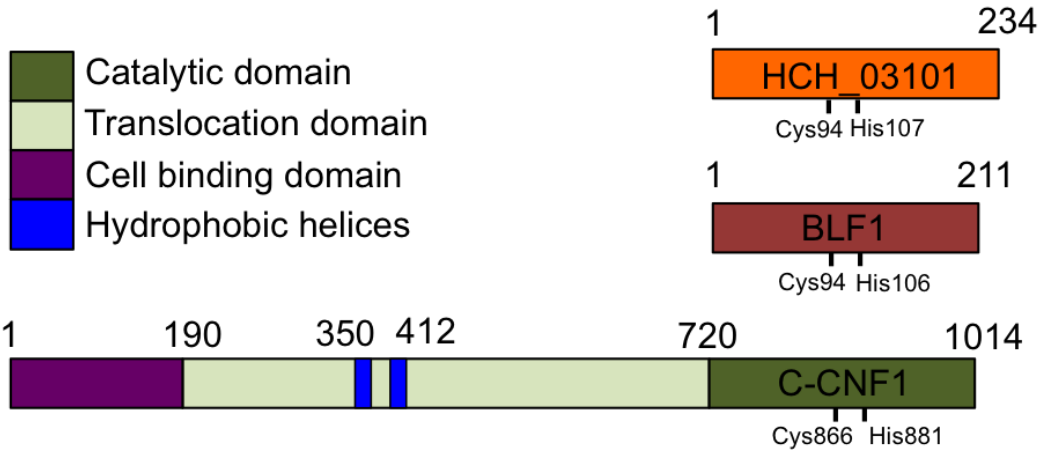


Fig. 1.10 **Domain organisation of the other members of the CNF1-like family of GDTs.** Unlike CNF1, BLF1 and HCH_03101 do not consist of host cell recognition and translocation domains.

1.5.2 HCH_03101 from *Hahella chejuensis* shares 16% sequence identity to BLF1

H. chejuensis is a putative pathogen isolated from a Korean shore

H. chejuensis is a gram-negative, aerobic, rod-shaped and motile bacterium belonging to the γ-*Proteobacteria* class. A bacterial sample isolated from a marine sediment, near Cheju island, Korea, was found to produce abundant extracellular polysaccharides, and red pigments (Bae et al., 2001). The bacteria also exhibit a slight halophilic property as growth is inhibited in the absence of NaCl, whereby optimum growth occurs in medium containing 2 % (w/v) NaCl concentration (Bae et al., 2001). Following a complete genomic DNA sequence determination of this bacterial strain, and identification of several genes in genomic islands possibly involved in encoding virulence factors including a Type III secretion system (T3SS), the possibility that *H. chejuensis* might be a potential marine pathogen emerged (Jeong et al., 2005). Purified red pigment produced by the bacteria, was found to be the antibiotic prodigiosin through ^{1}H- and ^{13}C- NMR and was shown to cause cell lysis of *Cochlodinium*

polykrikoides, a dinoflagellate micro-algae that is responsible for harmful algal blooms (HAB) or generally known as the red-tide (Jeong et al., 2005). Interestingly, the red pigment is also known for having a broad-range cytotoxic activity (Fürstner, 2003) as well as having anti-cancer properties inducing apoptosis in cancer cells (Pérez-Tomás et al., 2003).

HCH_03101 has a 3D structure almost identical to that of BLF1 but with an interesting added feature

Sequence searches on the genome of *H. chejuensis* identified a distant homolog of BLF1 encoded by the gene HCH_03101. This gene codes for a protein product of 26 kDa with 234 amino acids sharing 16% sequence identity to BLF1. Structural studies have shown that the conserved residues are concentrated around the active site, and include the conserved LSGC motif containing the catalytic cysteine residue and the associated histidine residue (Figure 1.11). In addition, structure-based sequence alignment between HCH_03101 and BLF1 revealed a 32-residues long insertion in the sequence of HCH_03101, between residues 165 and 166 in BLF1, which was shown to form an extended loop (see below) (Figure 1.11 & 1.12). Furthermore, HCH_03101 lacks the cell binding and membrane translocation domains suggesting that *H. chejuensis* may be an intracellular pathogen (Figure 1.10).

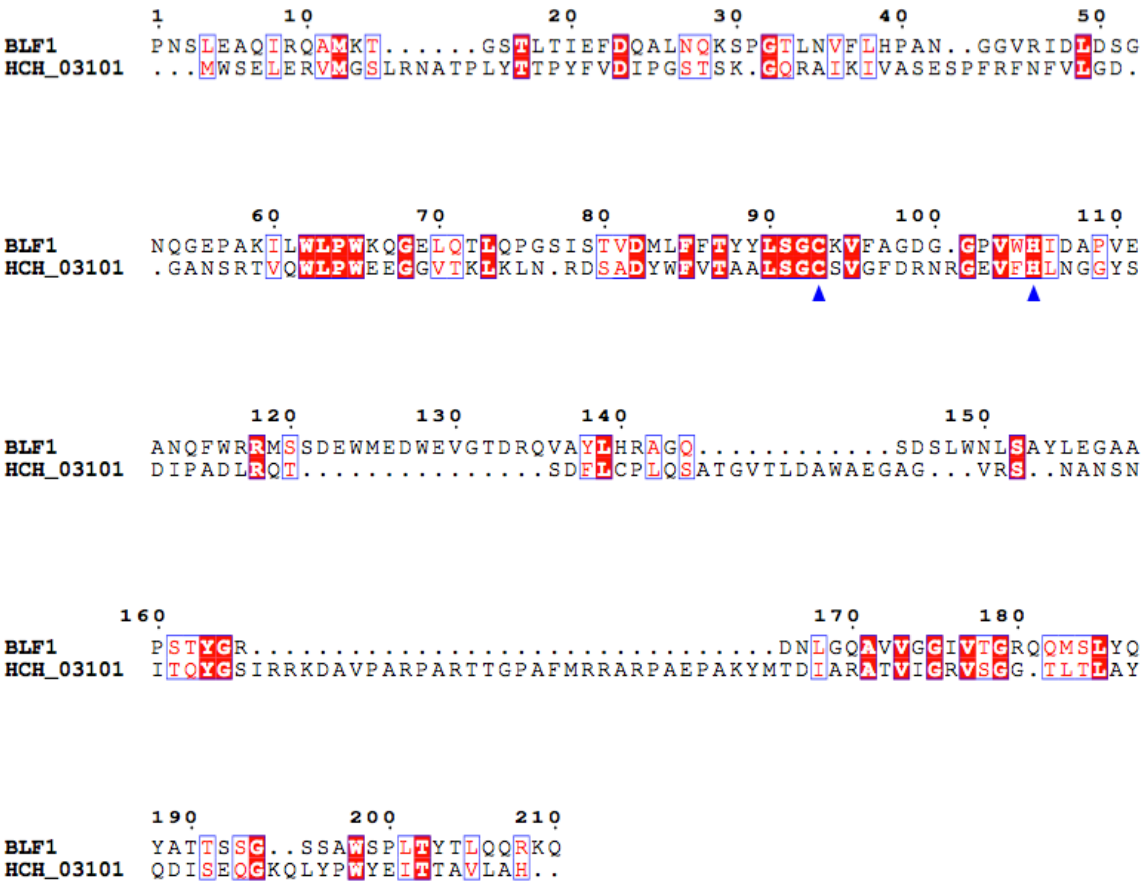


Fig. 1.11 **Structure-based sequence alignment between HCH_03101 and BLF1.** The LSGC motif, containing the cysteine, and the associated histidine, are conserved in HCH_03101 (blue triangles).

The crystal structure of HCH_03101 was determined via Se-Met derivative Single Anomalous Dispersion (Se-SAD) method (George Mobbs, PhD thesis, The University of Sheffield Library, 2014) (Mobbs, 2014) (Figure 1.12). Superposition of HCH_03101 with BLF1 gives an rmsd value of 2.1 Å across 148 equivalent C_α showing a very similar tertiary fold, consisting of two β-sheets surrounded by α-helices (Figure 1.12). The similar orientation of the catalytic Cys-His dyad in the active site cleft as that seen in BLF1 confirms the similarity between these two toxins being in the same family of GDTs (Figure 1.12). However, a fascinating extra feature can be seen in the crystal structure of HCH_03101, whereby there is a protrusion of an elongated loop between β-15 and β-16, which extends for a length of 44 Å almost as long as the diameter of the main globular domain (Figure 1.12). The protrusion is also seen to have a repeating proline-alanine rich motif, A-X-P-A-X-P-A, which possesses a pseudo two-fold symmetry (Figure 1.13). The nature of the protrusion that is rich in charged residues has led to a proposal that this region may act as an extension to the molecular surface that cradles the target binding partner, and could be implicated in substrate recognition in addition to providing structural support for the formation of enzyme-target complex (Figure 1.14).

Analysis of the structures of the WT and the C94S mutant of HCH_03101 indicates that they are very similar (rmsd value 0.74 Å over 221 equivalent C_α carbons) apart from minor changes as a result of the mutation. These include a change in the torsion angle of the mutated S94 which forms a hydrogen bond to the His107 in the catalytic dyad which in turn imposes minor structural shift in the surrounding region. An interesting feature in the structure of the C94S mutant is the slight motion observed in the protrusion when compared with that of the WT, which implies the presence of a possible hinge region that allows flexibility and therefore might be related to function (Figure 1.14).

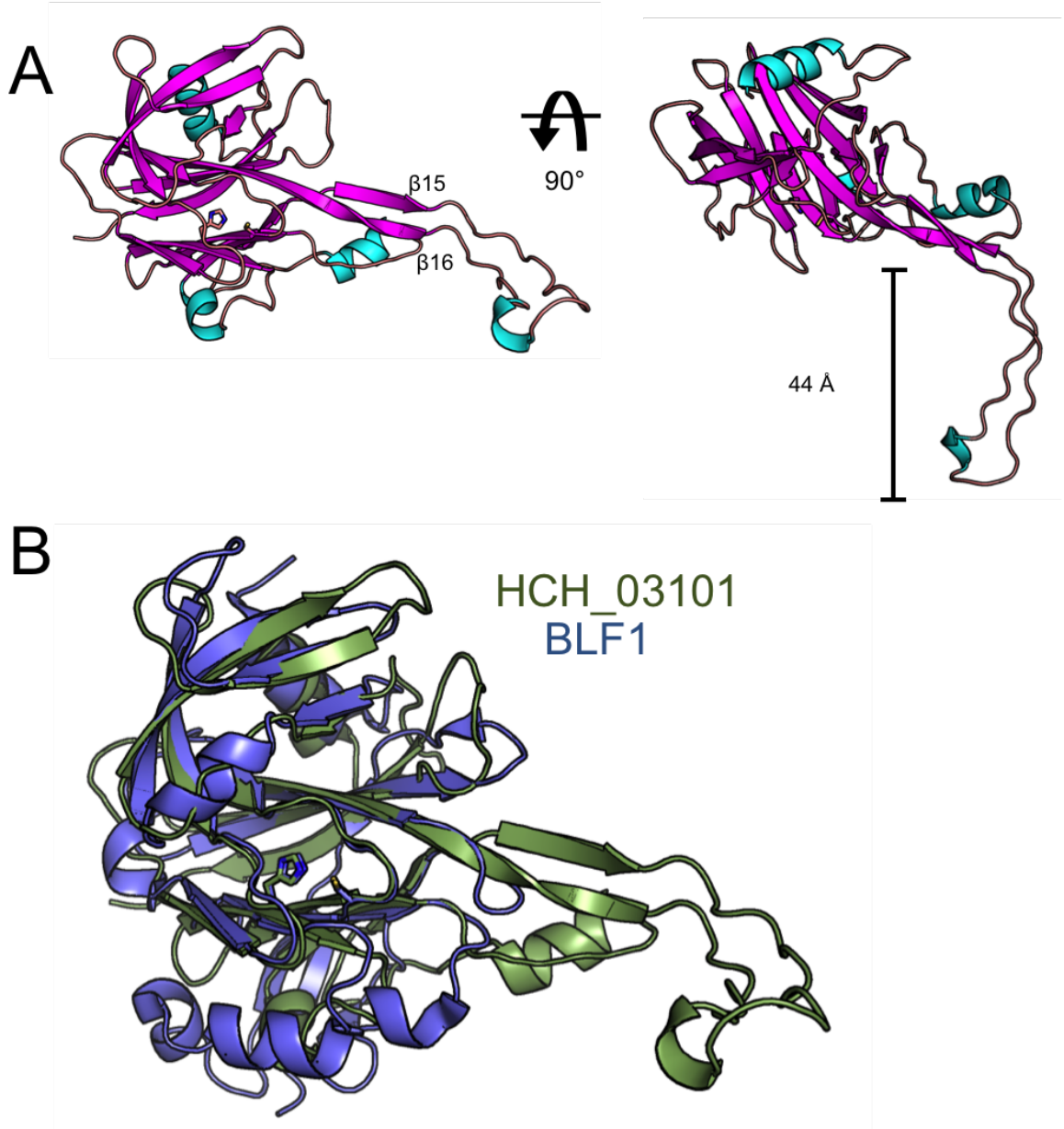


Fig. 1.12 **Comparison between the structures of HCH_03101 and BLF1.** A) Crystal structure of HCH_03101 reveals a CNF1-like fold with an extended protrusion between β-15 and β-16. B) Superposition of the structures of HCH_03101 and BLF1, showing almost similar fold. The Cys-His pair is shown as sticks.

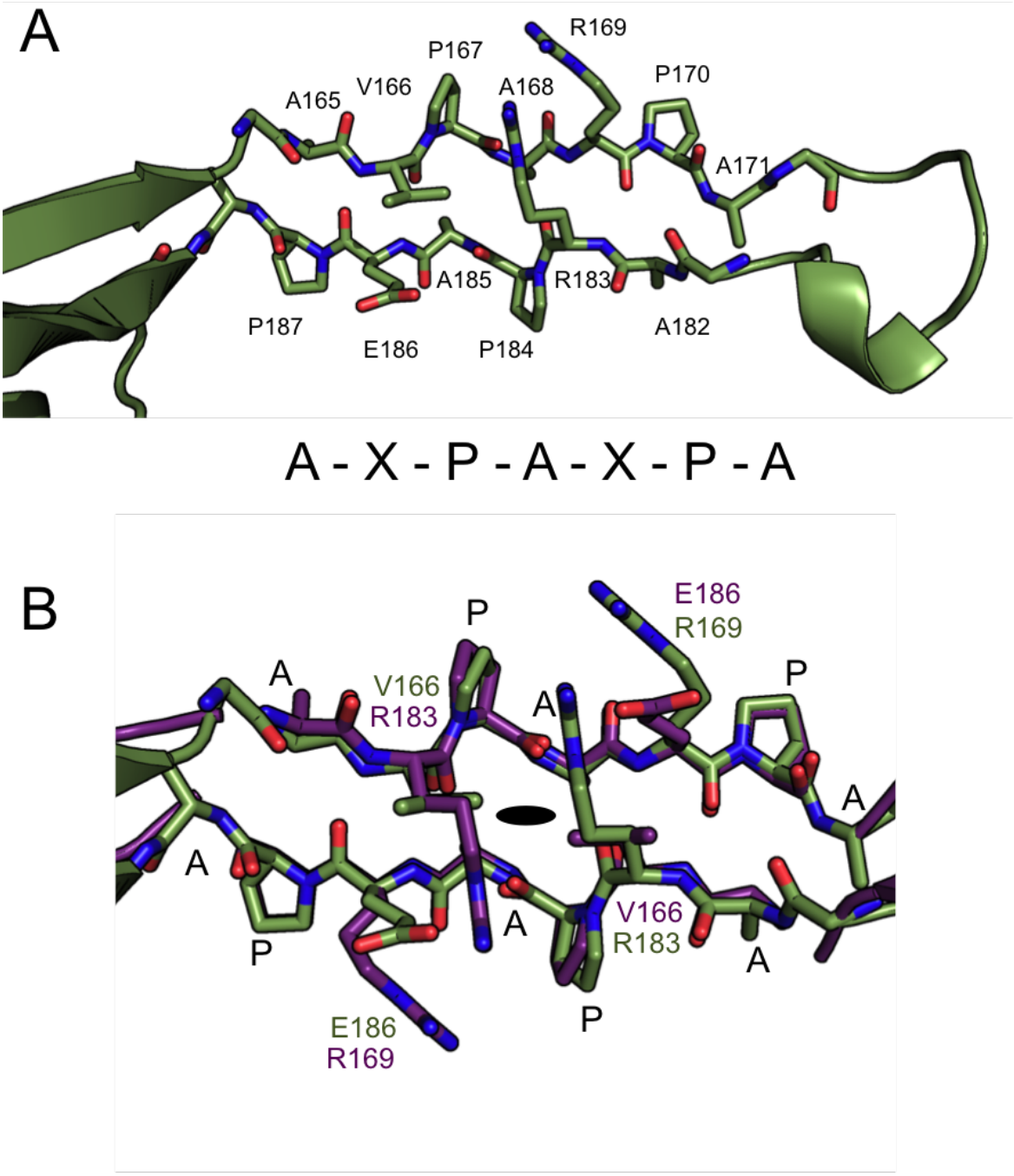


Fig. 1.13 **Features of the extended protrusion in HCH_03101.** A) The proline-alanine rich sequence repeat in the protrusion. B) Superposition of the repeating A-X-P-A-X-P-A motif showed a structural conservation with a pseudo two-fold symmetry (black oval).

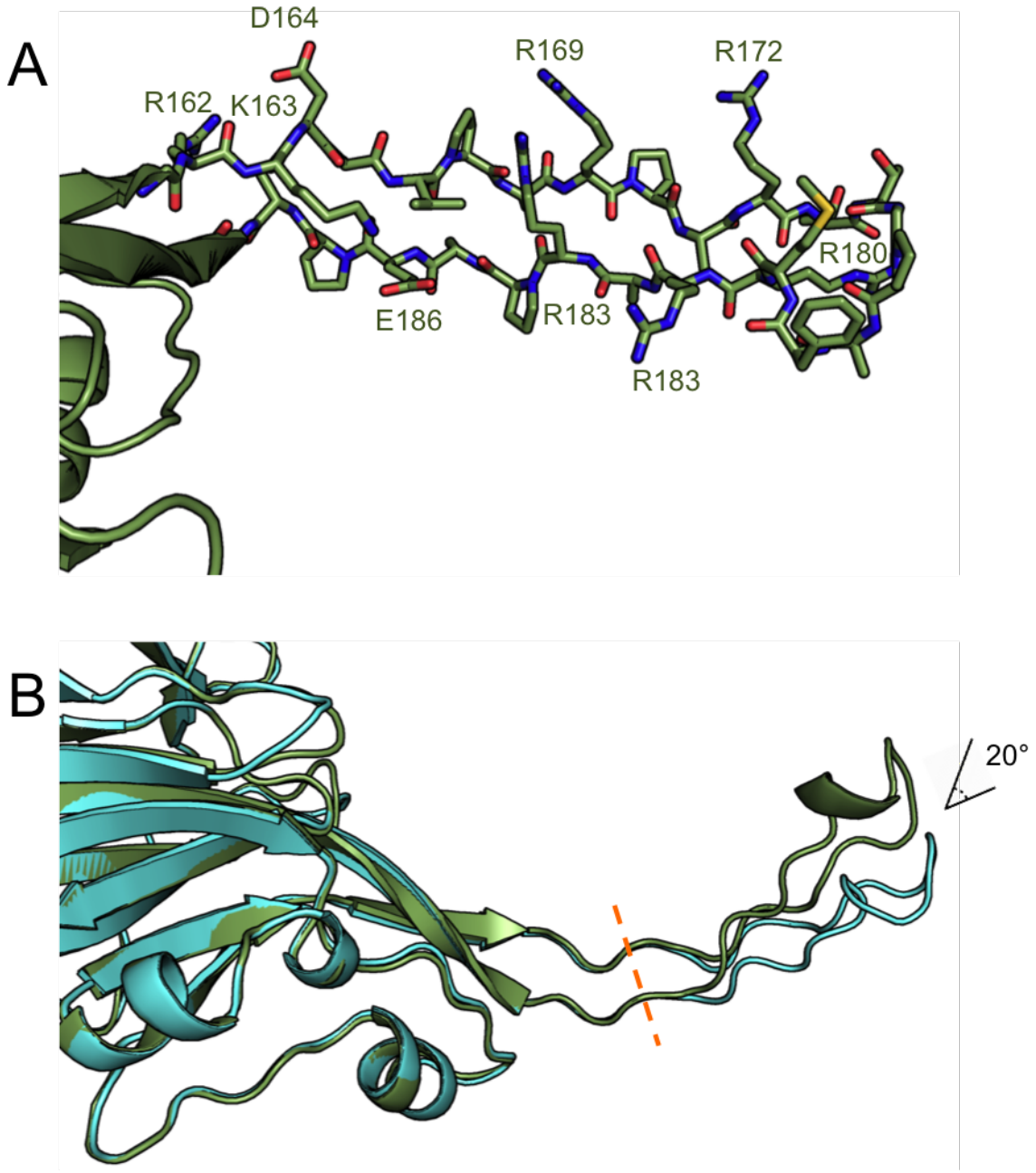


Fig. 1.14 **Features of the extended protrusion in HCH_03101.** A) A significant number of charged residues are located along the protrusion which may have roles in specificity and recognition. B) Superposition between WT HCH_03101 (green) and the C94S mutant (cyan) revealed a rotation of 20^{o} around a possible hinge region (orange dashes).

HCH_03101 binds algal initiation factor from algal extract pull-down assay

Due to the fact that HCH_03101 conserves the Cys-His dyad as well as sharing a substantial similarity in the tertiary structure with BLF1, it was proposed that HCH_03101 is a functional BLF1 homolog that may share a similar target protein. To investigate this, a pull down assay was done on J774 macrophages cell-free extract by immobilising His_6-tagged HCH_03101 on a Nickel-NTA resin column. The binding partners were then eluted with high salt buffer and run on SDS-PAGE gel before being analysed by MS/MS mass spectrometry on trypsin-digested gel bands and MASCOT searches to identify the actual peptides. The SDS-PAGE gel identified a binding partner similar in size to that of BLF1 protein target, eIF4A, and this was confirmed by the MASCOT searches (Mobbs, 2014).

Although HCH_03101 was found to bind human eIF4A, the coastal environment that *H. chejuensis* lives in would imply that, as with *B. pseudomallei*, man is not the main target of the organism. As *H. chejuensis* was found to cause the lysis of a red-tide microalgae (Jeong et al., 2005), algal species would be the obvious target host. To test this, a similar pull down assay approach was carried out on algal cell-free extract from a green algae *Tetraselmis cuecica*. However, due to a limited availability of a complete genomic sequence of *T. cuecica*, identification of actual targets proved difficult, although the presence of a conserved motif within the bound protein enabled MASCOT search to identify sequences related to other initiation-factor proteins from other completely sequenced algal species (Mobbs, 2014). This result suggests that HCH_03101 produced by *H. chejuensis* has affinity for eIf4A variants in algal species, and if the protein does deamidate these targets, it could well be toxic and may kill algal cells. Therefore, although the active component prodigiosin released by *H. chejuensis* was shown to have lytic activity on the microalgae *C. polykrikoides*, this does not exclude the possible involvement of HCH_03101 as part of *H. chejuensis* virulence.

1.6 Aims and objectives

Following the identification and characterisation of CNF1 as the first toxin that exhibits glutamine deamidase activity, a number of other GDTs were discovered in a variety of pathogenic bacteria targeting the deamidation of critical glutamine in various target proteins. However, these toxins were found to have a tertiary fold similar to that of papain and were classified as papain-like family of GDTs. Currently, there are only limited members in the CNF1-like family of GDTs, namely the C-CNF1 toxin from Uropathogenic *E. coli*, BLF1 from *B. pseudomallei* and HCH_03101 from *H. chejuensis*. The primary aim of this thesis is therefore to expand the CNF1-like family of GDTs by means of structural determination of new family member by X-ray crystallography. The structures of these toxins could then be analysed to enhance our understanding of the structure-function relationships of this toxin family.

At present, crystal structures of a number of the papain-like family of GDTs had been solved in complex with the biological target host proteins (Washington et al., 2013). The structures of these complexes have been central in elucidating the molecular basis of substrate specificity and thus the aspects of mechanism of these toxins. However, none of the toxins in the CNF1-like family of GDTs have their crystal structures solved in complex with their protein targets. Thus, information on substrate specificity and the mechanism of the toxins in the CNF1-like family is very scarce. In addition, pathogenesis of the bacteria producing these toxins, especially *B. pseudomallei* and *H. chejuensis*, is very poorly understood. This provides a very interesting gap to be explored and therefore the secondary aim of this study is to obtain crystal structures of the toxins in this family in complex with their target substrates to shed light on the molecular basis of substrate specificity as well as their mechanistic properties. The outcomes of this study would then help deepen our understanding of the pathogenesis of these bacteria and hopefully provide insights into development of therapeutics to counter the effects of these toxins and therefore help in the control of the disease they cause.

The aims of this study are therefore as follow:

1. Search for more members of the CNF1-like family of GDTs through BLASTp searches
2. Solve the structures of the proteins of interest by X-ray crystallography
3. Characterise the activity of these putative toxins via biochemical and biophysical techniques
4. Co-crystallise with target proteins to analyse the interactions that are formed between the toxins and their target to elucidate the molecular basis of specificity and aspects of the mechanism

Chapter 2

Materials and Methods

2.1 General recipes

2.1.1 Antibiotics

Kanamycin was used as the main antibiotic for this study to select for constructs harbouring kanamycin resistance marker and was prepared by dissolving 0.5 g of kanamycin sulphate salt in 10 ml of MilliQ® water making up a (1,000x) stock at 50 mg ml^{-1}.

2.1.2 Lysogeny Broth (LB) medium

LB medium (Miller, 1972) was prepared by mixing the components described in Table 2.1 in a total of 1 l of distilled water, followed by mixing before sterilising by autoclave at 121 ^{o}C for 15 minutes.

2.1.3 LB Agar

LB Agar was prepared by mixing the components as decribed in 2.1.2 in a 800 ml of distilled water, and mixed well before adding 15 g Bacto-Agar and distilled water up to 1 l. The medium was then sterilised by autoclave at 121 ^{o}C for 15 minutes.

Table 2.1 Recipe for LB medium

Components	Amount (g)
Tryptone	10
Yeast extract	5
NaCl	10

2.1.4 SOC media

Super optimal broth with catabolite repression (SOC) media was prepared as in the recipe in Table 2.2 in distilled water and sterilised by autoclave at 121 ^{o}C for 15 minutes (Hanahan, 1983).

Table 2.2 Recipe for SOC medium

Components	Final concentration
Tryptone	2 % (w/v)
Yeast extract	0.5 % (w/v)
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose*	20 mM

*Glucose was added after the other components were dissolved, autoclaved and allowed to cool.

2.1.5 Agarose gel

1 % agarose gels were used to analyse the size of DNA products against DNA markers and were prepared by mixing 0.3g Iberose in 30 ml TAE buffer (40 mM Tris pH 7.6, 20 mM acetic acid, and 1 mM EDTA) and dissolved by using a microwave at medium power. 3 μl of GelRed™ (Biotium) nucleic acids stain (10,000x stock) was added and the gel was let to solidify in the gel tank. 5 μl of Hyperladder 1kb (Thermo Fisher Scientific) was used as the marker and gels were run at 100 V for 45 minutes prior to visualisation under UV light.

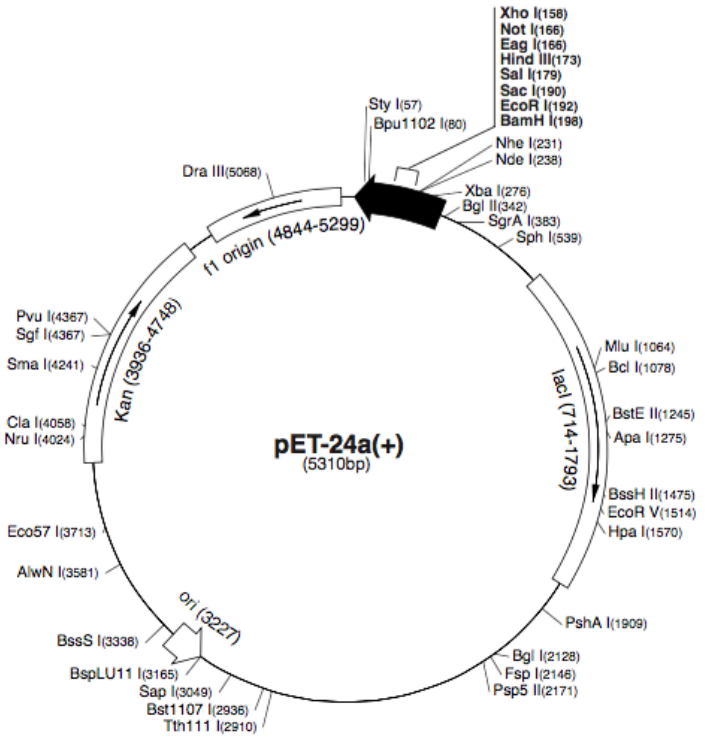
2.2 Strains and plasmid vector information

E. coli DH5α cells (NEB) were used to transform, propagate and purify plasmids. *E. coli* DH5α-silver efficiency cells (Bioline) were used during DNA cloning to increase transformation efficiency of ligated plasmid. *E. coli* BL21 (DE3) strain were used for T7 expression suitable for use with pET Recombinant Expression vectors (Novagen). pET-24a vector was used throughout the course of this study allowing tightly controlled expression inducible by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) and details on the vector is highlighted in Figure 2.1.

pET-24a(+) sequence landmarks

T7 promoter	311-327
T7 transcription start	310
T7•Tag coding sequence	207-239
Multiple cloning sites (*BamH* I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	714-1793
pBR322 origin	3227
Kan coding sequence	3936-4748
f1 origin	4844-5299

The maps for pET-24b(+), pET-24c(+) and pET-24d(+) are the same as pET-24a(+) (shown) with the following exceptions: pET-24b(+) is a 5309bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-24c(+) is a 5308bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198. pET-24d(+) is a 5307bp plasmid; the *Bam*H I site is in the same reading frame as in pET-24c(+). An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 238 of pET-24c(+). As a result, *Nco* I cuts pET24d(+) at 234, and *Nhe* I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-24a(+). *Nde* I does not cut pET-24d(+). Note also that *Sty* I is not unique in pET-24d(+).



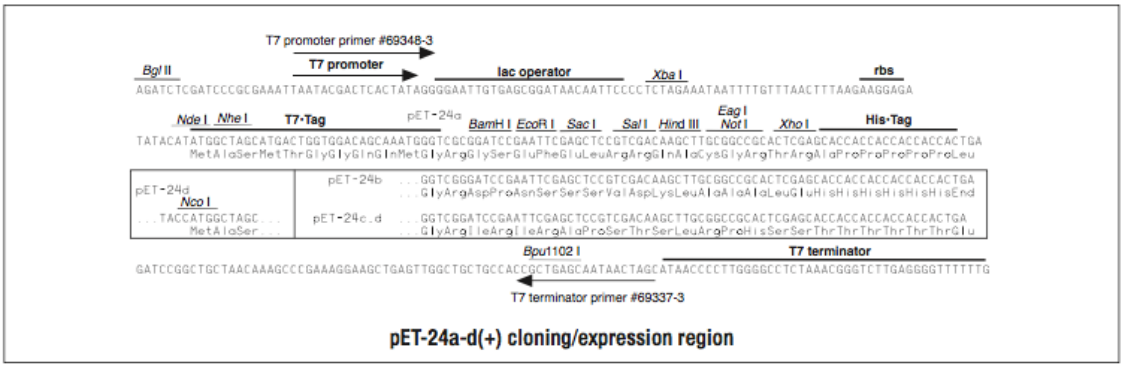


Fig. 2.1 Details on the pET-24a vector taken from Novagen

2.3 Gene Cloning

In order to clone the gene of interest either from genomic DNA or from a plasmid, forward and reverse primers were designed upstream and downstream of the gene respectively. The forward primer consists of a few nonsense nucleotide bases for restriction enzyme (RE) optimisation, followed by a restriction site of interest, an optional ATG start codon and a short length of about 20-30 bases of the gene at the 5' end. The reverse primer starts with 20-30 bases at the 3' end of the gene, followed by an optional TGA stop codon, a preferred restriction site and a few nonsense nucleotide bases. The melting temperature, T_m, of both primers were optimised by varying the length of the gene to obtain a T_m difference of < 5 ^oC with a value ranging from 50 to 72 ^oC. NEB Q5 T_m calculator online tool was used to calculate the T_m values and NEBCutter was used to check for and select compatible restriction enzymes.

2.3.1 Polymerase Chain Reaction (PCR)

PCR was conducted to clone the gene of interest using a Q5® High-Fidelity DNA Polymerase Kit (NEB™) according to the components in Table 2.3. All the components were mixed and collected to the bottom of PCR tubes by centrifugation. The PCR tubes were then transferred to a thermocycler and reaction conditions were set as Table 2.4.

Table 2.3 List of components and the required volumes for the PCR reactions by Q5® High-Fidelity DNA Polymerase Kit

Components	25µl reaction	50µl reaction	Final Concentration
5X Q5 Reaction Buffer	5 µl	10 µl	1X
10mM dNTPs	0.5 µl	1 µl	200 µM
10µM Forward Primer	1.25 µl	2.5 µl	0.5 µM
10µM Reverse Primer	1.25 µl	2.5 µl	0.5 µM
Template DNA	variable	variable	< 1000 ng
Q5 High-Fidelity DNA Polymerase	0.25µl	0.5 µl	0.02 U/µl
5X Q5 High GC Enhancer (optional)	(5 µl)	(10 µl)	1X
Nuclease-Free Water	Up to 25 µl	Up to 50 µl	

Table 2.4 Reaction temperature and time settings for the thermocycler

Step		Temperature	Time
Initial Denaturation		98 °C	30 seconds
25-35 Cycles	Denaturation	98 °C	5-10 seconds
	Annealing	*50-72 °C	10-30 seconds
	Elongation	72 °C	20-30 seconds
Final Extension		72 °C	2 minutes
Hold		4-10 °C	

*Annealing temperatures are chosen based on the T_m of primer

2.3.2 Restriction enzyme digest

Following PCR, the amplified gene was then treated with selected restriction enzymes to produce sticky overhangs to allow ligation to plasmid vectors of choice at the Multiple Cloning Site (MCS). The plasmid vector and the gene insert were separately treated with the corresponding restriction enzymes as per instructions in Table 2.5. If identification and isolation of specific length of digested DNA were required, samples following restriction enzyme digest were subjected to 1% Agarose Gel Electrophoresis against 1kb Hyperladder DNA Marker. The DNA bands were viewed under UV light and the correct DNA bands were purified using QIAquick Gel Extraction Kit (Qiagen).

Table 2.5 Instructions for restriction enzymes incubation

Restriction Enzyme	10 units, 1 µl
DNA	1 µg
10X NEBuffer	5 µl (1X)
Total Reaction Volume	50 µl
Incubation Time	1 hour
Incubation Temperature	Enzyme dependent

2.3.3 DNA Gel Extraction using QIAquick Gel Extraction Kit

DNA fragments of desired length were excised from the gel with a sterilised scalpel and mixed with 3 volumes of Buffer QG for 1 volume of gel (100 mg of gel counts to about 100 μl). The gel was then dissolved in the buffer by incubating the mixture at 50 ^{o}C for 10 minutes with intermittent vortexing. The colour of the dissolved mixture was checked to ensure a yellow colour indicating an acidic pH; in situations where the colour was orange or violet indicating a basic pH, 10 μl of 3 M sodium acetate pH 5.0 was added. 1 volume of

isopropanol was added to the sample, which was then mixed and applied to a spin column and centrifuged for 1 minute at 13k rpm. The flow-through was discarded and the spin column was washed with 0.75 ml Buffer PE and centrifuged for 1 minute. The flow-through was discarded and the column was centrifuged for another minute to get rid of residual ethanol from Buffer PE. The spin column was put into a microcentrifuge tube and 50 μl of autoclaved MilliQ® water was added to the center of the membrane and left to incubate for about 10 minutes. DNA was then eluted by centrifugation at 13k rpm for 1 minute.

2.3.4 PCR clean-up using QIAquick PCR Purification Kit

In situations where agarose gel purification was not necessary, PCR clean-up was done using PCR Purification Kit (Qiagen). 5 volumes of Bufffer PB was added to 1 volume of PCR sample. The sample was then mixed and transferred into a spin column and centrifuged. The bound DNA was then washed with 0.75 ml Buffer PE and centrifuged at 13k rpm for a minute. Flow-through was discarded and the column was centrifuged for an additional 1 minute. The column was then transferred into a clean microcentrifuge tube. 50 μl of autoclaved MilliQ® water was added to the membrane, let stand for 10 minutes and DNA was eluted by centrifugation.

2.3.5 Vector/Insert Ligation

To generate a full construct of recombinant plasmid, the vector and the gene insert were ligated utilising a T4 DNA Ligase Kit (NEB®). To ensure successful ligation, the insert was ligated to the vector in excess of at least 3:1 molar ratio. The required vector and insert masses were calculated using NEBioCalculator (available at http:// nebiocalculator. neb.com). The amount of vector and insert DNA used may vary, but in general, for a 4kb vector and a 1kb insert DNA, a 3:1 insert:vector ratio involves the mixture of the components as in Table 2.6. The reaction components were mixed and incubated either at 16 ^{o}C overnight or

at room temperature for 10 minutes. This was followed by heat inactivation at 65 °C for 10 minutes. 1-5 μl of the ligation mixture was then transformed into 50 μl DH5α silver efficiency competent cells.

Table 2.6 Components and required volumes for vector/insert ligation

Component	20 µl Reaction
10X T4 DNA Ligase Buffer	2 µl
Vector DNA (4kb)	50 ng (0.020 pmol)
Insert DNA (1kb)	37.5 ng (0.060 pmol)
Nuclease free water	Up to 20 µl
T4 DNA Ligase	1 µl

2.4 Plasmid DNA propagation, purification and sequencing

For successful gene cloning experiments, plasmid/vectors are often required at high concentration with significant levels of purity. Propagation of plasmids was carried out by transformation of the plasmid of interest into *E. coli* DH5α cells as described in Section 2.5 and grown overnight at 37 °C in 5-10 ml LB cultures supplemented with the appropriate antibiotics. Cells were then harvested for purification of the plasmid using Monarch® Plasmid Miniprep Kit (NEB), which implements alkaline lysis technique to extract plasmid DNA (protocol is provided with the kit or can be accessed online). For DNA sequencing, 5 μl of 60 ng μl^{-1} plasmid DNA was mixed with 5 μl of either T7 forward or reverse primer (compatible with pET vectors) and sent for Sanger sequencing with GATC Biotech (Germany).

2.5 Transformation of recombinant plasmids

The competent *E. coli* cells were thawed on ice for 10 minutes followed by finger-flicking to ensure complete resuspension. 1 μl of 1 pg to 100 ng plasmids or ligation mixture was added and gently mixed with finger-flicking. The cells were then incubated on ice for 5 minutes. The cells were then subjected to a heat-shock by submerging the tubes into a 42 ^{o}C water bath for 40 seconds. The cells were then incubated on ice for 10 minutes after which 200-950 μl of SOC medium was added and the mixture was incubated at 37 ^{o}C with 250 rpm shaking for 1 hour. 100 μl of the mixture was plated on LB Agar plates with selected antibiotics. The plates were then incubated at 37 ^{o}C overnight.

2.6 Colony PCR

Colony PCR allows a quick screening of colonies that potentially contain the plasmids with the correct recombinant gene. A number of colonies from transformation plates were selected and inoculated into 100 μl of LB each with selected antibiotic. The cultures were then incubated at 37 ^{o}C for 4 hours. 50 μl from each culture was isolated and boiled for a few minutes. The remaining culture was plated on LB antibiotic agar for further analysis. 4 μl of each boiled culture was then mixed with 0.5 μl of 5 μM T7 forward primer, 0.5 μl of 5 μM T7 reverse primer, and 5 μl of BioMix Red (Bioline), which is a master mix containing Taq polymerase, dNTPs, and dye for agarose gel analysis. The samples were then subjected to PCR with standard protocol and analysed by 1% agarose gel, stained with GelRed™ against DNA Hyperladder 1kb to compare gene insert sizes. Plasmids from the colony that possessed the correct gene size were then purified and sent for DNA sequencing with GATC Biotech (Germany).

2.7 Protein Overexpression trials

The protein expression system used in this project was the pET expression system where pET24a/d (Kanamycin resistance) plasmids were used as vectors carrying the recombinant gene. The recombinant plasmids were previously transformed into BL21 (DE3) from Novagen. Selected transformants from the agar plates were inoculated in 5 ml LB medium supplemented with antibiotics and incubated at 37 ^{o}C with 250 rpm shaking overnight. 0.5 to 1 ml of the overnight culture was added into 50 ml LB-antibiotic and incubated further to get an optical density at 600 nm ($O.D_{600}$) of 0.6. 1 mM of IPTG was added to induce protein expression. The cultures were then incubated with 250 rpm shaking across temperatures ranging from 12, 18, 25, 30 and 37 ^{o}C to select for the best condition that favours production of soluble proteins. 5 ml samples were taken at specific intervals (1-2 hour interval for higher temperatures and 8-12 hour interval for lower temperatures) to be subjected to Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) analysis. The samples were centrifuged at 4800 rpm to collect the cell pellet and and then stored in -20 ^{o}C freezer prior to SDS-PAGE analysis.

2.8 SDS PAGE analysis

The frozen cell pellets were thawed and resuspended in 500 μl of 50 mM Tris pH8 buffer ensuring complete resuspension with no cell clumps. The cells were then lysed on ice using a sonicator with two 4 seconds bursts. The lysates were centrifuged at 14800 rpm for 20 minutes at 4 ^{o}C to separate soluble fractions. The soluble fraction of each sample was separated into a new eppendorf tube while the cell debris was resuspended in 500 μl of 50 mM Tris pH8 buffer. The total protein content in the soluble fraction was then measured using Bradford's Assay. Next, 10 μg of proteins each from the soluble and insoluble fraction of the cell lysate was mixed with 5 μl Invitrogen NuPAGE® LDS Sample Buffer (40 %

Glycerol, 4 % Lithium dodecyl-sulphate, 0.025 % Phenol red, 0.025 % Coomassie G250, 2 mM EDTA, 0.8 M Tri-ethanolamine-Cl pH7.6), 2 μl reducing agent (10x 1 M Dithiothreithol (DTT)) and sterilised MilliQ® water to make up to 20 μl total volume. Then, the mixture was boiled at 95 ^{o}C for 3 minutes before loading it on an SDS-PAGE gel. The recipe used for SDS-PAGE was adapted from (Laemmli, 1970) and all the gels used for this project consisted of a 12 % resolving phase and 6 % stacking phase. The components required to make this gel are summarised in Table 2.7. The protein samples were run alongside Mark12 protein ladder first at 80 V for 10 minutes followed by 200 V for 40 minutes. The gel was taken out as soon as the dye reached the bottom of the gel, and stained with Coomassie Blue (10 % (v/v) acetic acid, 40 % (v/v) methanol and 0.1 % (w/v) Coomassie R250) for an hour to allow staining of protein bands. The gel was then destained in destaining solution (10 % (v/v) methanol and 10 % (v/v) acetic acid) overnight with light shaking to remove non-specific binding of the dye and thus background noise. The protein expression level was measured by comparing the size of the protein of interest against Mark12 protein ladder (Invitrogen^{TM}) and the intensity of the band across the soluble fractions.

Table 2.7 Components and the required volumes for making a 6-12% SDS-PAGE gel

Reagent	Volume for 6% Stack Phase	Volume for 12% Resolve Phase
1M Tris pH8.8		2.35 ml
1M Tris pH6.8	0.47 ml	
MilliQ® water	2.48 ml	1.28 ml
10% SDS	37.5 μl	62.5 μl
30 % Acrylamide solution	0.75 ml	2.5 ml
TEMED	3.75 μl	6.25 μl
10% Ammonium persulfate	37.5 μl	62.5 μl

2.9 Bradford's Assay

The approximate total protein concentration from the soluble fraction of the cell lysate was measured via Bradford's assay. This assay involves the use of Coomassie Brilliant Blue G-250 protein dye (Bio-Rad) that binds to basic residues on proteins forming protein-dye complexes which resulting in a change of colour from brown to blue. 1-20 μl of sample was mixed into 0.8 ml MilliQ® water in a plastic cuvette. 0.2 ml of Bradford's reagent was added and mixed carefully. Absorbance at 595 nm, OD_{595}, was measured using a spectrometer. For reliability, a reading of 0.1-0.7 has to be obtained. Calculation of protein concentration was based on the following formula, which had been optimised against Bovine Serum Albumin (BSA) standards:

$$Concentration\ (mg\ ml^{-1}) = \frac{OD_{595}\ x\ 15}{volume\ of\ sample\ (\mu l)}$$

2.10 Protein concentration measurement using UV absorbance at 280 nm

Unfortunately, the sensitivity of Bradford's assay varies on a case-by-case basis. A more reliable technique to measure protien concentration is by measuring the UV absorbance of aromatic residues at 280 nm, and is calculated as below:

$$Concentration\ (M) = \frac{OD_{280}\ x\ lid\ factor}{extinction\ coefficient\ (M^{-1}\ cm^{-1})}$$

The extinction coefficient of each protein was calculated using Protparam (Gasteiger et al., 2003).

2.11 Purification of overexpressed protein

The protein of interest was then overexpressed in a large-scale culture (1-2 l) using the same approach as in Section 2.7, but instead, inoculating 5 ml starter culture into 500 ml growth media in 2 l flasks. The cultures were grown in the condition found best to produce high-level soluble expression from the overexpression trial. In order to purify the protein to >90 % homogeneity, various chromatography techniques were carried out based on the characteristic of the protein constructs, including the Nickel-affinity column purification of His_6-tagged protein, ionic exchange chromatography for untagged proteins and gel filtration column for separating proteins based on their size.

2.11.1 Preparing cell-free extract

It is crucial to isolate soluble cell free extract from other components from the cell membrane and cell wall prior to purification. Cells were lysed by sonication using MSE Soniprep 150, which is a sonicator that converts electrical to mechanical energy in the form of high frequency ultrasonic vibrations via a titanium alloy probe with a 9.5 mm tip diameter. Every 2 g of frozen cell paste from the large-scale overexpression culture was resuspended with 10 ml of buffer. About 20 ml aliquots were transferred into sonication bottles and kept on ice throughout the sonication procedure which comprised 2 cycles of 20 seconds sonication at full power >15 microns with cooling in between. The lysed cells were then decanted into 40 ml centrifuge bottles and centrifuged at 24000 rpm for 10 minutes at 4 ^{o}C in Beckman JA 25-50 rotor. Following centrifugation, the cell free extract was isolated, ready for purification.

2.11.2 Nickel-affinity Chromatography

His_6-tagged proteins were purified using immobilised metal ion affinity chromatography (IMAC) utilising a nickel-affinity column. This technique employs immobilising Ni^{2+} ions

on highly cross-linked agarose matrix for affinity binding with poly-His clusters of the tagged proteins. This binding is reversible by competitive replacement of the poly-His clusters with increasing Imidazole concentration thus allowing selective purification of His_6-tagged protein. Throughout this study, a pre-packed HisTrap^{TM} HP 5 ml and 1ml columns and an AKTA pure purification machine from GE Healthcare were used. For this purification, the buffer used for sonication was 50 mM Tris pH 8.0 and 0.5 M NaCl to minimise non-specific interaction between contaminating proteins and the column matrix. The column was initially equilibrated with Binding Buffer (50 mM Tris pH 8.0, 0.5 M NaCl). Cell free lysates were then applied to the column to allow binding of tagged protein. The column was then washed with the same buffer to get rid of contaminants and unbound materials. Elution of tagged protein was done in a linear gradient of 0-70 % Elution Buffer (50 mM Tris pH 8.0, 0.5 M NaCl, 0.5 M Imidazole) in 10 column volumes (CV) with 3 ml fractionation. Fractions collected were analysed by SDS-PAGE gels to measure purity. Protein concentration was analysed using Bradford's Assay (BioRad®) or UV absorbance at 280nm.

2.11.3 Ion Exchange Chromatography (IEC)

Proteins possess an isoelectric point (pI), a pH value above which the protein would have an overall negative net charge and bind to a positively charged medium (anion exchanger), and below which the protein would have an overall positive net charge and thus bind to a negatively charged medium (cation exchanger). Based on the pI value of the proteins studied in this project, a weak anion exchanger, DEAE Sepharose FF 5 ml pre-packed column was used (GE Healthcare). Proteins will bind to the column at pH higher than the pI and can be eluted by changing ionic strength by varying NaCl concentrations. Cell free extract was obtained by sonicating cell suspension in 50 mM Tris pH 8.0 buffer followed by centrifugation as detailed above. The column was equilibrated with the same buffer followed by the application of cell free extract. The column was then washed with the same

buffer to remove unbound materials. Elution of bound protein was conducted in a linear gradient of 0-100 % Elution Buffer (50 mM Tris pH 8.0, 1 M NaCl) in 10 CV in 2.5 ml fractions. Fractions collected were analysed by SDS-PAGE gels to measure purity. Protein concentration was estimated using Bradford's Assay (BioRad®) or UV absorbance at 280nm.

2.11.4 Size Exclusion Chromatography (SEC) using gel filtration column

SEC, or generally known as gel filtration, is a technique of purifying protein by their molecular sizes. It deploys the use of a column of packed porous matrix of inert spherical particles normally made up of dextran beads. Based on the pore sizes, smaller proteins enter the matrix pores and therefore have longer retention time, whilst larger proteins cannot enter the pores and just pass in between the beads allowing quicker passage as they pass through the column. Superdex 200pg 16/600 Sepharose column was used throughout this study. The gel filtration column was generally used as the second chromatography step, in a 2-step purification after IEC or IMAC. The sample was concentrated to a volume of 2 ml and loaded to the system via a 2 ml loop. Elution was done using the same buffer at flow rate 1.5 ml/min. 2 ml fractions were collected across 1 CV and peaks profile were studied with SDS-PAGE to isolate fraction containing purified protein of interest.

2.12 Vivaspin Concentrator

Throughout the chromatography process, it is crucial to concentrate the protein solution either to reduce to a specific volume for sample loading, to conduct buffer exchange or to achieve a desired concentration for crystallisation trials. Vivaspin centrifugal concentrators (Sartorius®) were used to concentrate protein samples, whereby different molecular weight cut-offs (MWCO) were used according to the size of the protein under study. The commonly used MWCO in this project were the 30,000 Da and 10,000 Da. Protein solution in the selected buffer was transferred to the vivaspin and centrifuged continuously in a table-top centrifuge at 5000 rpm until the preferred volume or concentration was achieved.

2.13 Crystallisation trials and optimisation

Protein samples were typically concentrated to 10 mg ml^{-1} prior to crystallisation. Initial crystallisation trials were carried out via the sitting drop vapour diffusion technique using a 96-well MRC crystallisation plate. Subsequently, optimisation of crystal conditions was done by the hanging drop technique utilising a 24 big-well plate and siliconised cover slips.

2.13.1 Sitting drop vapour diffusion

Hydra II Plus One and TPP Labtech mosquito® LCP crystallisation robots were used in setting up sparse-matrix screening across a range of commercialised crystallisation screens including PACT, PEG, JCSG+, MPD, AmSO4, pH Clear and Proplex, each consisting of 96 different crystallisation conditions. 100 nl of precipitant was mixed with 100-200 nl drop of concentrated protein solution in each well, and when sealed, the mixed droplets would equilibrate against the precipitant, increasing the concentration of the protein to the point of supersaturation at which crystals may form. The plates were incubated at 17 ^{o}C and viewed under the microscope after a set time to observe the formation of crystals.

2.13.2 Hanging drop crystallisation

In order to generate bigger and better diffracting crystals, optimisation of crystallisation conditions was carried out by varying the pH and the concentration of the components. This was done via the hanging drop vapour diffusion technique involving a manual mixing of protein solution and optimised crystallisation reagent on a siliconised cover slip. The cover slip was then sealed on the well using grease or mineral oil in such a way that the droplet will hang above the precipitant in a sealed environment. The use of cover slips allow multiple droplets to be set according to preferred protein:precipitant ratio. The crystallisation plates were then incubated at 17 ^{o}C.

2.14 Data collection and structure determination

Crystals that were worthy of further analysis were harvested, mounted on a loop to facilitate cryo-cooling and subjected to a multi-level decision process of X-ray diffraction analysis.

2.14.1 Crystal harvesting and mounting at cryogenic temperature

Due to the damaging nature of the ionising X-ray photons, crystals are normally mounted at cryogenic temperatures using liquid nitrogen to reduce the effects of radiation damage. To avoid the formation of ice crystals from the water molecules in the mother liquor during cryomounting, crystals were briefly passed through a cryoprotectant solution before dipping into liquid nitrogen. Cryoprotectant solutions were prepared by making up the same components as the crystallisation condition with an addition of 25-30 % (w/v) ethylene glycol (for conditions containing poly-ethylene glycol (PEG) or glycerol). Crystals were kept cold in unipucks (Crystal Positioning Systems) stacked in a shipper prior to shipment to Diamond Light Source, Didcot, UK.

2.14.2 Data collection and processing

Details on the beamline and data collection strategies are highlighted in Paper 2. In general, initial test diffraction was done by taking 5, 0.1^{o} rotation, diffraction images at 45^{o} interval to ensure that the crystals lie in the X-ray beam and to provide the first look at the quality of the diffraction spots at different positions. These test images would then be auto-indexed by MOSFLM (Leslie and Powell, 2007) or EDNA (Incardona et al., 2009) in the ispyb SynchWeb interface (Fisher et al., 2015) to give details on the possible space group and unit cell dimensions, which will be used as a guide to assist the decision on data collection strategy. Following data collection, the integrated data reduction pipeline on ispyb allows high-throughput auto-processing of diffraction images utilising softwares including Fast DP (Winter and McAuley, 2011), XIA2 (Winter, 2010) and autoPROC (Global Phasing Ltd), that would cover the whole data processing steps from integration, scaling and merging. The resulting statistics were then analysed for quality indicators such as R_{merge}, R_{pim}, R_{meas}, signal-to-noise ratio, completeness and multiplicity. Data with poor statistics were then troubleshot for problems during processing by referring to the log files looking at possible regions with poor merging statistics across resolution and, when required, manual reprocessing of the diffraction images was carried out with preferred parameters using XIA2 (Winter, 2010).

2.14.3 Phasing and structure determination

Once the quality indicators were assessed and seen to have decent values, structural determination followed. Over the course of this study, a serendipitous Zn anomalous diffraction was obtained from crystals grown in $ZnSO_4$ and used to solve the crystal structure of the untagged SSR1 monomer. Details on the structure determination of the protein via Single wavelength anomalous dispersion (SAD) are included in Appendix B. For the majority of the other crystals, the molecular replacement technique using PhaserMR (McCoy et al., 2007)

was employed utilising structures of related proteins as the search model. Details on these structure determinations are highlighted in the relevant papers and appendices.

2.14.4 Model building, refinement, validation and analysis

Successful structural determination was followed by an iterative-fashion of model building in real-space by analysing the electron density map using *Coot* (Emsley et al., 2010), and refinement with REFMAC5 (Murshudov et al., 2011), with occasional bias check in between refinement cycles, until the R_{work} and R_{free} values reached convergence. The final model was then validated using MolProbity (Chen et al., 2010), to check for errors corresponding to Ramachandran analysis, bond length, angles, and chirality, favored side-chain rotamer positions, and clashes. Errors flagged by MolProbity were then fixed manually through several rounds of model building and refinement as described above. Figures were made mainly using PyMOL (Schrodinger, LLC), whilst figures including electron density map were either generated by FFT (Read and Schierbeek, 1988) or *Coot* (Emsley et al., 2010). Futher analysis including protein-protein interface and structural superposition were carried out using PDBePISA (Krissinel and Henrick, 2007) and Superpose in the CCP4 GUI (Krissinel and Henrick, 2004), respectively.

2.15 Other methods

Further details of additional techniques deployed during the project (eg. Mass spectrometry) are provided in the relevant papers and appendices. The theory underpinning X-ray crystallography can be found in standard textbooks such as 'Biomolecular crystallography: principles, practice and applications to structural biology' by Bernhard Rupp (Garland Science, Taylor & Francis Group, 2010) and 'Crystallography made crystal clear' by Gale Rhodes (Academic Press, 2006).

Publications

The publications included in this thesis are listed below.

1. Molecular basis of the specificity of the Burkholderia Lethal Factor 1 for its eukaryotic target, eIF4A. ***Prepared for submission.***

    - BLF1 was characterised as a GDT, deamidating Gln339 of eIF4A. BLF1C94S mutant was successfully co-crystallised with its substrate, eIF4A, to shed light on the molecular basis of specificity and mechanism of BLF1 fulfilling aims 3 and 4 of this study.

2. Identification of a glutamine deamidase toxin in *Skermanella stiibiresistens*. ***Prepared for submission.***

    - SSR1 was identified as another member of the CNF1-like GDTs and successfully characterised as having deamidase activity, followed by production of crystals fulfilling aims 1, 2 and 3.

3. Strain-induced clipping: a novel form of post-translational modification. ***Prepared for submission.***

    - Different crystal structures of SSR1 contribute towards aims 2 and 3 of this study, but also provide unprecedented insights on protein folding and evolution.

Chapter 3

Paper 1

3.1 Summary

This manuscript describes the structure determination of BLF1 C94S in complex with eIF4A in two crystal forms. Analysis of the structures indicates that BLF1 binds in the cleft between the two domains of eIF4A forming interactions consisting of two patches, one on each domain of eIF4A, involving extensive van der Waals forces, hydrogen bonds and salt bridges. Interestingly, about 75 % of these residues in eIF4A are conserved across a range of eukaryotes. In addition, the two domains of eIF4A are held in a novel conformation, placing the target Gln339 residue of eIF4A in the active site cleft of BLF1, where a thiolate-imidazolium ion pair was proposed to carry out its function. The utilisation of the thiolate-imidazolium ion pair combined with an oxyanion hole led to a proposal that glutamine deamidation by BLF1 shares a similar mechanism to that of papain.

3.2 Author contributions

D.W.R., M.J.D. conceived the investigations; G.W.M., A.A.A., T.C.M. designed the experiments; G.W.M., A.A.A., S.E.S., and T.C.M. conducted the experiments; G.W.M., A.A.A., S.E.S., T.C.M., M.J.D., G.M.B., and D.W.R. interpreted the data and G.W.M., A.A.A., P.J.B, and D.W.R wrote the manuscript. Experimentally, I carried out the crystallisation, data collection and structure determination of BLF1 C94S:eIF4A complex of form B. The coordinates of crystal form A was given by G.W.M. which was further refined and validated by me. I also carried out the crystallisation, data collection and structure determination of the new BLF1 WT dataset for the analysis of the oxidation state of the active site cysteine residue. I also analysed the interactions between BLF1 C94S and eIF4A by PDBePISA. The figures were made as a shared effort between G.W.M and me.

3.3 Paper 1

Paper 1 manuscript will start on the next page.

Molecular basis of the specificity of the Burkholderia Lethal Factor 1 for its eukaryotic target, eIF4A.

George W. Mobbs[a,b,1], Adli A. Aziz[a,1], G. M. Blackburn[a], Sveta E. Sedelnikova[a], Mark J. Dickman[c], Thomas C. Minshull[c], Patrick J. Baker[a], Sheila Nathan[d], Mohd F. Raih[d], David W. Rice[a,*]

[a]*Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK.*
[b]*Present address:Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA.*
[c]*Department of Chemical and Biological Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.*
[d]*School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia.*

Abstract

Burkholderia pseudomallei is an intracellular gram-negative pathogen of animals and man and the causative agent of Melioidosis, an endemic disease of Southeast Asia and Northern Australia. Burkholderia lethal factor 1 (BLF1) is the first lethal toxin from this organism to be characterized and is a member of the CNF1-like family of toxins exhibiting site-specific glutamine deamidase activity against the eukaryotic initiation factor 4A (eIF4A). Here, we report the structure of an inactive C94S mutant of BLF1, in complex with human eIF4A, to provide the first example of a complex between a CNF1-like toxin bound to its substrate thereby revealing the molecular basis of toxin specificity. In the structure of the complex, BLF1 binds in the cleft between the two RecA-like domains of eIF4A, forming largely polar interactions with residues from both and holding them in a novel twisted conformation that is further open than those previously observed. The reorientation of eIF4A leaves the target glutamine Gln339 accessible to the

active site of the toxin, placing the sidechain amide and its scissile C-N bond in an ideal arrangement for nucleophilic attack by the thiolate side chain of Cys94 as part of a mechanism that closely resembles that of the papain-like family of cysteine proteases.

Keywords: BLF1, eIF4A, glutamine deamidation, cysteine protease, Cys-His dyad, thiolate-imidazolium ion pair, papain, oxyanion hole, toxin specificity.

Highlights

- The structure has been solved of a complex between an inactive mutant of the glutamine deamidase toxin, BLF1, a member of the CNF1-like family of toxins, and its molecular target, eIF4A, to provide the first insights into the specificity of any CNF-1 like toxin for its substrate.
- BLF1 binds between the two domains of eIF4A holding them in a novel open conformation facilitating recognition of the target glutamine, Gln339, which lies close to the BLF1 catalytic thiolate-imidazolium pair.
- The specificity of BLF1 for eIF4A arises through complementary recognition by the toxin of largely polar binding surfaces on both domains of the target.

---

*Correspondence: d.rice@sheffield.ac.uk (D.W.R)
[1]These authors contributed equally to the work.

1. Introduction

B. pseudomallei is an intracellular gram-negative bacterial pathogen of man and animals (Ray et al., 2009) and the causative agent of the disease Melioidosis (Wiersinga et al., 2006). This pathogen can be found in moist soil and surface water (Nandi et al., 2010) in Southeast Asia and Northern Australia where the disease is endemic (Dance, 2000) with, for example, 80% of 4-year-old children in Northern Thailand testing serologically positive for exposure to *B. pseudomallei* (Kanaphun et al., 1993). More recently, cases of Melioidosis have been reported in China, India, Africa (Jabbar & Currie, 2013) and Bangladesh (Jilani et al., 2016). Melioidosis presents with a broad array of symptoms and is therefore commonly misdiagnosed as Tuberculosis (Vidyalakshmi et al., 2008), Typhoid fever or Malaria (Currie, 2008). A particular feature of *B. pseudomallei* is that following infection, the organism is capable of lying dormant such that patients appear asymptomatic but, subsequently, and for reasons that are still poorly understood, a full-blown infection can emerge decades later (Kronmann et al., 2009). The emergence of latent infections in veterans who fought in the Vietnam conflict has earned the disease the chilling nickname, The Vietnam Time Bomb (Ngauy et al., 2005). The ability of *B. pseudomallei* to survive harsh conditions and to persist in the environment (Aldhous, 2005), combined with its capacity for latency and to cause a devastating disease has led to this bacteria being classified as a category B bio-warfare agent (Rotz et al., 2002).

The structure determination of the 23 kDa protein of unknown function encoded by the *B. pseudomallei* gene, bpsl1549, revealed that its fold was based on a single domain comprising a sandwich of two mixed β-sheets flanked by α-helices that was similar to the fold, but not the sequence, of the C-terminal

catalytic domain of the Cytotoxic Necrotizing Factor 1 (C-CNF1). The latter is a glutamine deamidase toxin from Uropathogenic *E. coli* (Buetow et al., 2001) which deamidates Gln63 or its equivalent in the small GTPases RhoA, Rac and cdc42 and disrupts actin cytoskeleton assembly (Flatau et al., 1997). The identification of this structural similarity, which included the spatial conservation in the Burkholderia protein of two catalytic residues present in CNF1, Cys866 and His881, led to biochemical studies that showed that the role of the former in causing disease was to act as a toxin targeting the site specific deamidation of Gln339 of the eukaryotic initiation factor 4A (eIF4A) (Cruz-Migoni et al., 2011). eIF4A is a member of the DEAD box superfamily and is an essential component in the translation initiation complex, supplying RNA helicase activity responsible for melting the secondary structures present in mRNA prior to translation (Ray et al., 1985; Pestova et al., 1996). DEAD box family members invariably contain two domains, an N- and a C-terminal RecA-like domain (residues 21-236 and 247-406 respectively in human eIF4A), connected through a flexible linker region (Linder & Jankowsky, 2011) which possesses an inherent degree of conformational plasticity demonstrated by differences in the orientation of RecA-like domains relative to one another as they cycle between closed (active) and open (inactive) states regulated by the binding of ATP (Sengoku et al., 2006). Gln339 is a residue located in the C-terminal domain of eIF4A, between the conserved sequence motifs V and VI, a region believed to interface with RNA and ATP (Caruthers et al., 2000). Site-specific deamidation of eIF4A Gln339 results in extensive inhibition of protein synthesis in human cells (Cruz-Migoni et al., 2011). Further studies demonstrated that BLF1 is a key component for *B. pseudomallei* pathogenesis, with a ΔBPSL1549 strain being 100-fold less virulent than the wild type (Cruz-Migoni et al., 2011). As the first toxin to be

identified in Burkholderia, and only the second member of the CNF1-like family of glutamine deamidase toxins to be discovered, the product of the bpsl1549 gene was named as the Burkholderia Lethal Factor 1, BLF1.

Studies by others have previously identified further toxins which exert their biological activity by glutamine deamidation but which differ from CNF1 in their fold being more reminiscent of the cysteine protease, papain. These toxins are therefore classified as belonging to the papain-like family of glutamine deamidase toxins and include molecules such as the cycle inhibiting factors (Cif) from various pathogenic bacteria like enteropathogenic *E. coli*, *B. pseudomallei* and *Y. pseudotuberculosis*, a Type III effector from *S. flexineri*; Osp1, and a *P. multoxida* toxin, PMT (Washington et al., 2013). Interestingly, whilst both the CNF1-like and papain-like toxins exert their effect by the deamidation of a specific glutamine, recent studies have identified a Type VI effector from the opportunistic pathogen *B. cenocepacia*, TecA, which is also predicted to have a papain-like fold but which functions as an asparagine deamidase (Aubert et al., 2016).

The structures of the complexes of several papain-like family toxins with their substrates have been solved including the Cif-like homologues from *B. pseudomallei* (CHBP), *Y. pseudotuberculosis* and *P. luminiscence* in complex with NEDD8 (Washington et al., 2013). However, there are no structures of a complex between any CNF1-like family of toxins with their targets. In this paper, we present the structure of an inactive C94S mutant of BLF1 in complex with human eIF4A, which, for the first time, allows the mechanism and specificity of deamidation of a CNF1-like glutamine deamidase toxin to be examined.

2. Experimental Procedures

2.1. Protein expression and purification

An inactive mutant of BLF1, C94S, was expressed in *Escherichia coli* BL21 (DE3) cells (Novagen) grown in LB broth supplemented with 100 mg l^{-1} Ampicillin, with expression induced by the addition of 1 mM IPTG at an OD_{600} of 0.6 before incubation at 37 ^{o}C for 4 hours. A construct of human eIF4A missing 20 residues at the N-terminal end, eIF4A$^{\Delta 20}$, was expressed in *Escherichia coli* Tuner (DE3) cells (Novagen) in LB broth supplemented with 50 mg l^{-1} Kanamycin, with the cells grown to an OD_{600} of 1 then chilled to 4 ^{o}C for 30 minutes on ice before induction with 1 mM IPTG and incubation at 13 ^{o}C for 72 hours. Prior to purification, the *E. coli* cells (2 g dry weight) were lysed via three rounds of sonication at 15 microns for 20 seconds, whilst stored on ice in 20 ml of a 50 mM TRIS pH8 buffer and the lysate was clarified by centrifugation at 60,000 x g for 20 minutes. BLF1 was purified through three chromatography stages: (1) the clarified lysate was directly loaded onto a 5 ml DEAE FF column (GE) and eluted with a 50 mM TRIS buffer at pH 8 using a linear gradient of 0.5 M NaCl from 0-100% across 20 column volumes, (2) fractions containing BLF1 were diluted with 50 mM TRIS pH 8 and loaded onto a 5 ml Q HP column (GE) and eluted in an identical fashion, (3) samples were concentrated to a volume of 2 ml and loaded onto a Superdex 200 16/60 size exclusion column (GE). eIF4A was purified either using an equivalent protocol or by affinity chromatography using a nickel column Ni-HP 5 ml (GE) eluted with a 50 mM TRIS buffer at pH8 with 0.5 M NaCl and a linear gradient of 0.5 M Imidazole. To form a complex, a 4:1 molar ratio of BLF1 C94S to eIF4A$^{\Delta 20}$ was mixed in a 50 mM TRIS buffer at pH 7.5. The BLF1 C94S:eIF4A$^{\Delta 20}$ complex was separated from excess BLF1 by gel filtration using a Superdex 200 16/60 size exclusion column run

in a 50 mM TRIS pH 7.5, 0.1 M NaCl buffer. The fractions containing eIF4A and BLF1 run with an apparent molecular weight of ~70 kDa consistent with the formation of a stoichiometric 1:1 complex. The BLF1 C94S:eIF4A$^{\Delta 20}$ complex was then buffer exchanged into a 10 mM TRIS pH 7.5 buffer and concentrated to 12 mg ml^{-1} for crystallization trials.

2.2. Crystallization and structure determination

Two distinct crystal forms of co-crystals of BLF1 C94S:eIF4A$^{\Delta 20}$ were grown at 17 ^{o}C in a sitting drop trials containing 200 nl of a 0.1 M MES pH 6.5, 50 mM $MgCl_2$, 10 % (w/v) 2-propanol, 5 % (w/v) PEG 4000 solution and 500 nl purified complex at 12 mg ml^{-1} (form A); or 200 nl of a 0.1 M HEPES pH 7.5, 4 % (w/v) PEG 6000 solution mixed with 200 nl purified complex at 12 mg ml^{-1} (form B). Crystals were flash-cooled with liquid nitrogen in mother liquor to which 30 % (v/v) ethylene-glycol had been added as a cryoprotectant. X-ray diffraction data were collected at the Diamond Light Source (UK) on beamline I04 (form A) or I03 (form B) to a resolution of 2.5 Å and 2.24 Å, respectively. The data were handled using the CCP4 suite (Winn et al., 2011) and processed by Xia2 (Winter et al., 2013), XDS and AIMLESS (Kabsch, 2010; Evans & Murshudov, 2013). Although the diffraction of crystal form B extended to 2.24 Å, the observation of anisotropy in the data led to a decision to limit the resolution to 3 Å. The two crystal forms were solved independently but using a similar approach. For crystals of form A, initial phases were calculated following molecular replacement in PHASER (McCoy et al., 2007) using coordinates for the C94S mutant of BLF1 (3TUA), the N-terminal domain of eIf4A (PDB: 2ZU6 residues 63-231) and the C-terminal domain of eIf4A (PDB: 2ZU6 residues 240-389) as three separate search ensembles. For crystals of form B, the coordinates for the N-terminal do-

main of eIF4A from PDB entry 2G9N (20-238) replaced those from 2ZU6 in an equivalent molecular replacement search. Model building was carried out in an iterative fashion using *Coot* (Emsley & Cowtan, 2004) and the structures were refined using PHENIX (Adams et al., 2010) (form A) or REFMAC (Murshudov et al., 1997) (form B). Crystals of WT BLF1 isomorphous to those grown as previously described were cryoprotected and data collected to 0.99 Å on beamline I03 at the Diamond Light Source. The structure was refined using REFMAC. Details of the structure solution and refinement are provided in Table 1.

2.3. Mass spectrometry analysis

LC ESI MS analysis was performed as previously described (Cruz-Migoni et al., 2011). In brief, following in gel tryptic digestion peptides were analysed using LC ESI MS analysis using a maXis UHR TOF mass spectrometer (Bruker Daltonics) using an automated acquisition approach. MS and MS/MS scans (m/z 50-3000) were acquired in positive ion mode. Lock mass calibration was performed using HP 1221.990364. Line spectra data was then processed into peak list by Data analysis (Bruker Daltonics) and the sum peak finder algorithm was used for peak detection using a signal to noise (S/N) ratio of 10, a relative to base peak intensity of 0.1 % and an absolute intensity threshold of 100. Extracted ion chromatograms were generated for both the non deamidated and deamidated peptide in Data analysis. MS/MS spectra of the peptides identified were manually verified.

Data collection			
Protein	BLF1 C94S : eIF4A $^{\Delta NTD}$ (PDB:XXXX)	BLF1 C94S : eIF4A $^{\Delta NTD}$ (PDB:XXXX)	BLF1 WT (PDB:XXXX)
Beamline	Diamond light source (UK) I04	Diamond light source (UK) I03	Diamond light source (UK) I03
Wavelength (Å)	0.9795	0.9763	0.9763
Space group	C2	P63	P212121
a, *b*, *c* (Å)	135.9, 50.4, 95.7	199.53, 199.53, 51.34	37.25, 44.58, 116.87
α, β, γ (°)	90.0, 111.9, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
	Native	*Native*	*Native*
Resolution (Å)	63.1 – 2.5	65.3 – 3.0	44.6 – 1.0
R_{pim} (I)[a]	0.096 (0.441)	0.021 (0.109)	0.044 (0.246)
$\langle I \rangle$ / $\langle \sigma I \rangle$[a]	8.1 (1.7)	34.9 (9.0)	12.6 (1.6)
Completeness (%)[a]	99.2 (98.7)	99.7 (97.9)	86.7 (18.6)
No. observations	76,647	198,942	548,111
No. unique reflections [a]	20,378 (1457)	23,808 (1702)	93, 887 (1556)
Redundancy [a]	3.8 (3.6)	8.4 (8.4)	5.8 (1.3)
Refinement			
R_{work} / R_{free}	0.20 / 0.28	0.18 / 0.23	0.14 / 0.16
No. of non-H atoms	4,735	4716	2090
Protein / Water	4,638 / 97	4716 / 0	1638 / 442
Average B-factors (Å^2) (Protein / Water)	46.0 / 44.3	36.3 / -	11.2 / 22.9
Bond length rmsd (Å)	0.011	0.011	0.011
Bond angle rmsd (°)	1.67	1.55	1.62
Validation			
Ramachandran favored (%)	95.4	94.4	98.1
Additionally allowed (%)	4.6	5.1	1.2
Outliers (%)	0.0	0.5	0.0
Rotamer outliers (%)[b]	2.0	4.5	0.6
Cb outliers (%)[b]	0.0	0.2	0.0
Clash score [b]	5.78 (99th)	7.35 (98th)	2.79 (91st)
MolProbity score [b]	1.87 (97th)	2.28 (98th)	1.07 (96th)

[a] Highest resolution shell shown in parentheses.

[b] Determined by MolProbity (Chen *et al.,* 2010), with percentile scores ranked against structures deposited in the PDB within a ± 0.25 Å range, see parentheses.

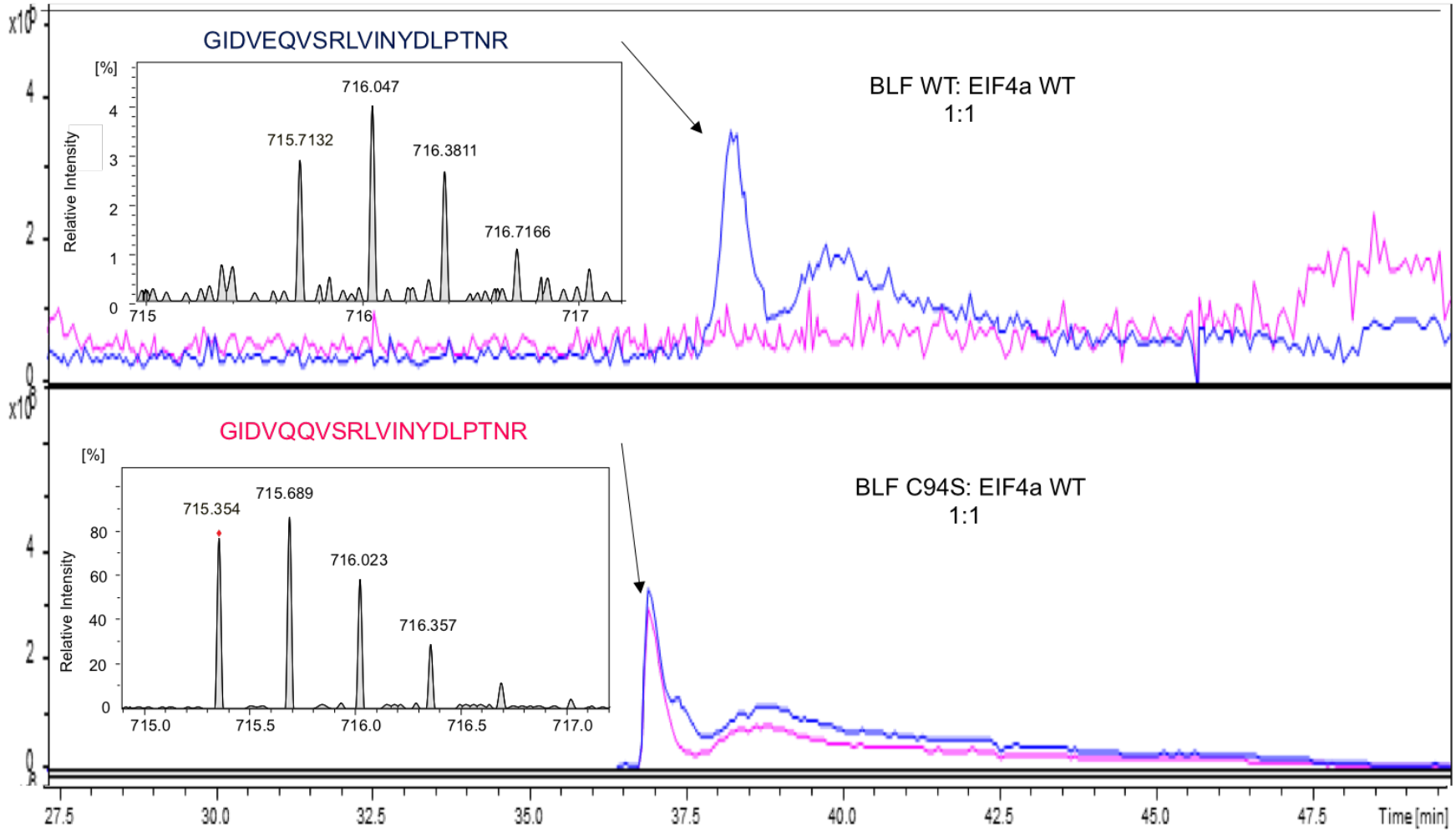
Table 1: **Data collection and refinement statistics.**

3. Results And Discussion

3.1. The structure of the inactive BLF1 C94S:eIF4A$^{\Delta 20}$ complex.

The complex between BLF1 C94S and eIF4A$^{\Delta 20}$ was purified by gel filtration prior to SDS-PAGE analysis. Subsequent in-gel tryptic digestion of eIF4A in conjunction with mass spectrometry analysis confirmed that the C94S mutant was inactive with no deamidation of Glutamine 339 observed in contrast to treatment with the WT toxin (see Figure S1). The structure of the catalytically inactive BLF1 C94S mutant in complex with human eIF4A$^{\Delta 20}$, in two distinct crystal forms, A and B, was solved and refined to 2.5 Å and 3.0 Å, respectively. The structures of crystal forms A and B show that they are very closely related with BLF1 C94S binding between and forming interactions with each of the N- and C-terminal RecA-like domains of eIF4A (Figure 1A). Superposition of the coordinates of these two complexes based on those of BLF1 reveals that the position of the C-terminal domain of eIF4A with respect to BLF1 is very similar with a maximal displacement of equivalent alpha carbon atoms of approximately 1.7 Å. The major difference lies in the position of the N-terminal domain which can be seen to have rotated by an angle of approximately 4^{o} against the BLF1 surface with a maximal displacement of 3.8 Å (Figure S2A). Whilst the conformation of the main chain of BLF1 C94S in both forms A and B is essentially the same, differences can be identified between the two structures including the ordering of residues 135-150 in the N-terminal domain of eIF4A and residues 331-335 in its C-terminal domain, which are disordered in form A but which form a loop and an alpha helix connecting β4 and β5 and a loop connecting β12 and β13 in form B, respectively. Further changes in the conformation of residues of eIF4A can also be observed in the loop between α-helices α1 and α2 (residues 50-55) and residues connecting β2 with α4 (77-86) in the N-terminal domain.

A


GIDVEQVSRLVINYDLPTNR
[%]
716.047
715.7132
716.3811
716.7166
Relative Intensity
715
716
717
BLF WT: EIF4a WT
1:1
GIDVQQVSRLVINYDLPTNR
[%]
715.354
715.689
716.023
716.357
Relative Intensity
715.0
715.5
716.0
716.5
717.0
BLF C94S: EIF4a WT
1:1
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
47.5
Time [min]

B

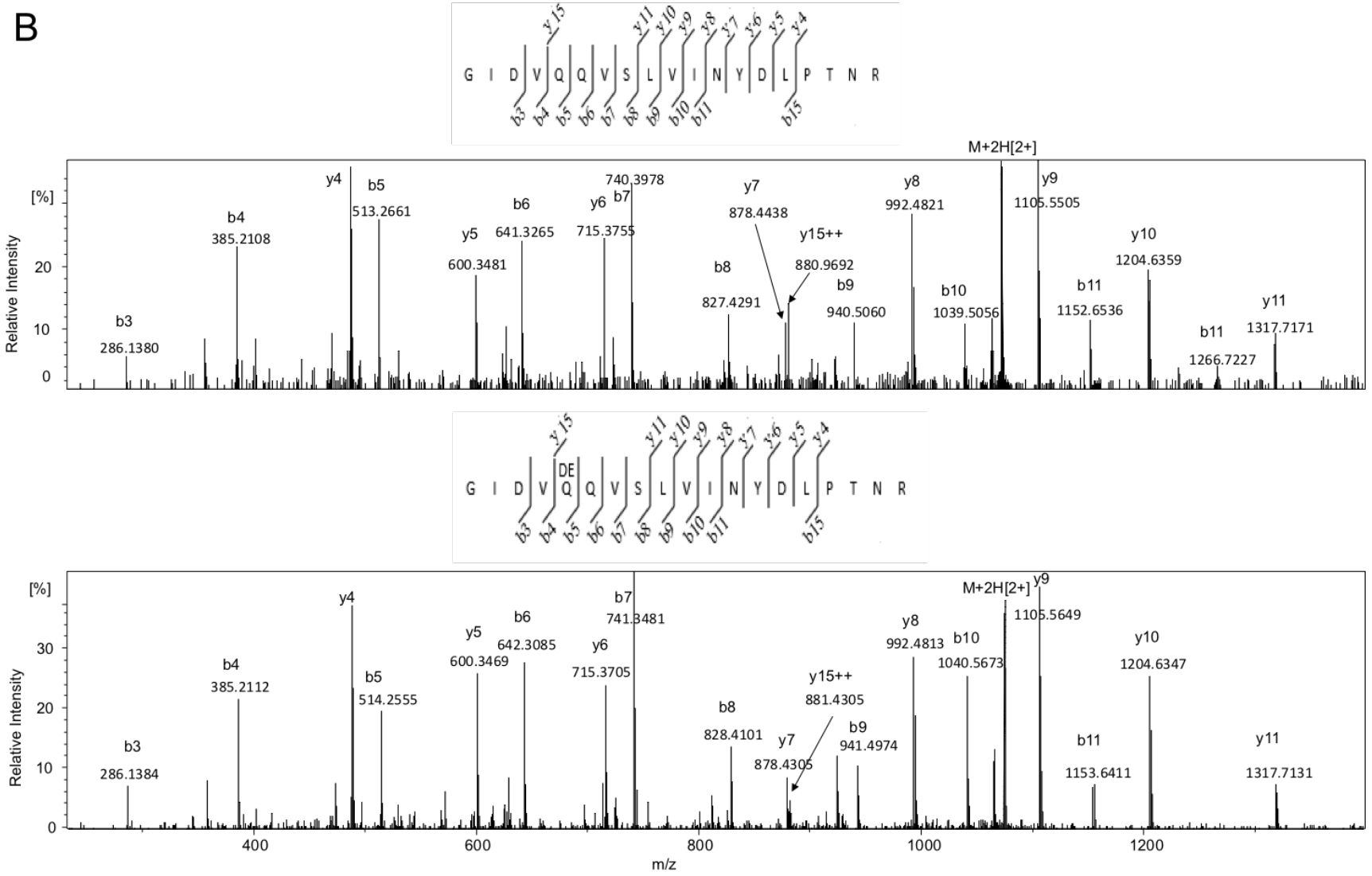

G I D V Q Q V S L V I N Y D L P T N R
b3 b4 b5 b6 b7 b8 b9 b10 b11 b15
y15 y11 y10 y9 y8 y7 y6 y5 y4
[%]
Relative Intensity
M+2H[2+]
y4
b5
513.2661
b4
385.2108
b6
641.3265
y6
715.3755
b7
740.3978
y5
600.3481
y7
878.4438
y8
992.4821
y9
1105.5505
y15++
880.9692
b8
827.4291
b9
940.5060
b10
1039.5056
y10
1204.6359
b11
1152.6536
y11
1317.7171
b3
286.1380
b11
1266.7227
G I D V DE Q Q V S L V I N Y D L P T N R
b3 b4 b5 b6 b7 b8 b9 b10 b11 b15
y15 y11 y10 y9 y8 y7 y6 y5 y4
[%]
Relative Intensity
y4
b7
741.3481
b6
642.3085
y5
600.3469
y6
715.3705
M+2H[2+]
y9
1105.5649
y8
992.4813
b10
1040.5673
y10
1204.6347
b4
385.2112
b5
514.2555
y15++
881.4305
b8
828.4101
y7
878.4305
b9
941.4974
b11
1153.6411
y11
1317.7131
b3
286.1384
m/z

Figure S1: **LC MS analysis of the Gln339 containing tryptic peptide from eIF4A.** A) Extracted ion chromatogram of the tryptic peptide GIDVEQVSLVINYDLPTNR $[M+3H]^{3+}$ from BLF1+eIF4A (top) and the extracted ion chromatogram of the peptide GIDVQQVSLVINYDLPTNR $[M+3H]^{3+}$ from BLF1 C94S+eIF4A, no significant deamidation of eIF4A was observed (bottom). MS1 spectra are shown inset. B) MS/MS analysis of the deamidated peptide GIDVQQVSLVINYDLPTNR $[M+3H]^{3+}$ from BLF1+eIF4A (top) and MS/MS analysis of GIDVEQVSLVINYDLPTNR $[M+3H]^{3+}$ from BLF1 C94S+eIF4A (bottom). The prominent b and y ions are highlighted confirming the deamidated and non-deamidated peptides respectively.

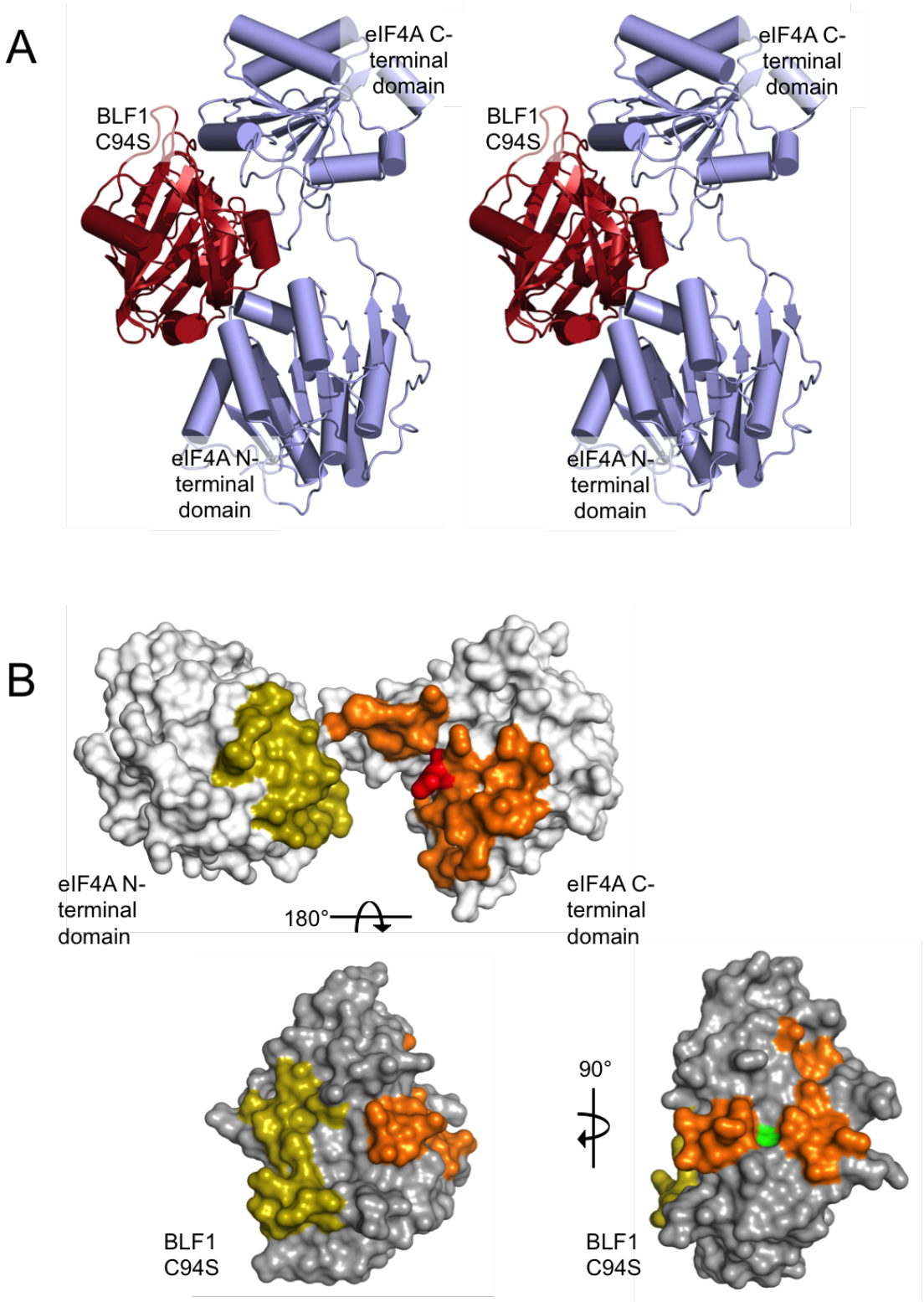


Figure 1: **BLF1 C94S binds in the cleft between the two domains of eIF4A.** A) Stereoview of the structure of the BLF1 C94S:eIF4A complex. B) Surface representation of the interacting patches between BLF1 C94S and eIF4A. Olive and orange patches on eIF4A (white) interact with the corresponding surface on BLF1 C94S (grey). Gln339 of eIF4A is colored red and the corresponding Cys/His pair in the active site cleft of BLF1 is colored green.

The comparison of forms A and B reveals a significant difference in the position of the target Gln339 in eIF4A with respect to the Cys/Ser-His pair in the BLF1 active site. Whilst in form A the orientation of Gln339 is consistent with the suggested involvement of these residues in the mechanism and the expected nucleophilic attack on the side-chain carbonyl group by Cys94 (see below), in form B it is not. This arises as a result of the binding of an unmodelled buffer component in form B close to the carboxyamide of Gln339. This leads to a shift in the position of the carboxyamide group of Gln339 away from Cys94 by approximately 1 Å which presumably triggers the slight opening in the domain orientation of eIF4A in form B compared to form A (Figure S2A). We note that a similar situation has previously been observed comparing the structures of the complex between the papain-like Cif glutamine deamidase from *Photorhabdus luminescence* (PDB: 4FBJ) and that of *Yersinia pseudotuberculosis* (PDB: 4F8C) with their substrate Nedd8, where the sidechain of the target Gln40 in the former adopts a position that is inappropriate for catalysis (Crow et al., 2012). Hence, we suggest that the orientation of Gln339 in form B represents a non-productive mode of binding. Since the relative position of the glutamine side chain to the active site in form A appears to represent one that is closely related to the expected catalytic conformation the following description will therefore focus on the interaction between BLF1 C94S and eIF4A seen in form A.

3.2. The structure of BLF1 C94S in complex with eIF4A closely resembles that of BLF1 on its own.

Analysis of the structure of the BLF1 C94S mutant in crystal form A in the complex with eIF4A compared to that of the mutant on its own (PDB: 3TUA) shows that its conformation is not significantly altered by complex formation

(r.m.s.d of 0.47 Å over 211 equivalent alpha carbon atoms in BLF1). The only differences are small shifts in the loops between β2-β3 (residues 41-45) and β13-β14 (residues 192-195), which have their origins in interactions between BLF1 C94S and the residues at its interface with eIF4A (Figure S2B). Comparison of the unbound wild-type (PDB: 3TU8) and the C94S mutant (PDB: 3TUA) of BLF1 shows that there is a difference in the torsion angle between the serine and the cysteine side chains at position 94, which is also observed in the structure of the C94S mutant in complex with eIF4A.

3.3. BLF1 binds between and forms interactions with both RecA-like domains of eIF4A, which are held in a novel, twisted conformation.

Comparison of the structure of eIF4A in the complex with BLF1 C94S with other eIF4A structures in the PDB reveals a significant change to the relative orientation of its N- and C-terminal domains. Thus, comparing the structures of the apo yeast eIF4A structure (PDB: 1FUU) (Caruthers et al., 2000) and the active yeast eIF4A:eIF4G complex (PDB: 2VSX) (Schütz et al., 2008), these two domains can be seen to adopt an open conformation in the former and a closed conformation in the latter with the difference corresponding to a relative rotation of the C-terminal domain with respect to the N-terminal domain about the hinge by approximately 50^{o} (Figure 2). Examination of the relative position of these two domains in the structure of the complex between human eIF4A and BLF1 reveals that they adopt a novel conformation not previously observed in which the two domains are not only some 10^{o} more open compared to the yeast structure (PDB:1FUU) but also twisted by 180^{o} such that the two domains face each other quite differently (Figure 2).

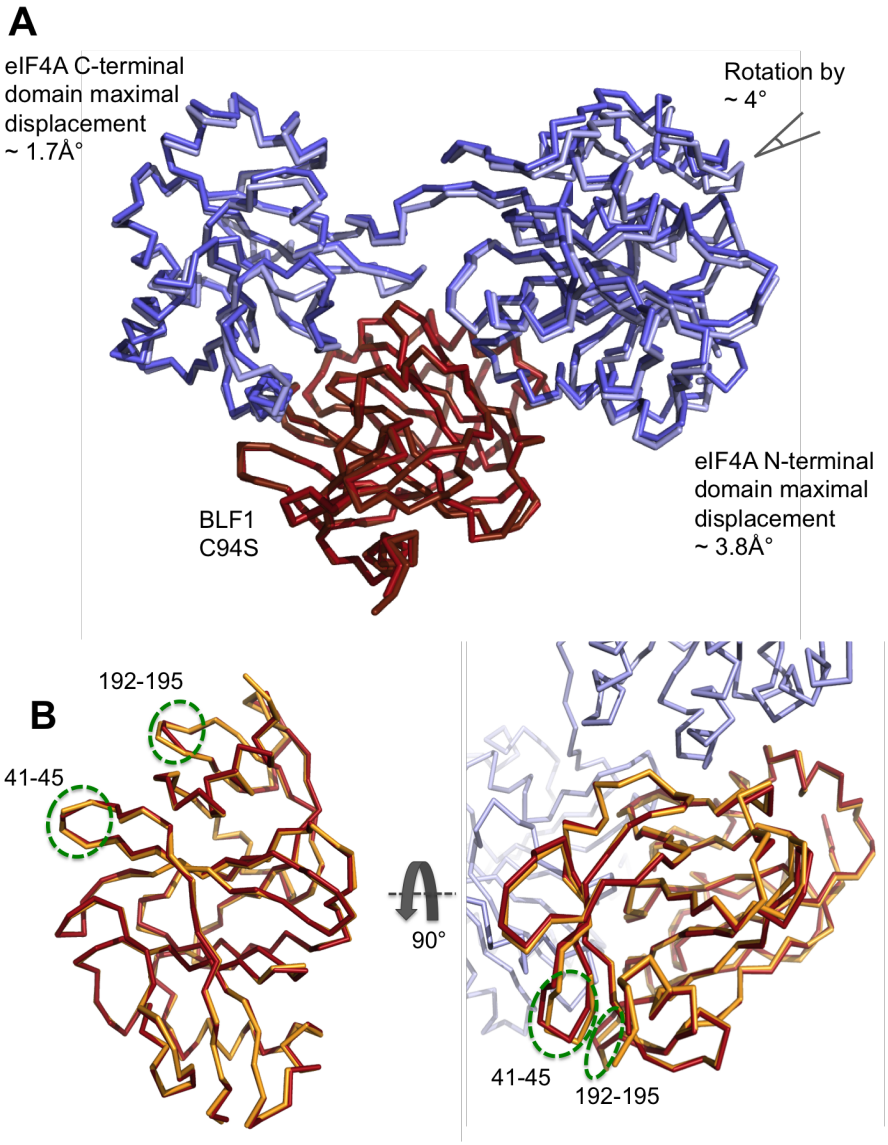


Figure S2: **BLF1C94S undergoes minimal rearrangement upon binding eIF4A.** A) Structural comparison of forms A (red : light blue) and B (chocolate: blue) of BLF1 C94S:eIF4A complex showing the small difference in the position of the two eIF4A domains. B) Superposition of BLF1 C94S of the complex structure of form A on the BLF1 C94S apo (PDB ID: 3TUA) showing very similar conformation, but small shifts in two loops (green dashed circles).

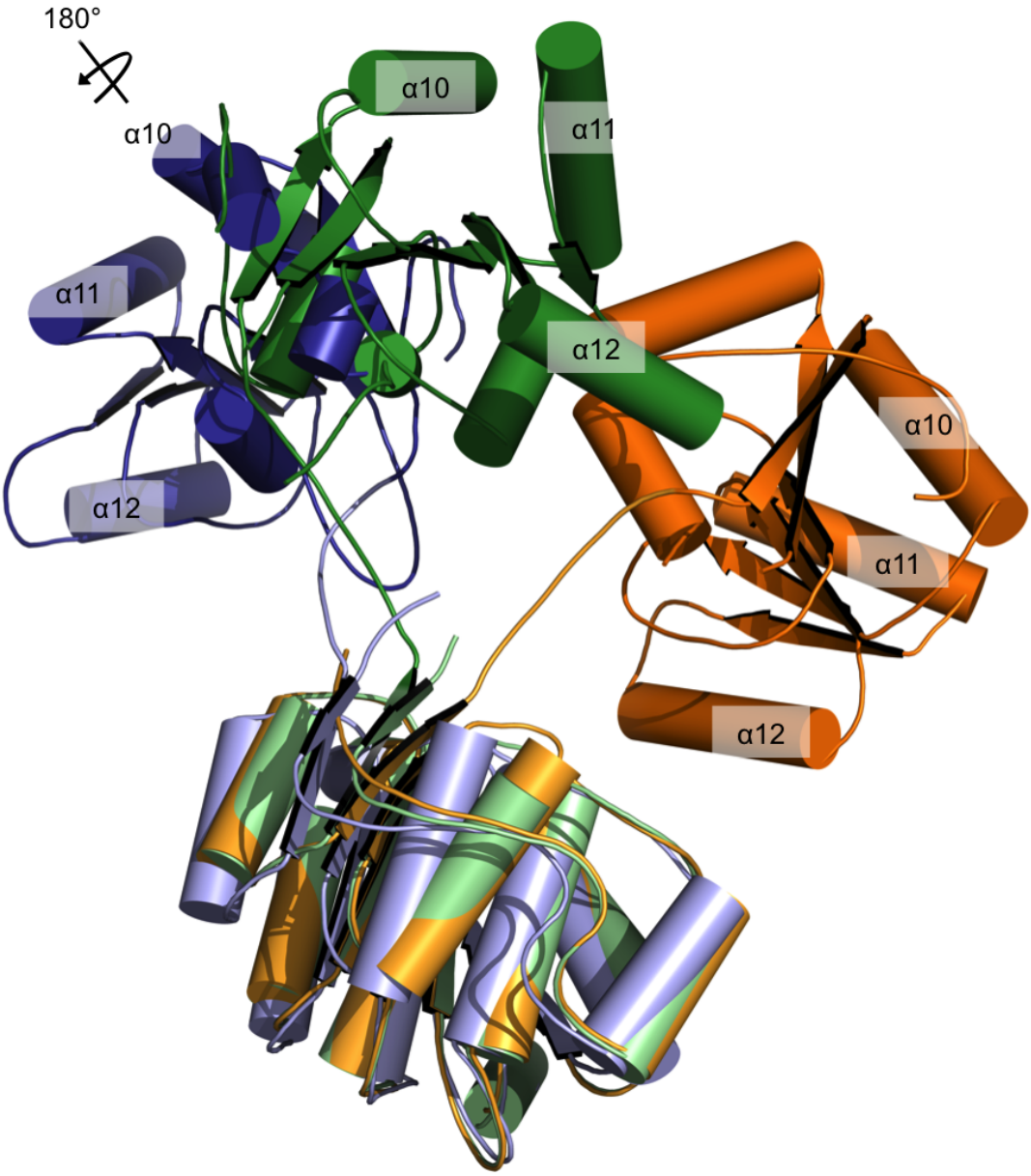


Figure 2: **BLF1 C94S binds to a novel open conformation of the two eIF4A domains.** Superposition of different eIF4A structures on the basis of their N-terminal domains reveals the quite different position of the C-terminal domain. In the apo yeast eIF4A, (Green; PDB: 1FUU) the two domains adopt an open conformation whilst a complex of yeast eIF4A with eIF4G (Orange; PDB: 2VSX) reveals a closed conformation for eIF4A. Human eIF4A in complex with BLF1 (Blue; PDB:XXXX) adopts an even more open and twisted configuration.

3.4. The molecular basis of specificity of BLF1 for eIF4A.

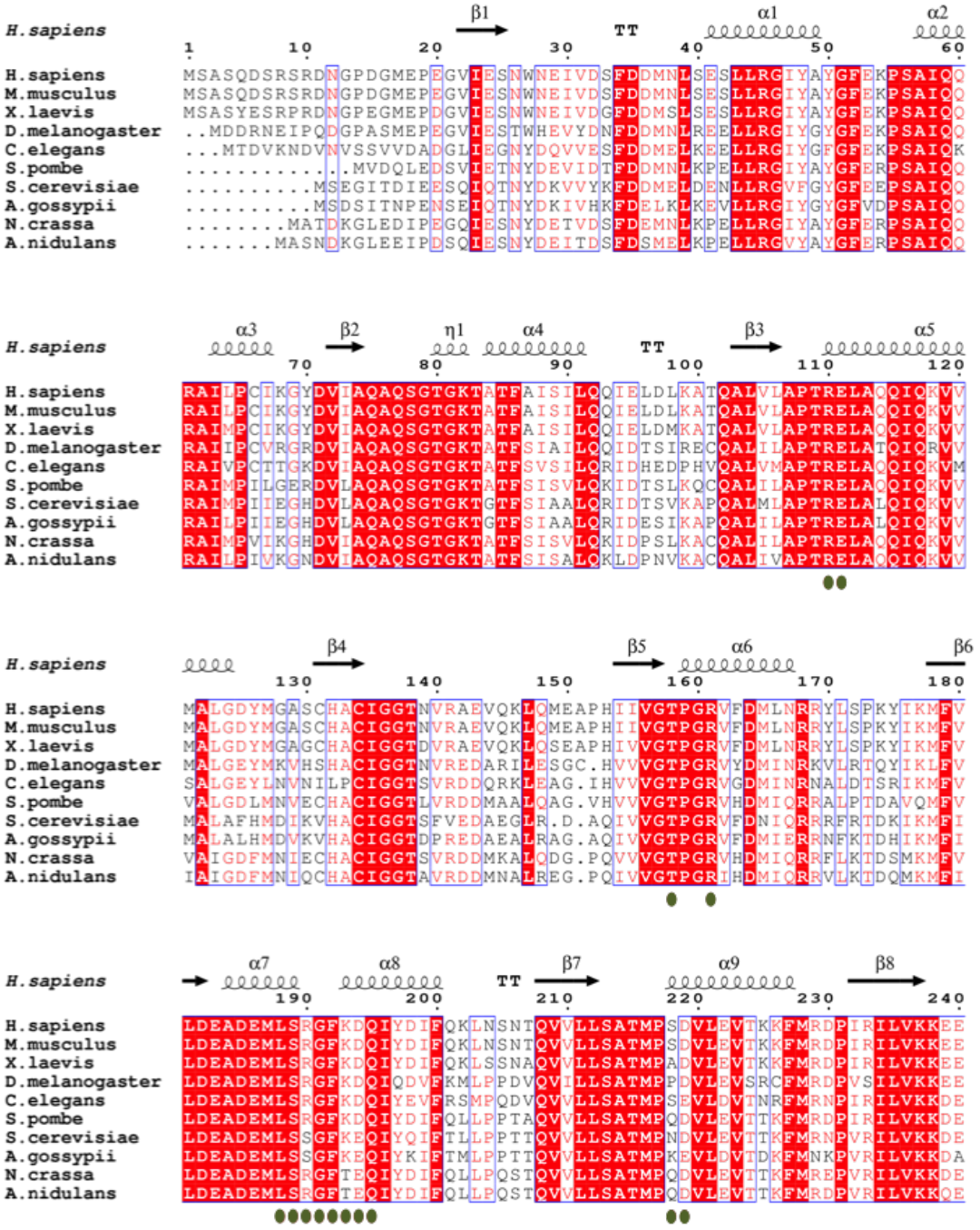
Analysis of the interface formed between BLF1 C94S and eIF4A indicates that a total surface area of 1900 Å^2 (figures rounded to the nearest 100 Å^2) is buried in the formation of the complex. This involves interaction of two separate patches involving 15 and 29 residues on the surface of BLF1 with 14 and 22 residues on the faces of the N- and C-terminal domains of eIF4A, burying 700 Å^2 and 1200 Å^2 surface area, respectively. Interestingly, more than 75 % of the residues in eIf4A buried in the interaction with BLF1 are strongly conserved across a representative set of eIF4A sequences (Table 2, Figure S3 and Figure 1B).

Examination of the residues on the surface of the N-terminal domain of eIF4A that are buried on binding BLF1 shows that they form a complementary network of largely polar interactions involving van der Waals packing, direct and water mediated hydrogen bonds and ion pairs (Table S1 & S2). A key feature of this interface is the interaction of a basic patch on the surface of eIF4A involving Arg110 and Arg161 with an acidic surface of BLF1 composed of Glu124, Glu127, and Asp128 (Figure S4A, Table S2). Other interactions involve Asp194 and Gln195 of eIF4A, forming hydrogen bonds with the backbone amide of Thr34 and the hydroxyl group of Ser78 on BLF1 (Figure S4A, Table S1).

The buried surface between BLF1 and the C-terminal domain of eIF4A again involves the interaction of largely polar surfaces which form a patch surrounding the active site (Table S1 & S2, Figure 1B, Figure S4B). An important feature of this polar interface is the interaction of a number of charged and polar residues (Arg316, Arg319, Arg362 and Arg365 in eIF4A; Glu70, Glu22, Asp108 and Gln171 in BLF1) (Figure S4B, Table S1 & S2). This interface also involves

eIF4A N-terminal domain		eIF4A C-terminal domain	
Residue Number	Surface Area Change (Å^2)	Residue Number	Surface Area Change (Å^2)
Arg110	-90	***Thr273***	-50
Glu111	-40	***Gln274***	-20
Thr158	-20	Lys309	-40
Arg161	-40	Asp312	-30
Leu188	-30	Arg316	-50
Ser189	-80	***Arg319***	-110
Arg190	-90	Ser320	-70
Gly191	-50	***Gly321***	-20
Phe192	-30	***Ser322***	-30
Lys193	-50	***Arg324***	-20
Asp194	-80	***Arg334***	-20
Gln195	-40	***Asp337***	-80
Ser218	-40	***Val338***	-20
Asp219	-50	***Gln339***	-170
Total	-730	***Gln340***	-120
		Arg362	-60
		Gly363	-30
		Gly364	-20
		Arg365	-30
		Phe366	-130
		Gly367	-30
		Lys369	-70
		Total	-1220

Table 2: **eIF4A residues buried by BLF1 C94S upon complex formation.** Surface areas are rounded to the nearest 10 Å^2. Residues in bold italic text are strongly conserved in a multiple sequence alignment of eIF4A (Figure S3).


H.sapiens
H.sapiens
M.musculus
X.laevis
D.melanogaster
C.elegans
S.pombe
S.cerevisiae
A.gossypii
N.crassa
A.nidulans

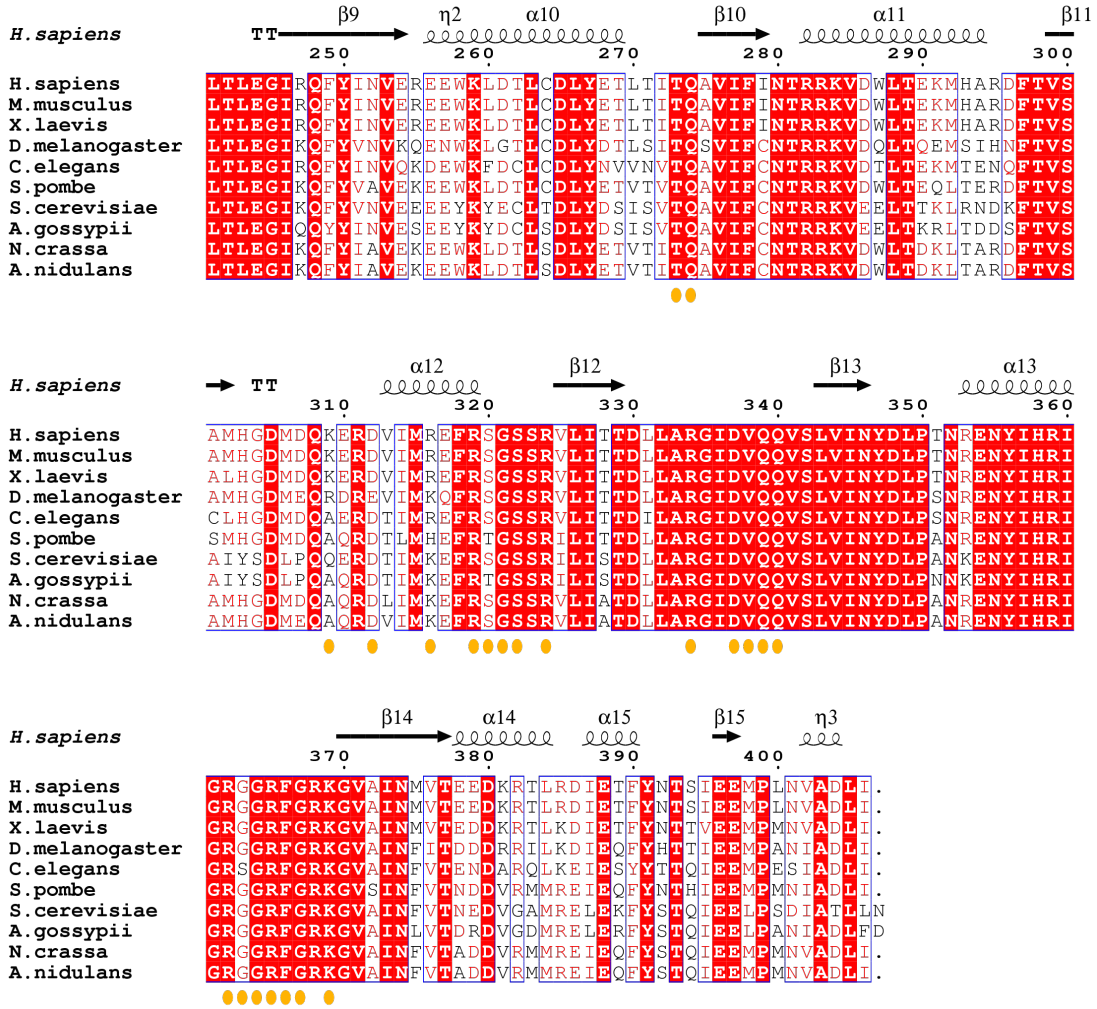
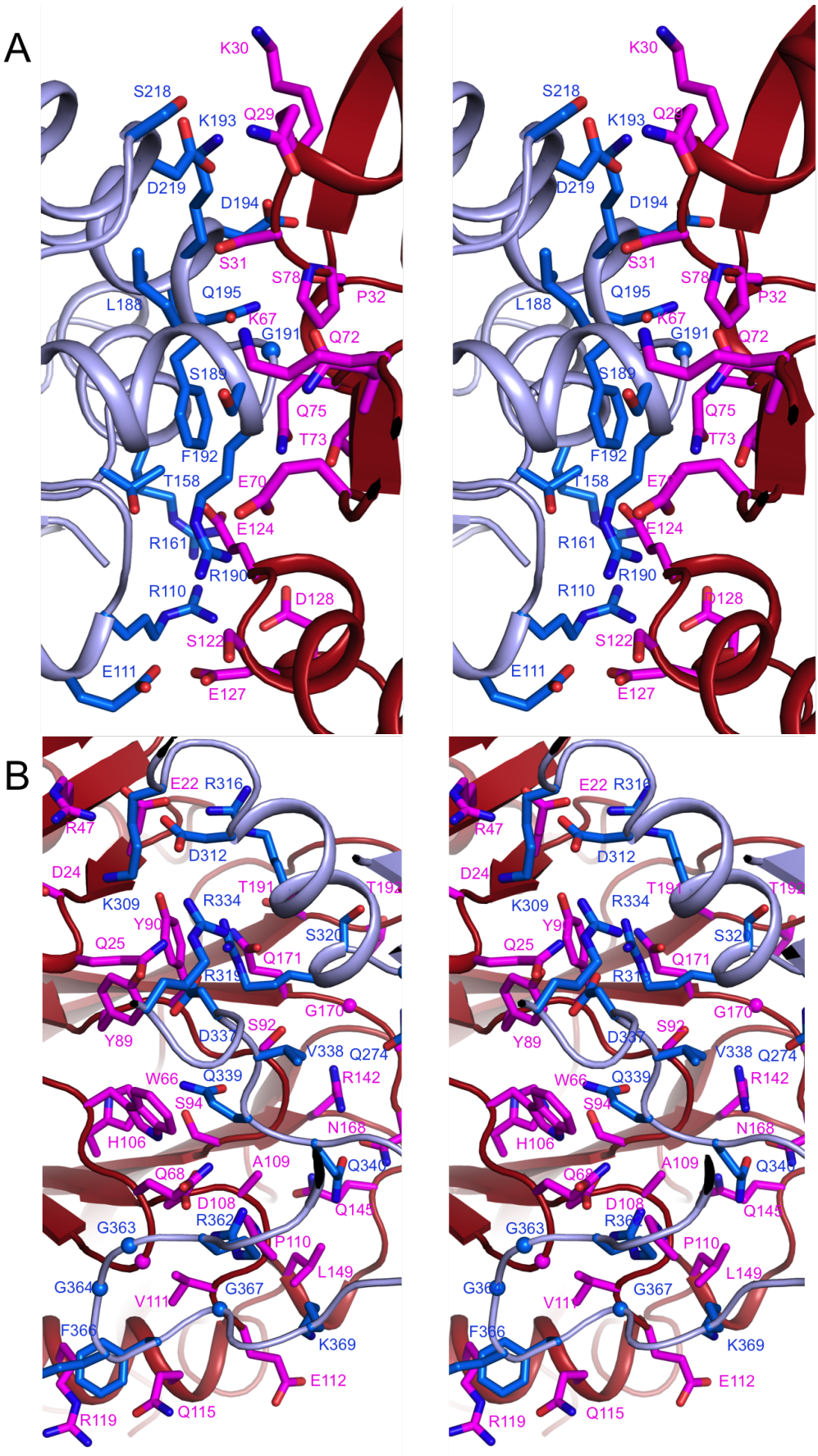


Figure S3: **A multi sequence alignment of eIF4A across a representative range of eukaryotes.** The top line is a cartoon representation of the secondary structures seen in the X-ray structure of human eIF4A. Below are shown a multi-species sequence alignment, that reveals that eIF4A is well conserved from fungi to man. The bottom row denotes exposed residues whose surface areas changed by more than 20 Å^2 on formation of the complex with by BLF1. Olive and orange dots represent residues in the patches on the surface of the N- and C-terminal domains of eIF4A respectively.

257 interactions between BLF1 and the residues in the loop containing the target
258 Gln339 of eIF4A (Figure S4C).

259 We note that the interaction of numerous negatively charged residues in BLF1
260 (Glu22, Glu70, Asp108, Glu124, Glu127, and Asp128) are a dominant feature
261 of its interaction with positively charged side chains of eIF4A (Arg110, Arg161,
262 Arg316, Arg362, and Arg365). Studies elsewhere have implicated a number of
263 these positively charged residues with ATP hydrolysis and RNA binding (Arg362
264 and Arg365 of eIF4A) (Pause et al., 1993). Given this, it seems plausible to
265 suggest that, at least in part, the binding of BLF1 mimics that of the eIF4A sub-
266 strates.

267 Examination of the interaction between BLF1 and Gln339 of eIF4A and
268 residues flanking it, reveals that their interactions are mediated almost exclu-
269 sively by sidechains. Those critical to the recognition include the sidechains of
270 Asp337, Gln339 and Gln340 in eIF4A which point outwards from its molecular
271 surface binding into the pocket provided by BLF1, where major contributions
272 arise from the sidechains of Tyr89, Arg142, Gln145 and Asn168, and the main
273 chain of Cys/Ser94, Ser92 and Tyr90 (Figure S4B & S4C, Table S1). These inter-
274 actions, together with the extensive interfaces formed by the two polar patches,
275 serve to orient the two molecules against each other trapping the two domains
276 of eIF4A in a novel conformation placing the critical side chain of Gln339 in
277 the groove of BLF1 recognising both the nitrogen and oxygen atoms of the car-
278 boxyamide of Gln339 as described below (Figure S4B & S4C).


A
K30
S218
K193
Q29
D219
D194
S31
S78
P32
L188
Q195
K67
G191
Q72
S189
Q75
T73
F192
T158
E70
E124
R161
R190
R110
D128
S122
E111
E127
B
E22
R316
R47
D312
D24
T191
K309
R334
Y90
Q25
S320
Q171
R319
G170
Y89
D337
S92
V338
Q274
W66
Q339
R142
S94
N168
H106
A109
Q340
Q68
D108
Q145
R362
G363
P110
L149
G364
V111
G367
F366
K369
E112
R119
Q115

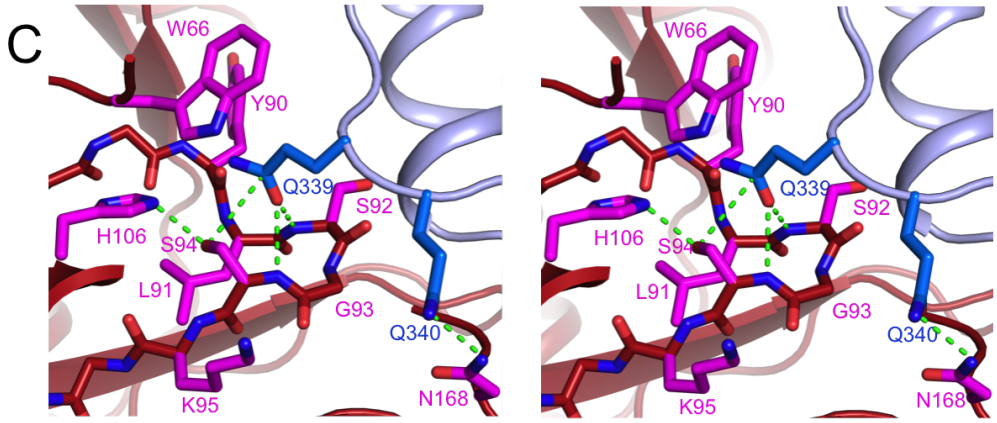


Figure S4: **Cartoon and sticks stereodiagrams to illustrate the specificity of BLF1 towards eIF4A showing an extensive intricate network of interactions forming interfaces with largely polar character.** A) Residues from the N-terminal domain of eIF4A involved in the interaction with BLF1. B) Residues involved at the C-terminal surface of eIF4A. C) Interaction between the residues in the loop containing the target Gln339 of eIF4A with the active site residues of BLF1. eIF4A: light blue, eIF4A side chains; marine, BLF1 C94S: brick red, BLF1 C94S sidechains: magenta.

eIF4A	Distance (Å)	BLF1 C94S
Arg110[NH1]	3.5	Glu124[O]
Arg110[NH1]	3.0	Glu127[OE2]
Agr110[NH2]	3.1	Asp128[OD1]
Arg161[NH1]	2.2	Glu124[OE2]
Ser189[O]	2.9	Thr73[N]
Arg190[O]	2.8	Gln75[NE2]
Asp194[OD1]	3.2	Thr34[N]
Asp194[OD2]	3.0	Ser78[OG]
Asp194[OD2]	2.8	Thr34[OG1]
Asp194[OD2]	3.0	Thr34[N]
Gln195[NE2]	3.6	Ser78[OG]
Asp219[OD2]	3.7	Ser31[N]
Lys309[NZ]	3.8	Asp24[OD1]
Arg316[NH1]	2.7	Glu22[OE2]
Arg319[NE]	2.9	Gln171[OE1]
Arg319[O]	3.6	Gln171[N]
Ser320[O]	3.9	Ser193[N]
Ser322[OG]	3.7	Thr192[O]
Gln339[NE2]	2.5	Tyr90[O]
Gln339[OE1]	3.4	Gly93[N]
Gln339[OE1]	2.9	Ser94[N]
Gln339[OE1]	2.5	Ser92[N]
Gln340[N]	3.1	Ser92[O]
Gln340[OE1]	2.9	Gln145[NE2]
Gln340[OE1]	2.6	Arg142[NH1]
Gln340[OE1]	3.0	Asn168[ND2]
Arg362[NH2]	2.9	Asp108[OD2]
Gly363[N]	2.7	Gln68[OE1]
Arg365[N]	3.4	Gly69[O]
Arg365[N]	2.9	Glu70[OE2]
Arg365[NE]	2.8	Glu70[OE2]
Arg365[NH2]	3.0	Glu70[OE1]
Phe366[N]	3.0	Gly69[O]
Asp337[OD1]	2.9	Tyr89[OH]

Table S1: **Output from PISA, identifying the H-bonds between BLF1 C94S and eIF4A in the complex.**

eIF4A	Distance(Å)	BLF1 C94S
Arg110[NH1]	3.5	Asp128[OD1]
Arg110[NH1]	3.0	Glu127[OE2]
Arg110[NH2]	3.1	Asp128[OD1]
Arg110[NH2]	3.9	Asp128[OD2]
Arg161[NE]	3.5	Glu124[OE2]
Arg161[NH1]	2.2	Glu124[OE2]
Lys309[NZ]	3.8	Asp24[OD1]
Arg316[NH1]	2.7	Glu22[OE2]
Arg362[NH2]	2.9	Asp108[OD2]
Arg365[NE]	3.7	Glu70[OE1]
Arg365[NE]	2.8	Glu70[OE2]
Arg365[NH2]	3.0	Glu70[OE1]
Arg365[NH2]	3.6	Glu70[OE2]

Table S2: **Output from PISA, identifying the salt bridges between BLF1 C94S and eIF4A in the complex.**

279 *3.5. Proposed catalytic mechanism for the glutamine deamidase toxins.*

280 A mechanistic similarity between catalysis by the glutamine deamidase ac-
281 tivity of BLF1 and the cysteine proteases has previously been suggested given
282 the similar orientation of a Cys-His dyad in the active site (Cruz-Migoni et al.,
283 2011). In the cysteine protease papain, the Cys-His dyad is proposed to exist as a
284 thiolate-imidazolium ion pair prior to nucleophilic attack on the substrate by the
285 thiolate anion to form the first anionic tetrahedral intermediate (Storer & Ménard,
286 1994). This proposal has been supported following biochemical and biophysical
287 studies on free and Cys25 thiol-derivatives of papain using spectroscopic tech-
288 niques, fluorometric titrations, kinetic deuterium isotope experiments and proton
289 nuclear magnetic resonance studies (Lewis et al., 1976; Creighton et al., 1980;
290 Johnson et al., 1981; Lewis et al., 1981; Brocklehurst et al., 1988), as reviewed
291 by (Polgár & Halász, 1982). However, more recently, the "ion pair" paradigm
292 had been challenged by theoretical QM/SCRF(VS) calculations suggesting that
293 the Cys-His pair in free papain consists of a protonated histidine associated with
294 a protonated cysteine (SH/NH^+) (Shokhen et al., 2009).

295 Whilst BLF1 and CNF1 share a common β-sandwich tertiary fold, they show
296 only 5 % sequence identity (12 of 211 equivalent residues in a structure-based
297 sequence alignment). However, 6 of these conserved residues are located in
298 their respective active sites including a conserved L-S-G-C motif located on a
299 short loop containing the putative cysteine nucleophile (BLF1: Cys94; CNF1:
300 Cys866). In C-CNF1 the active site cysteine occupies two positions (PDB:
301 1HQ0) complicating an analysis of the role of the cysteine in the mechanism.
302 Moreover, investigations of the mechanism of the cysteine proteases in general
303 are also complicated as a result of the oxidation of the active site cysteine, a

feature that has been commonly noted (PDB: 9PAP, 2CIO, 3P5U, 3P5V, 3P5X) (Kamphuis et al., 1984; Alphey & Hunter, 2006; Yogavel et al., 2010). The presence of one or more oxygen atoms bound to the sulphur of the cysteine side chain clearly has the potential to alter the local geometry and therefore hampers an understanding of the interactions between the residues of the catalytic Cys-His dyad and the substrate. Whilst partial oxidation of Cys94 was a feature of the initial BLF1 WT structure (PDB:3TU8), more recently, to clarify the extent and origin of the oxidation, a new dataset was collected to 0.99 Å on a BLF1 WT crystal in a manner designed to minimise the radiation damage and oxidation. Analysis of the refined structure indicated that the sulphur had not been modified by the oxidation of its side chain, but the position of the cysteine side chain and the surrounding residues was essentially unchanged compared to that seen earlier.

In this reduced BLF1 WT structure, the S_γ of Cys94 is 3.12 Å away from the $N_{\epsilon 2}$ of His106 strongly suggestive of a hydrogen bond between them. Analysis of the geometry of the hydrogen bond donor/acceptor between the two residues indicates that the hydrogen bond has a more linear arrangement if the proton resides on the $N_{\epsilon 2}$ of His106 (160^o) rather than on the S_γ of Cys94 (144^o). Thus, we assume that His106 is protonated indicating the presence of a thiolate-imidazolium ion pair stabilised by hydrogen bonding consistent with the "ion pair" proposal for the mechanism of papain as discussed above. This situation resembles that seen in the serine proteases where a hydrogen bond is observed between the -OH of the serine and the $N_{\epsilon 2}$ of the histidine in the catalytic triad with the key difference being that in BLF1 and the cysteine proteases the proton resides on the nitrogen whereas in the serine proteases it resides on the oxygen

of the serine (Hedstrom, 2002).

We have previously established that the structure of BLF1 and the C94S mutant (PDB:3TUA) are very similar except for the minor local changes arising as a result of the mutation, which induce a 120^{o} rotation in the torsion angle about the bond between C_{α} and C_{β} and which lead to the formation of a hydrogen bond between the O_{γ} of Ser94 and the $N_{\epsilon 2}$ of His106. Thus it is reasonable to assume that the interactions seen between the enzyme and the substrate in the form A structure of the BLF1 C94S:eIF4A$^{\Delta 20}$ complex resemble those in the fully active enzyme. In this structure, Gln339 is located at the active site of BLF1 with the carbonyl carbon of its carboxyamide group lying within 3.4 Å from the O_{γ} of Ser94. Analysis of the relative position of the Ser94 O_{γ} and the carbonyl carbon of the side chain of Gln339 showed that it is not ideally placed for the nucleophilic attack on the carbonyl (Burgi et al., 1974) possibly explaining the inability of BLF1 C94S mutant to deamidate Gln339 of eIF4A (Cruz-Migoni et al., 2011). Modeling the Ser94 as cysteine while maintaining the local geometry of the surrounding residues and applying the torsion angle change of the cysteine as discussed above places the S_{γ} atom approximately 2.7 Å away from the carbonyl carbon making an angle of 88^{o} to the plane of the Gln339 carboxyamide, a conformation which resembles that of Near-Attack Conformers (NAC) (Bruice, 2002). In an active BLF1 WT:eIF4A complex, the change in torsion angle of Cys94 as described above combined with only small movements of the side chain of Gln339 would move the sulphur atom of Cys94 to an appropriate Burgi-Dunitz angle for nucleophilic attack (Burgi et al., 1974). Critically, the side chain carbonyl of Gln339 is hydrogen bonded to the backbone amide moieties of Ser92 and Ser94 which form an oxyanion hole (Figure 3A). This ar-

rangement of an attacking thiolate-imidazolium ion pair in combination with a coordinating oxyanion hole in BLF1 is analogous to the layout of the active site in the papain family of cysteine proteases albeit that the oxyanion hole in papain is formed through hydrogen bonding to the NH_2 sidechain group of Gln19 and a backbone amide (Cys25) (Figure 3B) (Menard et al., 1991).

Superposition of the BLF1 C94S:eIF4A$^{\Delta 20}$ complex with the inactive mutant C151A, of a chymotrypsin-like cysteine protease from Tobacco Etch Virus (TEV), in complex with a peptide substrate (PDB:1LVB) reveals that the δ-carboxyamide of Gln339 in eIF4A is located at a position almost identical to the position of the target peptide bond of the TEV substrate with respect to the Cys-His dyad (Figure 3C). Furthermore, superposition of the papain-like Glutamine deamidase CHYP (Cif-like homologues in *Yersinia pseudotuberculosis*) in complex with NEDD8 (PDB:4F8C) places the N and the O atoms of the amide group of the substrate glutamine Gln40 carboxyamide at an equivalent position to that of Gln339 of eIF4A, albeit with the residue approaching from the opposite direction (Figure 3D). In the structure of the BLF1 C94S:eIF4A$^{\Delta 20}$ complex, the target Gln339 of eIF4A is only some 3.6 Å away from His106 of BLF1, supporting the suggestion that it can act as an acid and donate the proton required for the formation of the acyl-enzyme intermediate which follows the cleavage of the C-N bond. This situation is analogous to the proposed mechanism of papain (Storer & Ménard, 1994). Following the formation of the tetrahedral intermediate in BLF1, the protonated active site imidazolium ion, His106, is then believed to initiate the acylation step of the catalytic cycle by transferring its proton to the substrate amide nitrogen, triggering the release ammonia. This leads to the breakdown of the tetrahedral intermediate, with the cleaved carbonyl group now

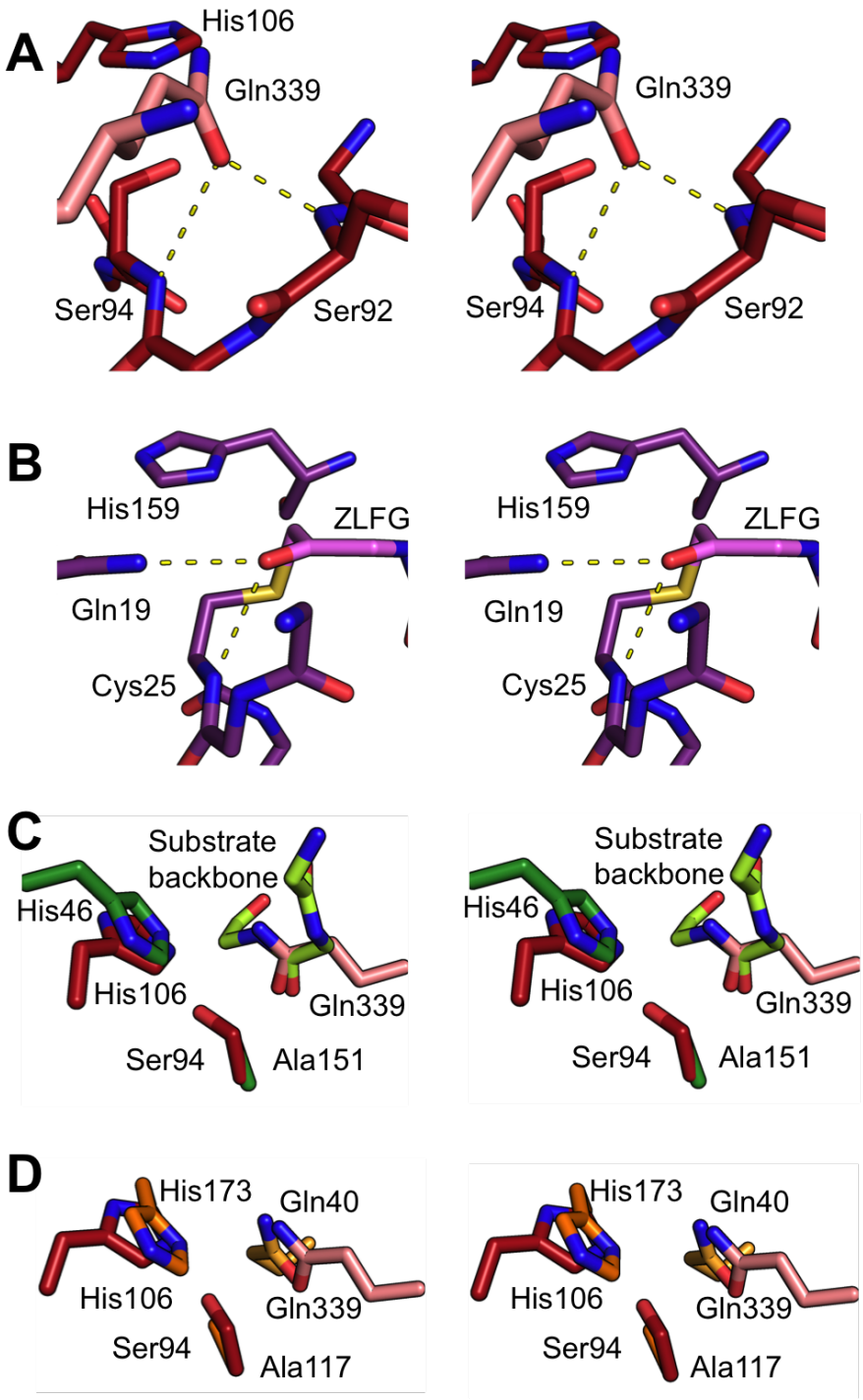


Figure 3: **The relative positions of the oxyanion hole and the substrate in BLF1 C94S:eIF4A compared to other structures.** Stereo images of the comparison between the oxyanion hole of A) BLF1 C94S:eIF4A complex and B) the papain-ZFLG inhibitor complex (Purple; PDB: 1KHP). C) Structural superposition between BLF1 C94S:eIF4A and TEV C151A protease with its substrate (PDB code 1LVB). The target carboxyamide group of the sidechain of Q339 of eIF4A superposes well with the scissile peptide of the TEV substrate. D) Superposition between BLF1 C94S:eIF4A complex and Cif C117A:Nedd8 (PDB code: 4F8C) showing the papain-like glutamine deamidase orientating Q40 into the active site from the opposite side while conserving the position of the carboxyamide oxygen and nitrogen atoms close to the catalytic residues.

bonded to the cysteine sulphur as a covalent acyl-enzyme intermediate (Storer & Ménard, 1994).

Given the above similarities, it is reasonable to propose that the deacylation step of the reaction catalysed by BLF1 involves the imidazole group of His106 acting as a general base extracting a proton from an attacking water molecule resulting in the formation of a second tetrahedral intermediate as in papain (Szawelski & Wharton, 1981). The collapse of this intermediate would then result in the release of the thiolate ion of Cys94 in BLF1 to regenerate the thiolate-imidazolium ion pair with His106 ready for the cycle to restart (Figure S5). The hydrogen bond between of the NH_2 group of the carboxyamide of the Gln339 and the main-chain carbonyl of Tyr90 would clearly be lost upon deamidation. We presume that the conversion of the amide sidechain to a carboxyl group would destabilise the interaction of the product with the enzyme stimulating product release (Figure S4C). The similarities between BLF1 and papain represents an interesting example of convergent evolution.

4. Concluding remarks.

Current therapeutic strategies for Melioidosis entail upwards of 5 months treatment on a cocktail of antibiotics, with 20-50 % of patients not surviving the prescribed treatment (Wuthiekanun & Peacock, 2006). This startlingly poor prognosis upon infection by *B. pseudomallei* is predominantly due to antibiotic resistance and high levels of secreted anti-microbial compounds (Mima & Schweizer, 2010). Therefore, the development of specific therapeutics against melioidosis is critical. A deep understanding of the mechanism of papain/cysteine proteases has been important in the development of these protease inhibitors in a

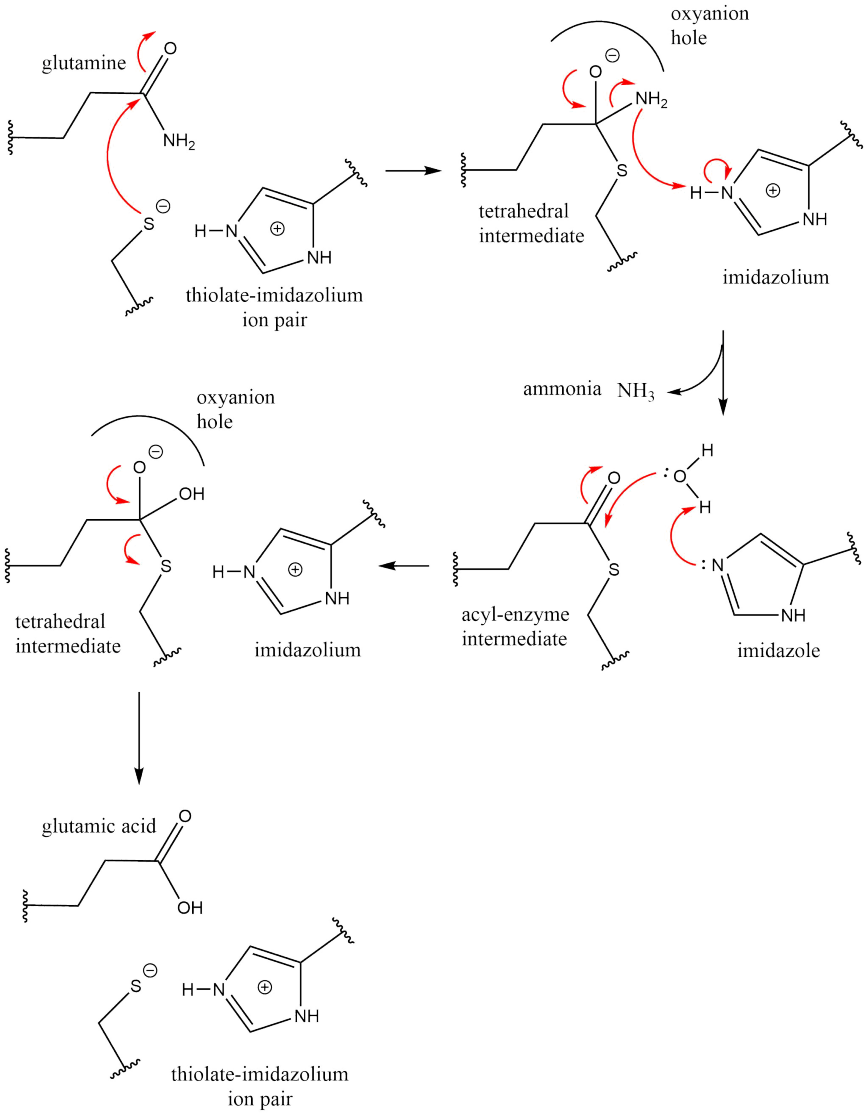


Figure S5: **Proposed Mechanism for Glutamine Deamidase Toxins.** Schematic diagram to show the mechanism of glutamine deamidation by BLF1. From top left in a clockwise direction; the thiolate ion from the thiolate-imidazolium ion pair attacks the carbonyl group of the target Gln forming a tetrahedral intermediate with the oxyanion component stabilised by an oxyanion hole. This unstable intermediate then collapses into an acyl-enzyme intermediate releasing NH_3. The acyl-enzyme intermediate is then hydrolysed forming a second tetrahedral intermediate, which collapses releasing the eIF4A now with Glu339. The thiolate-imidazolium ion pair is now fully restored for the next round of catalysis.

number of therapeutics, with a wide array of commercial and clinical inhibitors either available or in development, particularly against the lysosomal caspases (Rozman-Pungerčar et al., 2003). The work described here has highlighted the molecular factors which control the specificity of the BLF1 toxin for its eukaryotic eIF4A target identifying that the toxin may mimic some elements of the binding of eIF4A for its substrates. Mechanistically, the close similarity of the active site layout in BLF1 to papain including the orientation of the scissile C-N bond, the position of the oxyanion hole and the Cys-His catalytic pair demonstrate their global equivalence. Moreover, the identification of a hydrogen bond between the catalytic Cys-His residues serves, in part, to unify the cysteine and serine protease mechanisms which differ mainly in the charge separation at the active site. In the future, it is hoped that these insights may contribute to the development of inhibitors for the control of melioidosis.

Author Contribution

G.W.M., D.W.R. and P.J.B., designed the research. G.W.M., and A.A.A, Performed protein expression, purification, crystallisation, structure solution, model building and data analysis. S.E.S. analysed and aided protein expression and purification. T.C.M and M.J.D. performed the mass spectrometry experiments and data analysis. G.W.M., A.A.A, D.W.R. and P.J.B. wrote the paper. G.M.B, contributed to the discussion of the mechanism of BLF1.

Acknowledgements

A.A.A acknowledges a PhD scholarship from Majlis Amanah Rakyat (MARA) Malaysia, M.J.D acknowledges support from the Biotechnology and Biological

426 Sciences Research Council UK (BBSRC) (BB/M012166/1), and S.N, F.R, P.J.B
427 and D.W.R acknowledge support from a Royal Society International Collabora-
428 tion Award IC170306.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, *66*, 213–221.

Aldhous, P. (2005). Melioidosis? Never heard of it... *Nature*, *434*, 692–693.

Alphey, M. S., & Hunter, W. N. (2006). High-resolution complex of papain with remnants of a cysteine protease inhibitor derived from Trypanosoma brucei. *Acta crystallographica. Section F, Structural biology and crystallization communications*, *62*, 504–8.

Aubert, D. F., Xu, H., Yang, J., Shi, X., Gao, W., Li, L., Bisaro, F., Chen, S., Valvano, M. A., & Shao, F. (2016). A Burkholderia Type VI Effector Deamidates Rho GTPases to Activate the Pyrin Inflammasome and Trigger Inflammation. *Cell Host & Microbe*, *19*, 664–674.

Brocklehurst, K., Kowlessur, D., Patel, G., Templeton, W., Quigley, K., Thomas, E. W., Wharton, C. W., Willenbrock, F., & Szawelski, R. J. (1988). Consequences of molecular recognition in the S1-S2 intersubsite region of papain for catalytic-site chemistry. Change in pH-dependence characteristics and generation of an inverse solvent kinetic isotope effect by introduction of a P1-P2 amide bond into a two-protonic-state reactivity probe. *The Biochemical journal*, *250*, 761–72.

Bruice, T. C. (2002). A View at the Millennium: the Efficiency of Enzymatic Catalysis. *Accounts of Chemical Research*, *35*, 139–148.

Buetow, L., Flatau, G., Chiu, K., Boquet, P., & Ghosh, P. (2001). Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1. *Nature structural biology*, *8*, 584–588.

Burgi, H., Dunitz, J., Lehn, J., & Wipff, G. (1974). Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron*, *30*, 1563–1572.

Caruthers, J. M., Johnson, E. R., & McKay, D. B. (2000). Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 13080–5.

Creighton, D. J., Gessouroun, M. S., & Heapes, J. M. (1980). Is the thiolate-imidazolium ion pair the catalytically important form of papain? *FEBS Letters*, *110*, 319–322.

460 Crow, A., Hughes, R. K., Taieb, F., Oswald, E., & Banfield, M. J. (2012). The molecular basis of
461 ubiquitin-like protein NEDD8 deamidation by the bacterial effector protein Cif. *Proceedings*
462 *of the National Academy of Sciences of the United States of America*, *109*, E1830–8.
463 Cruz-Migoni, a., Hautbergue, G. M., Artymiuk, P. J., Baker, P. J., Bokori-Brown, M., Chang,
464 C.-T., Dickman, M. J., Essex-Lopresti, A., Harding, S. V., Mahadi, N. M., Marshall, L. E.,
465 Mobbs, G. W., Mohamed, R., Nathan, S., Ngugi, S. a., Ong, C., Ooi, W. F., Partridge, L. J.,
466 Phillips, H. L., Raih, M. F., Ruzheinikov, S., Sarkar-Tyson, M., Sedelnikova, S. E., Smither,
467 S. J., Tan, P., Titball, R. W., Wilson, S. a., & Rice, D. W. (2011). A Burkholderia pseudomallei
468 Toxin Inhibits Helicase Activity of Translation Factor eIF4A. *Science*, *334*, 821–824.
469 Currie, B. J. (2008). Advances and remaining uncertainties in the epidemiology of Burkholderia
470 pseudomallei and melioidosis. *Transactions of the Royal Society of Tropical Medicine and*
471 *Hygiene*, *102*, 225–227.
472 Dance, D. A. (2000). Ecology of Burkholderia pseudomallei and the interactions between envi-
473 ronmental Burkholderia spp. and human-animal hosts. *Acta tropica*, *74*, 159–68.
474 Emsley, P., & Cowtan, K. (2004). *Coot* : model-building tools for molecular graphics. *Acta*
475 *Crystallographica Section D Biological Crystallography*, *60*, 2126–2132.
476 Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution?
477 *Acta crystallographica. Section D, Biological crystallography*, *69*, 1204–14.
478 Flatau, G., Lemichez, E., Gauthier, M., Chardin, P., Paris, S., Fiorentini, C., & Boquet, P. (1997).
479 Toxin-induced activation of the G protein p21Rho by deamidation of glutamine. *Nature*, *387*,
480 729–733.
481 Hedstrom, L. (2002). Serine Protease Mechanism and Specificity. *Chem. Rev.*, *102*, 4501–4524.
482 Jabbar, Z., & Currie, B. J. (2013). Melioidosis and the kidney. *Nephrology*, *18*, 169–175.
483 Jilani, M. S. A., Robayet, J. A. M., Mohiuddin, M., Hasan, M. R., Ahsan, C. R., & Haq, J. A.
484 (2016). Burkholderia pseudomallei: Its Detection in Soil and Seroprevalence in Bangladesh.
485 *PLOS Neglected Tropical Diseases*, *10*, e0004301.
486 Johnson, F. A., Lewis, S. D., & Shafer, J. A. (1981). Determination of a low pK for histidine-159
487 in the S-methylthio derivative of papain by proton nuclear magnetic resonance spectroscopy.
488 *Biochemistry*, *20*, 44–48.
489 Kabsch, W. (2010). XDS. *Acta crystallographica. Section D, Biological crystallography*, *66*,
490 125–32.

491 Kamphuis, I., Kalk, K., Swarte, M., & Drenth, J. (1984). Structure of papain refined at 1.65 Å
492 resolution. *Journal of Molecular Biology*, *179*, 233–256.
493 Kanaphun, P., Thirawattanasuk, N., Suputtamongkol, Y., Naigowit, P., Dance, D. A., Smith,
494 M. D., & White, N. J. (1993). Serology and carriage of Pseudomonas pseudomallei: a
495 prospective study in 1000 hospitalized children in northeast Thailand. *The Journal of in-*
496 *fectious diseases*, *167*, 230–3.
497 Kronmann, K. C., Truett, A. A., Hale, B. R., & Crum-Cianflone, N. F. (2009). Melioidosis after
498 brief exposure: a serologic survey in US Marines. *The American journal of tropical medicine*
499 *and hygiene*, *80*, 182–4.
500 Lewis, S. D., Johnson, F. A., & Shafer, J. A. (1976). Potentiometric determination of ionizations
501 at the active site of papain. *Biochemistry*, *15*, 5009–5017.
502 Lewis, S. D., Johnson, F. A., & Shafer, J. A. (1981). Effect of cysteine-25 on the ionization
503 of histidine-159 in papain as determined by proton nuclear magnetic resonance spectroscopy.
504 Evidence for a histidine-159-cysteine-25 ion pair and its possible role in catalysis. *Biochem-*
505 *istry*, *20*, 48–51.
506 Linder, P., & Jankowsky, E. (2011). From unwinding to clamping the DEAD box RNA helicase
507 family. *Nature Reviews Molecular Cell Biology*, *12*, 505–516.
508 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J.
509 (2007). Phaser crystallographic software. *Journal of applied crystallography*, *40*, 658–674.
510 Menard, R., Carriere, J., Laflamme, P., Plouffe, C., Khouri, H. E., Vernet, T., Tessier, D. C.,
511 Thomas, D. Y., & Storer, A. C. (1991). Contribution of the glutamine 19 side chain to
512 transition-state stabilization in the oxyanion hole of papain. *Biochemistry*, *30*, 8924–8928.
513 Mima, T., & Schweizer, H. P. (2010). The BpeAB-OprB Efflux Pump of Burkholderia pseu-
514 domallei 1026b Does Not Play a Role in Quorum Sensing, Virulence Factor Production, or
515 Extrusion of Aminoglycosides but Is a Broad-Spectrum Drug Efflux System. *Antimicrobial*
516 *Agents and Chemotherapy*, *54*, 3113–3120.
517 Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of Macromolecular Struc-
518 tures by the Maximum-Likelihood Method. *Acta Crystallographica Section D Biological*
519 *Crystallography*, *53*, 240–255.
520 Nandi, T., Ong, C., Singh, A. P., Boddey, J., Atkins, T., Sarkar-Tyson, M., Essex-Lopresti, A. E.,
521 Chua, H. H., Pearson, T., Kreisberg, J. F., Nilsson, C., Ariyaratne, P., Ronning, C., Losada, L.,

Ruan, Y., Sung, W.-K., Woods, D., Titball, R. W., Beacham, I., Peak, I., Keim, P., Nierman, W. C., & Tan, P. (2010). A Genomic Survey of Positive Selection in Burkholderia pseudomallei Provides Insights into the Evolution of Accidental Virulence. *PLoS Pathogens*, *6*, e1000845.

Ngauy, V., Lemeshev, Y., Sadkowski, L., & Crawford, G. (2005). Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II. *Journal of clinical microbiology*, *43*, 970–2.

Pause, A., Méthot, N., & Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Molecular and cellular biology*, *13*, 6789–98.

Pestova, T. V., Shatsky, I. N., & Hellen, C. U. (1996). Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Molecular and cellular biology*, *16*, 6870–8.

Polgár, L., & Halász, P. (1982). Current problems in mechanistic studies of serine and cysteine proteinases. *The Biochemical journal*, *207*, 1–10.

Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C., & Thach, R. E. (1985). ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *The Journal of biological chemistry*, *260*, 7651–8.

Ray, K., Marteyn, B., Sansonetti, P. J., & Tang, C. M. (2009). Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nature Reviews Microbiology*, *7*, 333–340.

Rotz, L. D., Khan, A. S., Lillibridge, S. R., Ostroff, S. M., & Hughes, J. M. (2002). Public health assessment of potential biological terrorism agents. *Emerging infectious diseases*, *8*, 225–30.

Rozman-Pungerčar, J., Kopitar-Jerala, N., Bogyo, M., Turk, D., Vasiljeva, O., Štefe, I., Vandenabeele, P., Brömme, D., Puizdar, V., Fonović, M., Trstenjak-Prebanda, M., Dolenc, I., Turk, V., & Turk, B. (2003). Inhibition of papain-like cysteine proteases and legumain by caspase-specific inhibitors: when reaction mechanism is more important than specificity. *Cell Death & Differentiation*, *10*, 881–888.

Schütz, P., Bumann, M., Oberholzer, A. E., Bieniossek, C., Trachsel, H., Altmann, M., & Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proceedings of the National Academy of Sciences*

of the United States of America, *105*, 9564–9.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., & Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell*, *125*, 287–300.

Shokhen, M., Khazanov, N., & Albeck, A. (2009). Challenging a paradigm: theoretical calculations of the protonation state of the Cys25-His159 catalytic diad in free papain. *Proteins*, *77*, 916–26.

Storer, A., & Ménard, R. (1994). Catalytic mechanism in papain family of cysteine peptidases. *Methods in Enzymology*, *244*, 486–500.

Szawelski, R. J., & Wharton, C. W. (1981). Kinetic solvent isotope effects on the deacylation of specific acyl-papains Proton inventory studies on the papain-catalysed hydrolyses of specific ester substrates: analysis of possible transition state structures. *Biochem. J*, (pp. 681–692).

Vidyalakshmi, K., Chakrapani, M., Shrikala, B., Damodar, S., Lipika, S., & Vishal, S. (2008). Tuberculosis mimicked by melioidosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, *12*, 1209–15.

Washington, E. J., Banfield, M. J., & Dangl, J. L. (2013). What a difference a Dalton makes: bacterial virulence factors modulate eukaryotic host cell signaling systems via deamidation. *Microbiology and molecular biology reviews : MMBR*, *77*, 527–39.

Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P., & Peacock, S. J. (2006). Melioidosis: insights into the pathogenicity of Burkholderia pseudomallei. *Nature reviews. Microbiology*, *4*, 272–282.

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. W., 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 *CCP* 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, *67*, 235–242.

Winter, G., Lobley, C. M. C., & Prince, S. M. (2013). Decision making in xia2. *Acta crystallographica. Section D, Biological crystallography*, *69*, 1260–73.

Wuthiekanun, V., & Peacock, S. J. (2006). Management of melioidosis. *Expert Review of Anti-infective Therapy*, *4*, 445–455.

Yogavel, M., Nithya, N., Suzuki, A., Sugiyama, Y., Yamane, T., Velmurugan, D., & Sharma, A.

584 (2010). Structural analysis of actinidin and a comparison of cadmium and sulfur anomalous
585 signals from actinidin crystals measured using in-house copper- and chromium-anode X-ray
586 sources. *Acta Crystallographica Section D Biological Crystallography*, *66*, 1323–1333.

Appendix A

Extended Methodology - BLF1C94S:eIF4A$^{\Delta 20}$ complex

This appendix serves as an overview of the body of work done that was omitted from the analysis and discussion in Paper 1. This includes details on the purification, crystallisation and subsequent structure determination of BLF1 C94S:eIF4A$^{\Delta 20}$ in crystal form B. The corresponding details on crystal form A of this complex are detailed in the thesis of George Mobbs (Mobbs, 2014). Several aspects on the analysis of the two crystal forms of BLF1 C94S:eIF4A$^{\Delta 20}$ will also be described here.

A.1 Details of the BLF1 C94S:eIF4A$^{\Delta 20}$ complex structure of crystal form B

A.1.1 Purification, crystallisation and data collection of BLF1 C94S in complex with eIF4A$^{\Delta 20}$

The purified sample of the BLF1 C94S:eIF4A$^{\Delta 20}$ complex was obtained as described in Paper 1, but a 2:1 molar ratio of BLF1 C94S to eIF4A was used during the gel filtration

(Figure A.1). Subsequent crystallisation trials and optimisation led to the production of crystals (Figure A.2), which following data collection, produced diffraction spots to 2.2 Å (Figure A.2) and processing statistics are shown in Figure A.3.

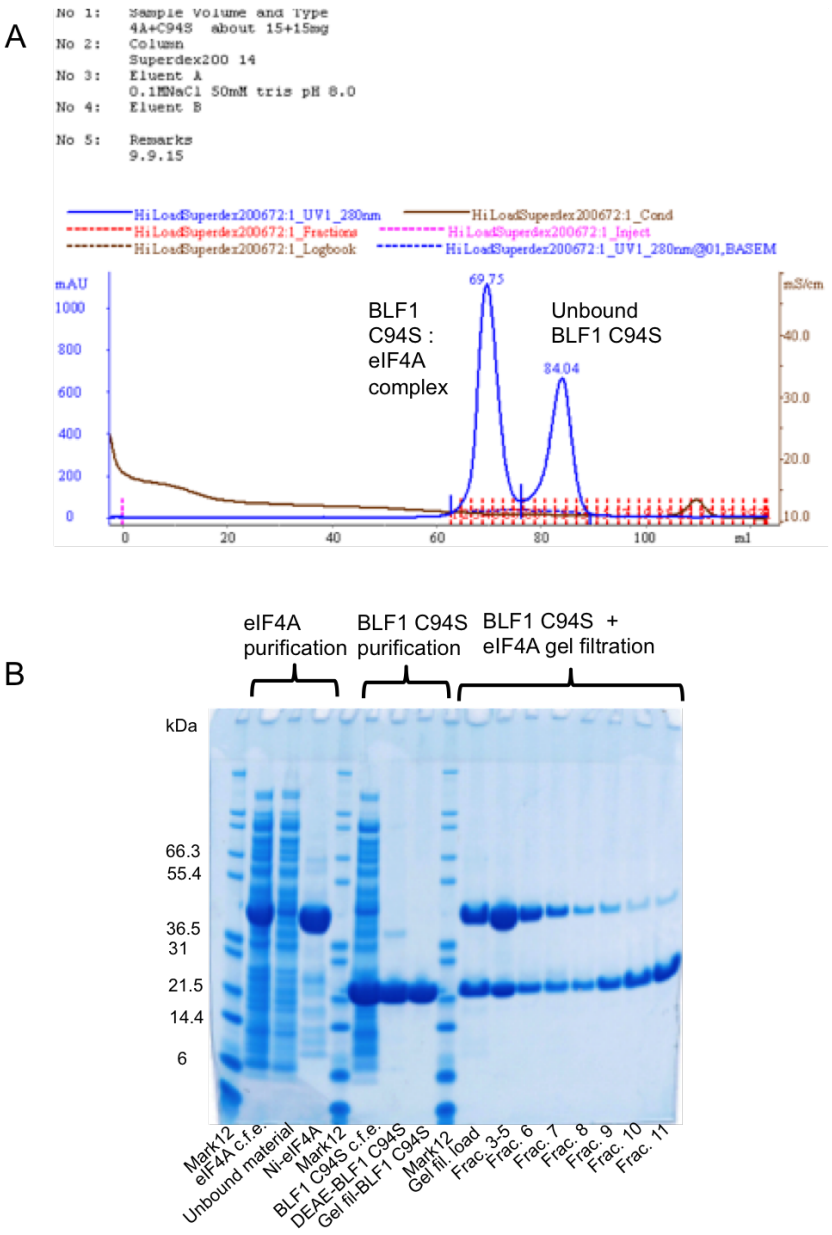


Fig. A.1 **Purification of BLF1 C94S:eIF4A$^{\Delta 20}$ complex.** A) Chromatogram trace of gel filtration of BLF1 C94S:eIF4A$^{\Delta 20}$ complex showing a good resolution between the complex and unbound BLF1 proteins. B) SDS-PAGE analysis of the various purification stages in producing a 1:1 complex of BLF1 C94S:eIF4A$^{\Delta 20}$ (Frac. 3-5).

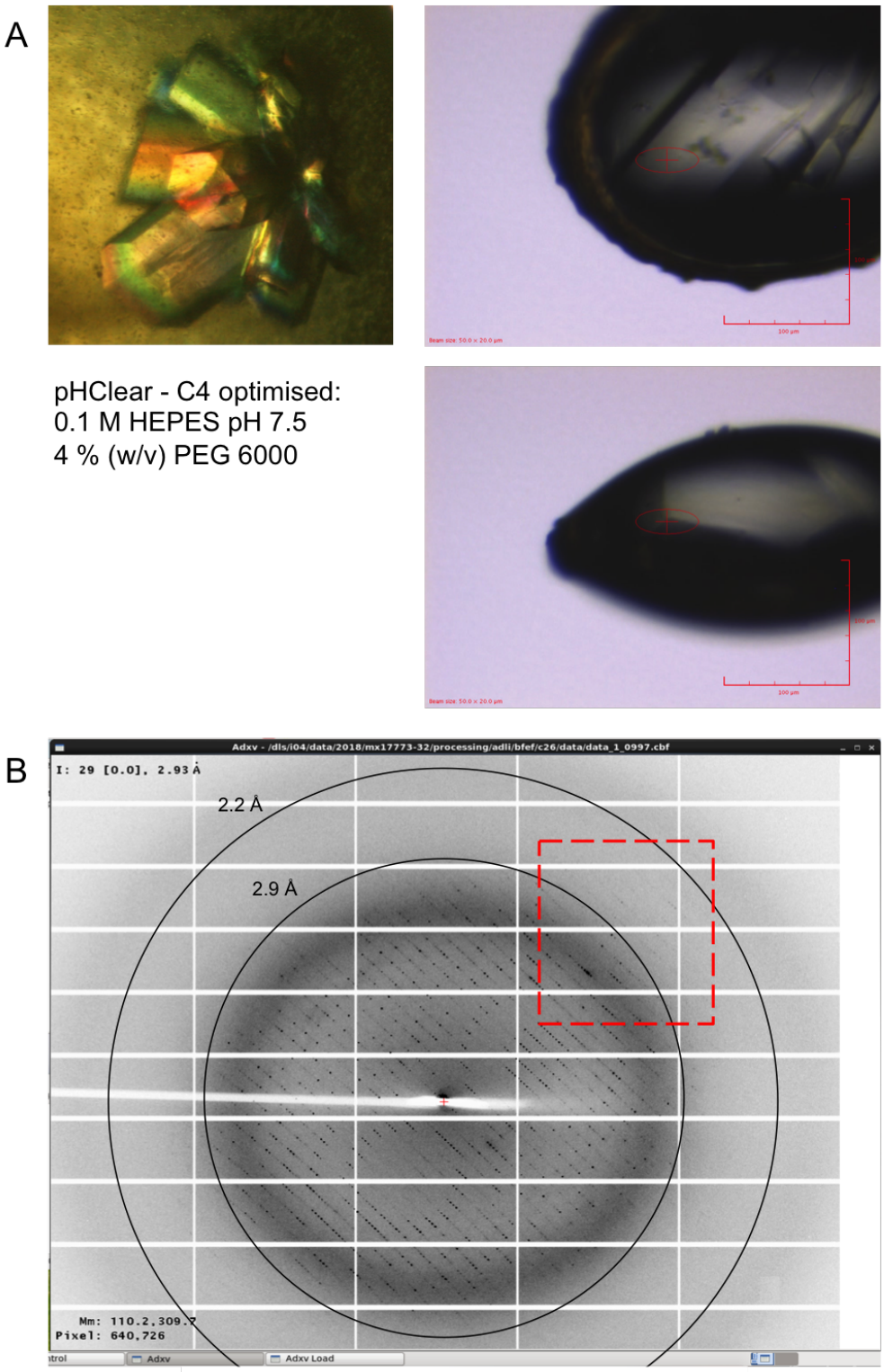


Fig. A.2 **Crystals of BLF1 C94S:eIF4A$^{\Delta 20}$ complex in form B and subsequent X-ray diffraction analysis.** A) Crystals were obtained from an optimised condition and subsequently mounted on a loop for analysis. B) A sample of a diffraction image at 0.1^{o} revealing anisotropy (red dashed-square).

```
============================================================

<!--SUMMARY_BEGIN--> $TEXT:Result: $$ $$
Summary data for        Project: mx12788v4 Crystal: xdata1 Dataset: SAD

                                         Overall  InnerShell  OuterShell
Low resolution limit                      65.30     65.30      2.30
High resolution limit                      2.24     10.02      2.24

Rmerge  (within I+/I-)                    0.078     0.028     2.191
Rmerge  (all I+ and I-)                   0.083     0.029     2.316
Rmeas (within I+/I-)                      0.087     0.031     2.437
Rmeas (all I+ & I-)                       0.088     0.032     2.439
Rpim (within I+/I-)                       0.038     0.014     1.062
Rpim (all I+ & I-)                        0.028     0.011     0.760
Rmerge in top intensity bin               0.027         -         -
Total number of observations             570143      5923     41991
Total number unique                       56669       702      4099
Mean((I)/sd(I))                            16.6      68.5       1.1
Mn(I) half-set correlation CC(1/2)        0.999     0.998     0.626
Completeness                               99.9      99.6      99.8
Multiplicity                               10.1       8.4      10.2

Anomalous completeness                     99.8      99.4      99.7
Anomalous multiplicity                      5.0       4.8       5.1
DelAnom correlation between half-sets     0.167    -0.002     0.047
Mid-Slope of Anom Normal Probability      1.078         -         -

Estimate of maximum resolution for significant anomalous signal =  3.28A, from CCanom >  0.15

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  2.24A  == maximum resolution
   from Mn(I/sd) >  1.50:                         limit =  2.36A

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  2.57A
   from Mn(I/sd) >  1.50:                         limit =  2.66A
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  2.24A  == maximum resolution
   from Mn(I/sd) >  1.50:                         limit =  2.24A  == maximum resolution

Anisotropic deltaB (i.e. range of principal components), A^2: 29.02

Average unit cell:  199.49  199.49   51.33   90.00   90.00  120.00
Space group: P 63
Average mosaicity:   0.00

Minimum and maximum SD correction factors: Fulls   0.75   2.40 Partials   0.00   0.00
Anomalous flag switched ON in input, strong anomalous signal found

$$ <!--SUMMARY_END-->
============================================================
```

Fig. A.3 **Data processing statistics of BLF1 C94S:eIF4A$^{\Delta 20}$ complex in form B from automated Xia2 on ispyb SyncWeb.** Initial processing results from automated Xia2 3dii pipeline showing reasonable values for the majority of quality indicators except for slightly high R_{merge} & R_{meas} in the highest resolution shell.

A.1.2 Structure determination, model building and refinement of BLF1 C94S:eIF4A$^{\Delta 20}$ complex in crystal form B

Structure determination of the crystal form B of BLF1 C94S:eIF4A$^{\Delta 20}$ complex at 2.24 Å was carried out using PhaserMR (McCoy et al., 2007) as described in Paper 1 (Figure A.5). Preliminary cycles of model building and refinement revealed relatively poor electron density for some regions of eIF4A (Figure A.5). Subsequent model building and refinement including TLS (Torsion/Liberation/Screw) refinement did not improve the electron density map, and the model building process was hampered by inappropriate R_{work} and R_{free} values. Analysis of the R_{merge} against batches at the last 30^{o} rotation diffraction images showed higher values relative to that of the first 150^{o}, which could have been the result of radiation damage (Figure A.4). In addition, investigations on the diffraction pattern displayed some degree of anisotropy (Figure A.2), which subsequently led to the decision to reprocess the dataset with a resolution cut-off of 3.0 Å and excluding the last 30^{o} rotation diffraction images, resulting in better statistics (Figure A.6). The final data reduction statistics, refinement and validation are detailed in Paper 1. Unfortunately, the low resolution nature of the structure of crystal form B limits the accuracy of the analysis of interactions formed between BLF1 C94S and eIF4A. Nevertheless, the crystallisation of the complex in a different crystal form provided a level of confidence that the formation of the complex was not merely a crystallisation artefact.

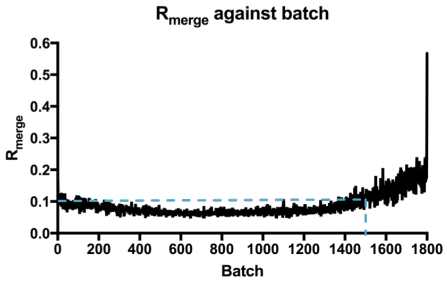


Fig. A.4 **A graph of R_{merge} against batch.** R_{merge} values following batch 1500 (150^{o} rotation) are relatively higher than earlier batches.

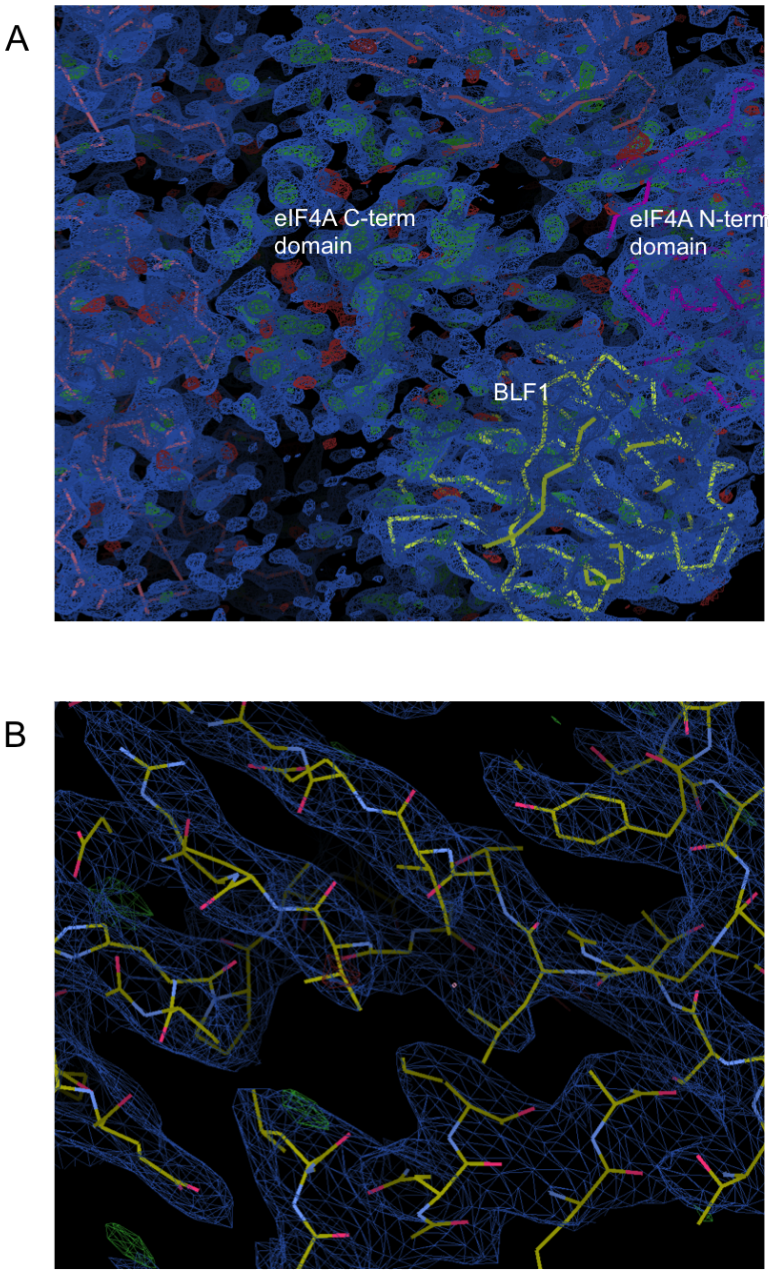


Fig. A.5 **Structure determination of BLF1 C94S:eIF4A$^{\Delta 20}$ complex in crystal form B.** A) Preliminary electron density map during molecular replacement using coordinates of BLF1 and the N-terminal domain of eIF4A revealing information on the location of the C-terminal domain of eIF4A. B) The subsequent model building and refinement produced poor electron density maps in certain regions inconsistent with the higher resolution limit of the data at 2.2 Å, and suggesting an inherent problem with the dataset. The electron density maps were contoured to 1.0 σ level.

```
============================================================

<!--SUMMARY_BEGIN--> $TEXT:Result: $$ $$
Summary data for        Project: AUTOMATIC Crystal: DEFAULT Dataset: NATIVE

                                         Overall  InnerShell  OuterShell
Low resolution limit                     65.31     65.31       3.08
High resolution limit                     3.00     13.42       3.00

Rmerge  (within I+/I-)                   0.039     0.028      0.204
Rmerge  (all I+ and I-)                  0.041     0.029      0.216
Rmeas (within I+/I-)                     0.044     0.033      0.232
Rmeas (all I+ & I-)                      0.044     0.032      0.230
Rpim (within I+/I-)                      0.021     0.016      0.109
Rpim (all I+ & I-)                       0.015     0.013      0.078
Rmerge in top intensity bin              0.024       -          -
Total number of observations            198942      2006      14304
Total number unique                      23808       301       1702
Mean((I)/sd(I))                           34.9      67.9        9.0
Mn(I) half-set correlation CC(1/2)       0.999     0.997      0.988
Completeness                              99.7      98.9       97.9
Multiplicity                               8.4       6.7        8.4

Anomalous completeness                    99.4      98.0       95.6
Anomalous multiplicity                     4.3       4.0        4.3
DelAnom correlation between half-sets    -0.094    -0.213     -0.016
Mid-Slope of Anom Normal Probability      0.929       -          -

No significant anomalous signal

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.00A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  3.00A  == maximum resolution
   from Mn(I/sd) >  2.00:                       limit =  3.00A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.00A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  3.00A  == maximum resolution
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.00A  == maximum resolution
   from Mn(I/sd) >  1.50:                       limit =  3.00A  == maximum resolution

Anisotropic deltaB (i.e. range of principal components), A^2: 36.32

Average unit cell:  199.53  199.53   51.34   90.00   90.00  120.00
Space group: P 63
Average mosaicity:   0.00

Minimum and maximum SD correction factors: Fulls   0.78   2.69 Partials   0.00   0.00
Anomalous flag switched OFF in input, anomalous signal is weak

$$ <!--SUMMARY_END-->
============================================================
```

Fig. A.6 **Data processing statistics of BLF1 C94S:eIF4A$^{\Delta 20}$ complex in form B from xia2 3dii at 3.0 Å.** The R_{merge} & R_{meas} have improved by a factor of 10 compared to the data processed at 2.2 Å.

A.2 Details on the BLF1 C94S:eIF4A$^{\Delta 20}$ complex structure of form A

The initial model of the structure of crystal form A provided by George Mobbs, was further refined and validated as described in Section 2.14.4. The final refinement statistics can be seen in Paper 1. The structure of form A was then compared to that of form B and analysed for the interactions formed between BLF1 C94S and eIF4A to shed light on the substrate specificity and the aspects of mechanism, full details of which, were discussed in Paper 1.

A.3 Crystallisation, structure determination and analysis of BLF1 WT

Included in the discussion on the mechanism of BLF1 in Paper 1 were issues regarding the protonation state of the Cys-His pair in the active site. This analysis was difficult due to the cysteine residue, Cys94, being oxidised in the BLF1 WT structure that was previously solved (PDB:3TU8) (Cruz-Migoni et al., 2011), possibly due to radiation damage. This led to a decision to collect a new datasets aiming to minimise radiation damage and to obtain a crystal structure with the cysteine unoxidised. Production and crystallisation of BLF1 WT was carried out as previously described (Figure A.7) (Cruz-Migoni et al., 2011). Data were collected at 0.99 Å and the structure was determined by molecular replacement using PhaserMR (McCoy et al., 2007) with the previous BLF1 WT model (PDB:3TU8) being used as a search model. The model was then refined and validated, see Paper 1, revealing an unoxidised cysteine at position 94 (Figure A.8). The distance between the Sγ atom of Cys94 and the Nε2 of His106 was found to be 3.1 Å suggestive of a hydrogen bond (Figure A.8). Although the resolution precluded the observation of the position of hydrogen atoms, their position could be inferred based on neighbouring residues. Analysis of the hydrogen bond

angles between Cys94 and His106 suggests that the hydrogen atom resides on the His106 Nε2 (Figure A.9). Further arguments on the "ion pair" proposal are detailed in Paper 1.

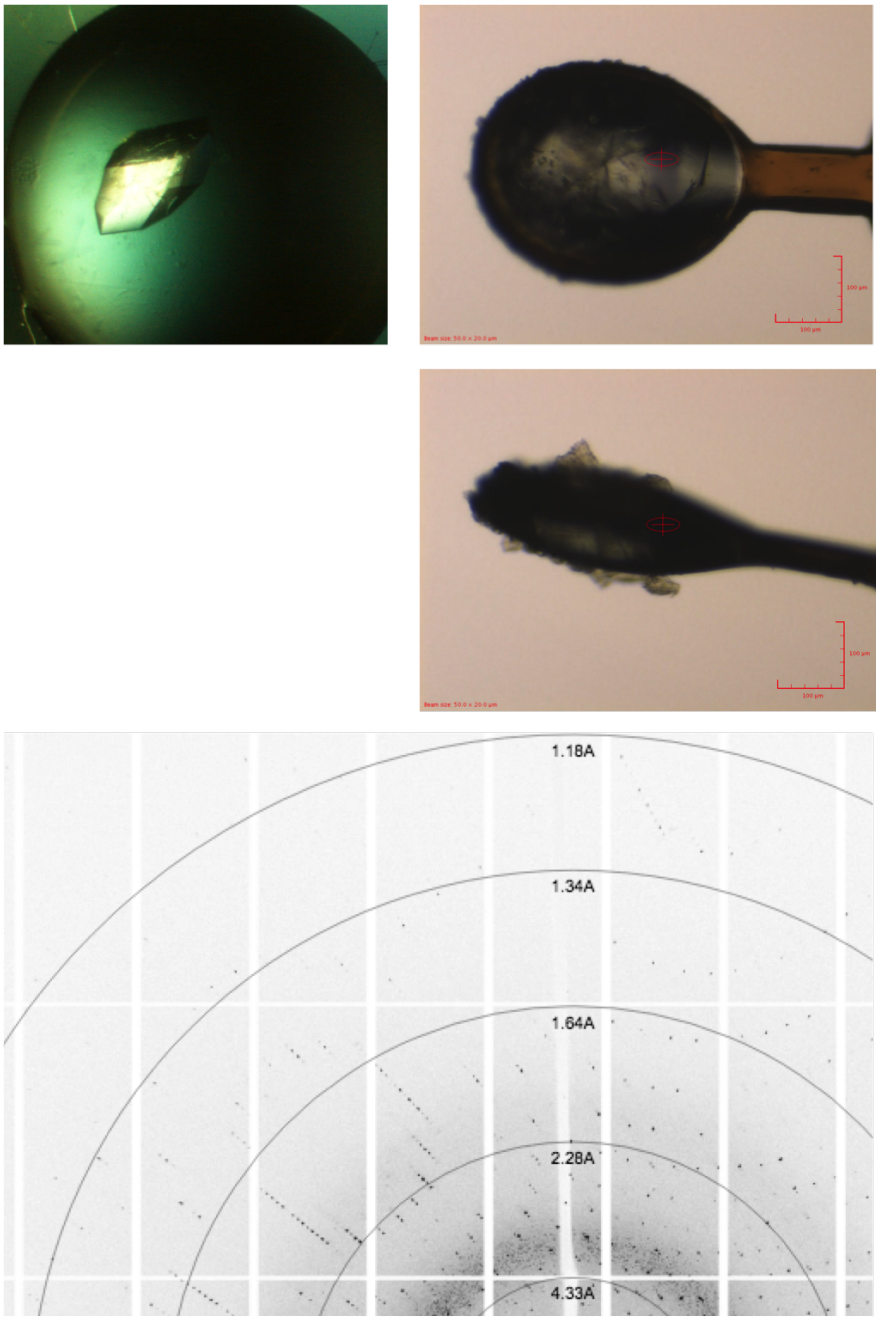


Fig. A.7 **Crystals of BLF1 WT and the resulting X-ray diffraction pattern.** A) BLF1 WT was crystallised as described previously and crystals were subjected to X-ray diffraction analysis. B) A diffraction image for 0.1^{o} rotation with resolution rings labelled.

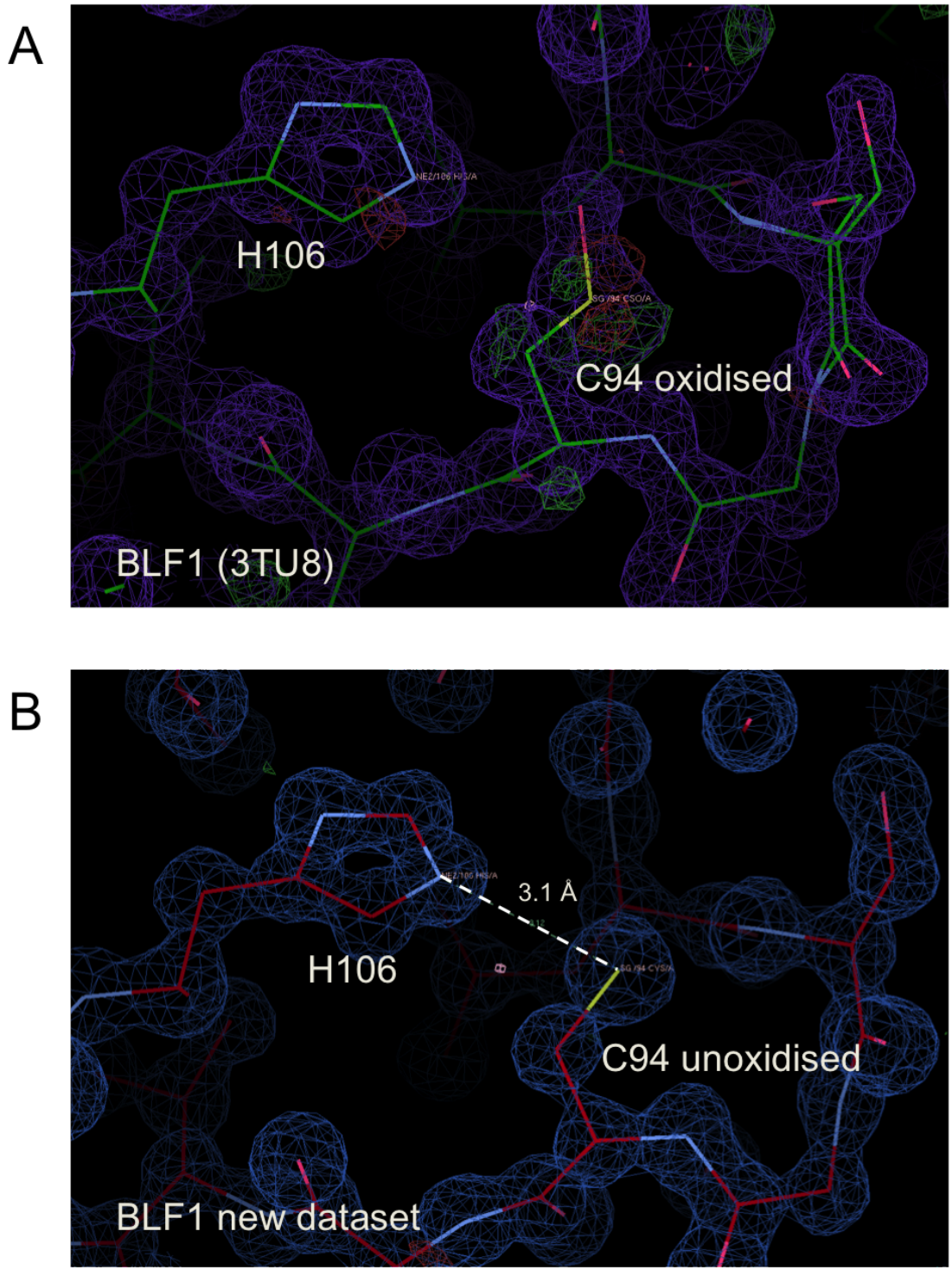


Fig. A.8 **Comparison of the Cys94 residue in the two structures of BLF1 WT.** A) Cys94 is oxidised in the previous crystal structure of BLF1 WT (3TU8). B) Clear electron density confirming the absence of oxidation in the Cys94 in the new crystals with the Sγ of Cys94 located only 3.1 Å away from the Nε2 of His106. The electron density maps were contoured to 1.0 σ level.

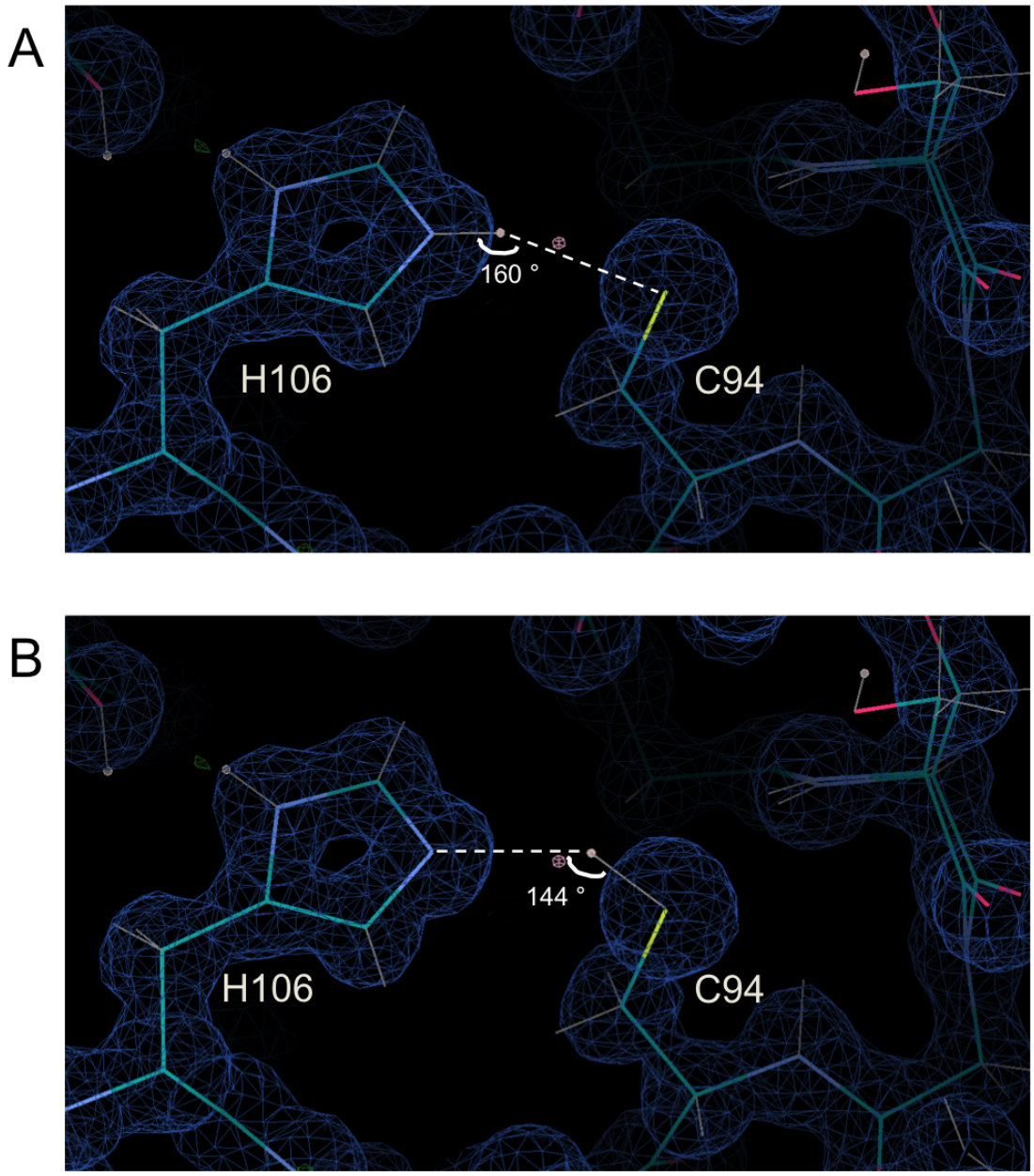


Fig. A.9 **The protonation states of the Cys-His pair.** A) The hydrogen bond angle if the proton resides on the Nε2 of His106. B) The hydrogen bond angle if the proton resides on the Sγ of Cys94. The electron density maps were contoured to 1.0 σ level.

A.4 Comparison between crystal structures of form A and B of BLF1 C94S:eIF4A$^{\Delta 20}$ and further investigations on papain, TEV protease and Cif deamidase

Details of the comparison between the two crystal forms of BLF1 C94S:eIF4A$^{\Delta 20}$ have been covered in Paper 1. As mentioned previously, the data for crystal form B of BLF1 C94S:eIF4A complex were reprocessed to 3.0 Å. This resulted in a slight improvement in electron density (Figure A.10) and better correlation between R_{work}/R_{free} values at this resolution. Critically, the Gln339 of eIF4A in crystal form B is shifted, presumably as a result of the binding of a buffer component in the active site, into a conformation not likely to be appropriate for catalysis (Figure A.10). This switched the focus of the studies on specificity and mechanism of BLF1 towards eIF4A to the crystal structure of form A. Subsequent analysis on the conformation of Ser94 in crystal form A revealed that the Burgi-Dunitz angle between the Sγ of Ser94 towards the plane of the side chain carbonyl group of Gln339 as shown in Figure A.11, is inappropriate for nucleophilic attack.

Multiple sequence alignment of a range of eukaryotic eIF4A proteins was carried out via Multalin (Corpet, 1988) and was manually edited with ESPript3.0 (http://espript.ibcp.fr) (Robert and Gouet, 2014). The schematic diagram for the proposed mechanism of GDT was made using ChemDraw Prime 16.0.1.4-AV5 (PerkinElmer Informatics).

Superpositions of coordinates were carried out using Superpose (Secondary-Structure Matching) (Krissinel and Henrick, 2004) or LSKQAB on the CCP4 GUI (Kabsch, 1976) and manually observed on *Coot* (Emsley et al., 2010). Figures were generated by PyMOL (Schrodinger, LLC). Analysis of the interacting surfaces between BLF1 C94S and eIF4A in crystal form A was carried out using PDBePISA (Krissinel and Henrick, 2007) that disseminated details such as surface area buried by interactions, interacting residues in the interface, as well as the type of interactions and the corresponding bond distances (Paper 1).

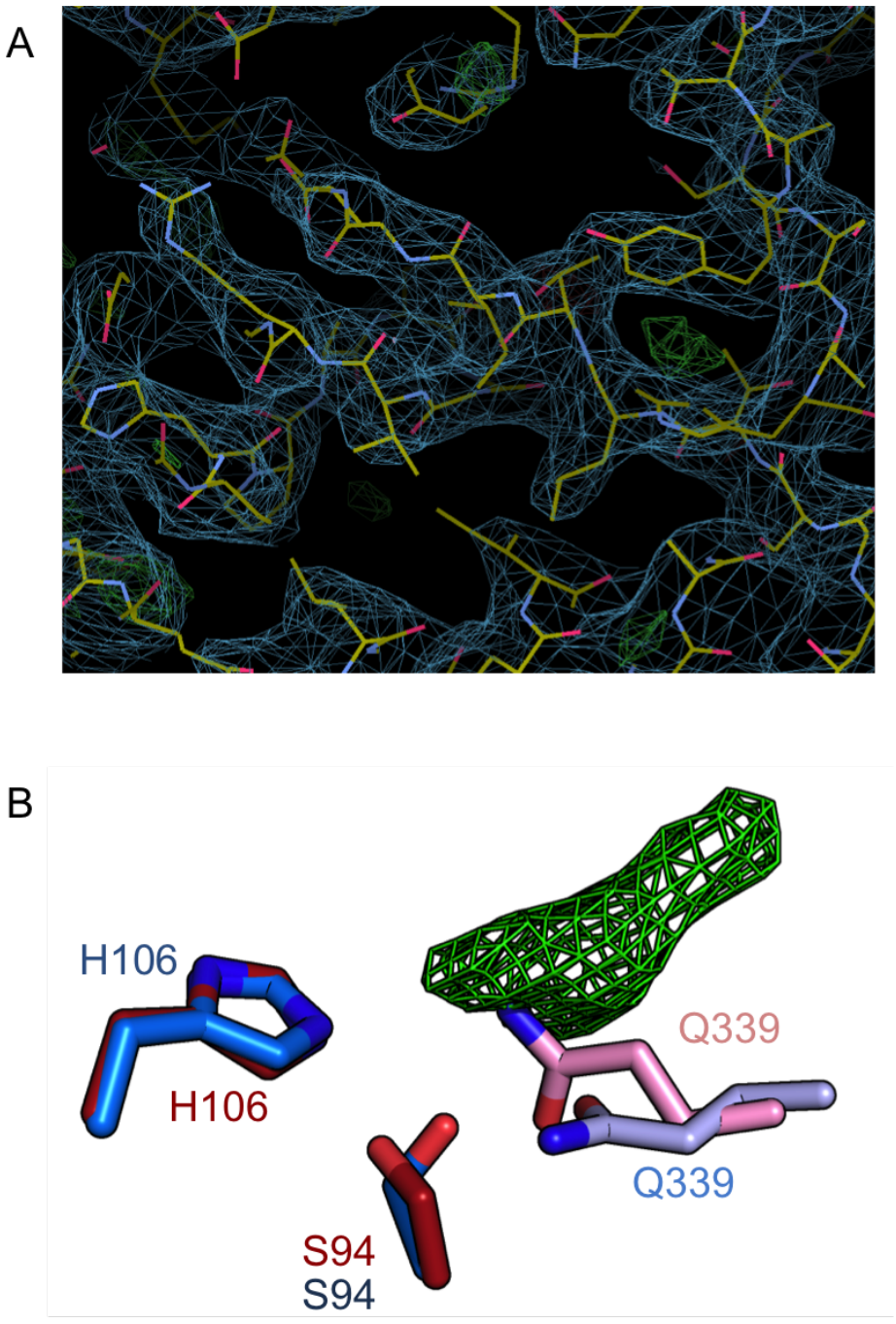


Fig. A.10 **The refined map at 3.0 Å of BLF1 C94S:eIF4A crystal form B.** A) The electron density map of crystal form B at 3.0 Å contoured at 1.0 σ level, showing minor improvement than that processed at 2.2 Å. B) Superposition of crystal forms A and B based on the dyad of BLF1 and the target Gln339 of eIF4A (red:pink and blue:light blue for form A and B, respectively). The binding of an un-modelled buffer component in the latter shifted the Gln339 of eIF4A into a non-productive conformation for nucleophilic attack.

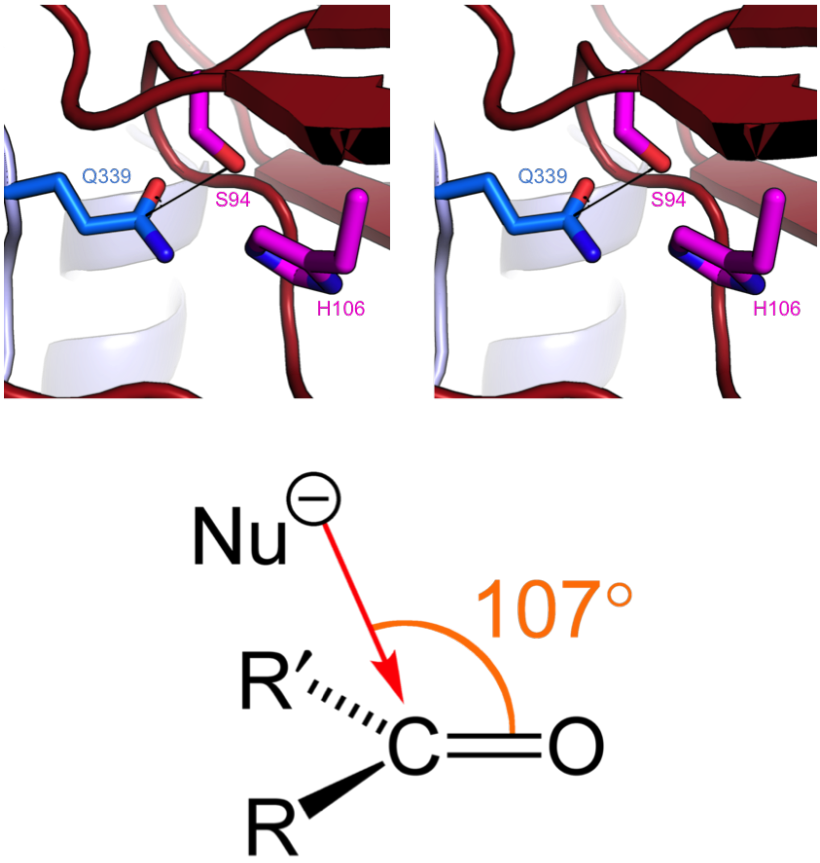


Fig. A.11 **Burgi-Dunitz angle for nucleophilic attack.** Stereodiagram of the position of the Ser94 Oγ relative to the C=O of Gln339 (top) indicating a non-ideal Burgi-Dunitz angle for nucleophilic attack. The generally accepted Burgi-Dunitz angle for a carbonyl is an obtuse angle (bottom) (Burgi et al., 1974). This suggests that Ser94 is in a Near-Attack conformation instead (Bruice., 2002).

Chapter 4

Paper 2

4.1 Summary

This manuscript describes the identification of a BLF1 homolog, SSR1, from *S. stiibiresistens* as another member of the CNF1-like family of GDTs. This is followed by an interesting observation of a mutually-exclusive peptide clipping in the monomeric form or dimerisation of the unclipped variant in solution during purification. Mass spectrometry analysis of SSR1 samples suggests that the clip site lies between residues Ser-191 and Ser-192. The clipped monomeric form of SSR1 was then shown to deamidate Gln339 of eIF4A, like BLF1. Finally, the paper describes the crystallisation and data collection of crystals of the monomeric and dimeric forms of SSR1.

4.2 Author contributions

D.W.R., M.J.D. conceived the investigations; A.A.A., C.E. designed the experiments; A.A.A., S.E.S., and C.E. conducted the experiments; A.A.A., S.E.S., C.E., M.J.D., and D.W.R. interpreted the data and A.A.A., P.J.B, and D.W.R wrote the manuscript with additional help from F.R. and G.W. Experimentally, I carried out the cloning, expression, purification and crystallisation of the monomeric and the dimeric forms of SSR1. In addition, data collection of these crystals were also carried out. I also prepared the samples for the biochemical assay with help by S.E.S and together with C.E. carried out LC-MS mass spectrometry analysis.

4.3 Paper 2

Paper 2 manuscript will start on the next page.

Identification of a glutamine deamidase toxin in *Skermanella stibiiresistens.*

Adli Aziz,[a]* Sveta Sedelnikova,[a] Caroline Evans,[b] Mark Dickman,[b] Firdaus Raih,[c] Patrick Baker,[a] Gejiao Wang[d] and David Rice[a]

[a]*Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, S10 2TN, Sheffield. England,* [b]*Department of Chemical and Biological Engineering, The University of Sheffield, Mappin Street, S1 3JD, Sheffield. England,* [c]*School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor. Malaysia, and* [d]*Institue of Agricultural Reasearch, 3425246 Wuhan, Hubei.China. E-mail: d.rice@sheffield.ac.uk*

Abstract

Skermanella stibiiresistens is a gram-negative, antimony-resistant, motile bacterium isolated from a soil sample in a coal mine in Jixi, China. Analysis of the genome of *S. stibiiresistens* has led to the identification of a homolog of the Burkholderia Lethal Factor 1 (BLF1), one of the armoury of effectors of the intracellular pathogen, *Burkholderia pseudomallei* that causes the disease melioidosis in human. BLF1 deamidates a critical glutamine residue, Gln339, in the eukaryotic initiation factor 4A (eIF4A) blocking initiation-dependent protein translation. This paper reports the cloning, purification, crystallization and preliminary experiment on SSR1

demonstrating that it shows identical toxic activity to BLF1 suggesting that *S. stibiiresistens* may be an intracellular eukaryotic pathogen.

1. Introduction

Glutamine deamidase toxins are responsible for the modification of the side chain of specific glutamine residues in target host proteins, converting them to the corresponding glutamic acid and thereby affecting their function. The first glutamine deamidase toxin to be discovered was the cytotoxic necrotizing factor 1 (CNF1), a 1014 residue protein from uropathogenic strains of *Escherichia coli*. To carry out its function, CNF1 is secreted by *E. coli* and some 700 residues at the N-terminal region are responsible for binding and translocation of the C-terminal catalytic domain (C-CNF1) harboring the toxic activity across the eukaryotic cell membrane to the cytosol (Falbo *et al.*, 1993; Lemichez *et al.*, 1997; Fabbri *et al.*, 1999). C-CNF1 deamidates a critical glutamine residue in the small GTPases including RhoA, Rac, and cdc42 (Flatau *et al.*, 1997; Schmidt *et al.*, 1997; Lerm *et al.*, 1999), inducing cytoskeleton reorganisation leading to membrane ruffling, the formation of stress fibers and multinucleated cells (Caprioli *et al.*, 1983). The crystal structure of C-CNF1 revealed the conservation of a Cys-His pair in the active site of the catalytic domain reminiscent of that seen in cysteine proteases, most notably papain, but with C-CNF1 having a completely novel fold, featuring two β-sheets surrounded by α-helices (Buetow *et al.*, 2001).

Following the discovery of glutamine deamidation activity as a potent factor in bacterial pathogenicity, further glutamine deamidase toxins having a fold similar to papain rather than C-CNF1 were identified leading to their classification as members of a family of papain-like glutamine deamidase toxins (Washington *et al.*, 2013). These toxins include but are not limited to Pasteurella multoxida toxin (PMT)

from *P. multocida*, Osp1 from *S. flexineri* and several Cycle-inhibiting factor (Cif) from a range of pathogenic bacterial strains including *E. coli*, *B. pseudomallei*, *Y. pseudotuberculosis* and *P. luminiscence* (Washington *et al.*, 2013).

Burkholderia Lethal Factor 1 (BLF1) is a potent toxin isolated from *Burkholderia pseudomallei*, a gram-negative pathogen responsible for causing the disease melioidosis in humans and mammals (Cruz-Migoni *et al.*, 2011) that is endemic in tropical and subtropical regions particularly in northern Australia and Southeast Asia (Wiersinga *et al.*, 2018). The disease itself is more widespread than previously recognised with reports of sporadic cases elsewhere including India, Bangladesh, and South America (Currie, 2008). Furthermore, melioidosis is thought to be underreported in endemic countries and, more recently, it has been estimated that there are a total of 165,000 worldwide cases per year, many of which would be fatal (Limmathurotsakul *et al.*, 2016). Structural studies on BLF1 revealed that despite sharing only 5% sequence identity to C-CNF1, it has a closely related fold (Cruz-Migoni *et al.*, 2011), leading to its classification as a member of the CNF1-like family of glutamine deamidase toxins. Comparison of these two toxins reveals that their active sites include the conservation of an LSGC motif containing the catalytic cysteine residue (Cys866 in CNF1 and Cys94 in BLF1) and the downstream histidine (His881 in CNF1 and His106 in BLF1) that make up a thiolate-imidazolium ion pair thought to be involved in the site specific deamidation.

Biochemical studies have identified that the target for BLF1 is the deamidation of Gln339 of human eukaryotic initiation factor 4A (eIF4A), blocking its helicase function thus inhibiting initiation-dependant protein translation (Cruz-Migoni *et al.*, 2011). However, BLF1 is based on a much smaller subunit, with only 211 residues compared to CNF1 with the similarity between them being limited to the catalytic domain of the latter. Thus, BLF1 lacks the cell binding and membrane

incorporation domains of CNF1, a difference that can be explained as *B. pseudomallei* is an intracellular pathogen, and hence there is no requirement for BLF1 toxin to cross the eukaryotic cell membrane.

S. stibiiresistens is a gram-negative, rod-shaped, motile bacterium that was isolated from a coal mine in Jixi, China (Luo *et al.*, 2012). A particular phenotype of this bacterium is its resistance to antimony (Sb), a Group 15 element with metalloid properties similar to those of arsenic (As) that is generally regarded as being toxic (Luo *et al.*, 2012). The completion of the genome sequence of *S. stibiiresistens* (Zhu *et al.*, 2014) revealed that it encodes many proteins involved in cell metabolism and motility but did not identify any toxins in this organism. In this paper, we describe the identification of a homolog of BLF1 in *S. stibiiresistens*, N825_12680, the cloning of the gene, and overexpression, purification, characterisation and crystallization of the protein, which will hereinafter be referred to as, SSR1. Since *S. stibiiresistens* has not yet been identified as a pathogen, and given the close similarity of SSR1 to BLF1, including the absence of the cell binding and membrane spanning domains seen in CNF1, we propose that *S. stibiiresistens* might be an intracellular eukaryotic pathogen.

2. Materials and Methods.

2.1. Macromolecule production.

The SSR1 gene was cloned and amplified via PCR from gDNA of *S. stibiiresistens* strain SB22^{T} (Luo *et al.*, 2012), using the appropriate primer pairs as shown in **Table 1** to produce an untagged and a C-terminal His_6-tagged constructs. The amplified DNA fragments were ligated into pET24a plasmid encoding a C-terminal His_6 tag, followed by transformation of *E.coli* DH5α strain. The positive clones were identified by colony PCR and the sequences of the plasmids were verified against

Table 1. ***Construct information.***

Construct	C-terminal His_6-tagged SSR1	Untagged SSR1
Tag details	....LEHHHHHH	-
Forward primer	GTTTTGTAAC**CATATG**CCGACACC GATCCAGGACAAACTCC	GTTTTGTAAC**CATATG**CCGACACCG ATCCAGGACAAACTCC
Reverse Primer	CTATTAA**CTCGAG**GTTCTTCCGCG TCTGCCTGGTGTAG	CTATTAA**CTCGAG**TCAGTTCTTCCGC GTCTGCCTGGTG
Restriction enzymes	Nde1 and Xho1	Nde1 and Xho1
Expression vector	pET24a (Kan^R marker)	pET24a (Kan^R marker)
Expression host	*E. coli* BL21 (DE3)	*E. coli* BL21 (DE3)

the NCBI database.

For protein overexpression, purified plasmids were transformed into *E. coli* BL21 (DE3) cells, from which a single colony was inoculated into a starter culture containing 5 ml of LB supplemented with 50 μg ml^{-1} kanamycin and grown overnight at 310 K with shaking at 250 rev min^{-1}. The starter culture was then inoculated into 500 ml LB medium with kanamycin as above. The cells were grown in a similar fashion to reach an OD_{600} of 0.6, at which they were harvested by centrifugation at 4000 rpm for 35 min using Beckman J-6B centrifuge and the LB medium was decanted. The cell pellet was then resuspended and washed in M9 minimal medium (6 g l^{-1} Na_2HPO_4, 3 g l^{-1} KH_2PO_4, 0.5 g l^{-1} NaCl, pH set to 7.4) twice before adding the suspension to 500 ml of M9 minimal medium supplemented with 325 μl trace elements stock solution **(Table S1)**, 7.5 ml 20 % (w/v) glucose, 50 μl thiamine (10 mg ml^{-1}), 1 ml NH_4Cl (0.3 g ml^{-1}), 0.5 ml of 1 M $MgSO_4$, 500 μl Kanamycin and 50 μl 1 M $CaCl_2$. The cells were grown at 310 K with shaking until OD_{600} reached 1.0, whereupon, the culture was incubated on ice for 15 minutes with intermittent shaking. Overexpression was induced by the addition of 1 mM IPTG and the cells were grown at 285 K for 72 h.

2.2. Purification.

For purification of the untagged SSR1, the cell pellet was resuspended in a buffer containing 50 mM Tris-HCl pH 8.0 and disrupted by sonication. The cell lysate was clarified by centrifugation before being applied to DEAE-Sepharose Fast Flow column (GE Healthcare) equilibrated with the same buffer. Proteins were eluted with a linear gradient of 0-0.5 M NaCl in the same buffer, and SSR1 eluted at approximately 0.12 M NaCl. Fractions included in the elution peak of the protein were combined and concentrated to 2 ml volume using a VivaSpin 20 concentrator

Table S1. ***Components for making the stock solution of trace elements.*** *To make 100 ml, the components above were dissolved into 70 ml distilled water. The pH was then adjusted to pH 8.0 before adding 500 mg EDTA. The pH was then readjusted to pH 8.0 followed by the addition of 375 mg of $FeSO_4.7H_2O$. Water was added to make it up to 100 ml after which the stock solution was autoclaved.*

Chemical components	**Amount (mg)**
$CaCl_2.2H_2O$	550
$MnSO_4.H_2O$	140
$CuSO_4.5H_2O$	40
$ZnSO_4.7H_2O$	220
$CoCl_2.6H_2O$	45
$Na_2MoO_4.2H_2O$	26
H_3BO_4	40
KI	26

with a MW cutoff of 10 000 Da. The concentrated sample was then applied to the Superdex 200 size exclusion chromatography column (GE healthcare) equilibrated with 50 mM Tris pH8, 0.5 M NaCl and eluted with the same buffer. The protein was buffer exchanged into 10 mM Tris-HCl pH8 and concentrated using a VivaSpin 20 concentrator fitted with a diafiltration cup.

For purification of the C-terminal His_6-tagged SSR1, cell pellet was resuspended in a buffer containing 50 mM Tris-HCl pH 8.0, 0.5 M NaCl and disrupted by sonication. The cell lysate was clarified by centrifugation before being applied to HisTrap-HP 5 ml (GE Healthcare) equilibrated with the same buffer. Proteins were eluted with a linear gradient of 0-0.5 M Imidazole in the same buffer, and SSR1 eluted at approximately 0.1 M Imidazole. Fractions containing C-terminal His_6-tagged SSR1 were then combined and subjected to size exclusion chromatography as detailed above.

2.3. Crystallization.

Preliminary crystallization trials were carried out using the sitting-drop vapour diffusion protocol, using Matrix Hydra II (Thermo Fisher Scientific, USA) and a variety of 96-well format screens including PACT, PHClear, and MPD. 100 nl of concentrated protein samples (as detailed above) were added to an equal volume of the precipitant and the drop was equilibrated against 50 μl of the precipitant at 290 K. Initial crystal hits were then optimized using the hanging-drop vapour-diffusion technique **(Table 2)**. For data collection, the crystals were cryoprotected in the same precipitant and flash-cooled at 100 K. Data were collected on beamline I03 at the Diamond Light Source synchrotron with strategies and processing statistics highlighted in **Table 3**.

Table 2. ***Crystallisation conditions for SSR1 WT crystals.***

Construct	**Untagged SSR1**		**C-terminal His_6-tagged SSR1**
Oligomeric form	Monomer	Monomer	Dimer
Buffer solution	10 mM Tris-HCL pH8	10 mM Tris-HCL pH8	50 mM Tris-HCL pH8, 0.5 M NaCl
Method	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion	Sitting-drop vapour diffusion
Reservoir components	0.25 M zinc sulfate, 41 % (w/v) MPD	0.1 M HEPES pH 6.5, 65 % (w/v) MPD	0.1 M Bis-Tris Propane pH 6.5, 0.2 M sodium formate, 20 % (w/v) PEG 3350
Volume (protein:precipitant)	1 µl : 1 µl	1 µl : 1 µl	100 nl : 100 nl
Volume of reservoir	500 µl	500 µl	50 µl

Table 3. ***Data collection and processing statistics.***

Construct	**Untagged SSR1 monomer (Crystal form 1)**	**Untagged SSR1 monomer (Crystal form 2)**	**C-terminal His_6-tagged SSR1 dimer**
Beamline	I03	I03	I03
Wavelength (°)	0.97629	0.97629	0.97629
Detector	Pilatus3 6M	Pilatus3 6M	Pilatus3 6M
Exposure time per image (s)	0.1	0.1	0.05
Rotation range per image (°)	0.2	0.2	0.1
Total rotation range (°)	200	200	200
Data-processing package	xia2-dials	xia2-dials	xia2-dials
Space group	$P3_121$ / $P3_221$	$P3_121$ / $P3_221$	$P2_12_12_1$
a, b, c (Å)	69.7, 69.7, 104.5	71.1, 71.1, 96.2	39.2, 83.3, 115.8
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution range (Å)	60.33-1.18 (1.21-1.18)	61.61-1.61 (1.65-1.61)	83.31-2.02 (2.07-2.02)
R_{merge}♩	0.173 (2.937)	0.046 (1.704)	0.214 (1.383)
$\langle I/\sigma(I)\rangle$	7.9 (1.1)	17.7 (1.3)	5.3 (1.4)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Multiplicity	10.0 (9.5)	10.4 (10.4)	7.0 (7.2)
Total reflections	968990	384125	180150
Unique reflections	96928	37108	25760

♩ $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_{i(hkl),}$ where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed intensities and the mean intensity of related reflections, respectively.

2.4. Biochemical assay.

LC ESI MS analysis was performed as previously described (Cruz-Migoni *et al.*, 2011). In brief following in gel tryptic digestion peptides were analysed using LC ESI MS analysis using a maXis UHR TOF mass spectrometer (Bruker Daltonics) using an automated acquisition approach. MS and MS/MS scans (m/z 50-3000) were acquired in positive ion mode. Lock mass calibration was performed using HP 1221.990364. Line spectra data was then processed into peak list by Data analysis (Bruker Daltonics) using the following settings. The sum peak finder algorithm was used for peak detection using a signal to noise (S/N) ratio of 10, a relative to base peak intensity of 0.1% and an absolute intensity threshold of 100. Extracted ion chromatograms were generated for both the non deamidated and deamidated peptide in Data analysis. MS/MS spectra of the peptides identified were manually verified.

3. Results and Discussion.

3.1. Identification of BLF1 homolog in the genome of S. stibiiresistens.

A BLASTp search of the genome of *S. stibiiresistens* led to the identification of a 206 amino-acid protein homologous to BLF1, SSR1, with 55 % sequence identity including the conservation of the LSGC motif and the associated catalytic histidine that form the critical components of BLF1 activity **(Figure 1)** (Cruz-Migoni *et al.*, 2011). The alignment of their sequences indicated that there were 3 positions of minor insertions or deletions between their sequences in loop regions of BLF1, together with a major 6 residue insertion in the region spanning β13-14 in BLF1 compared to SSR1 **(Figure 1)**.

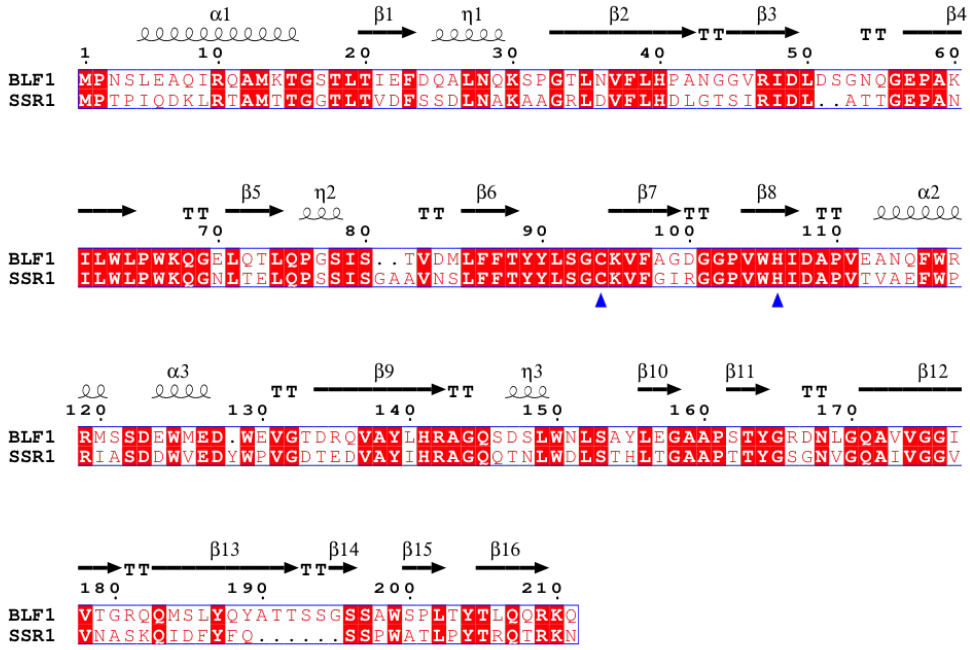


Fig. 1. **Sequence alignment between BLF1 and SSR1.** Top row indicates the secondary structure in the fold of BLF1 based on an atomic resolution crystal structure (PDB:3TU8). SSR1 shares 55 % sequence identity to BLF1 conserving the LSGC motif containing the catalytic cysteine, Cys94 (blue triangle). The associated histidine residue in the dyad, His106 (blue triangle), is also conserved in SSR1.

3.2. SSR1 over-production produced clipped proteins.

Following purification with a nickel affinity column chromatography of the C-terminal His_6-tagged SSR1, SDS-PAGE analysis indicated that a significant fraction of the 24 kDa protein (80%) had been clipped into 21 kDa and 3 kDa fragments which remain associated during purification **(Figure 2b)**. Subsequent purification via size-exclusion chromatography, resolved the protein into a major peak running with an apparent MW of 24 kDa consistent with a "monomeric" form of SSR1, and a minor peak running with an apparent MW of an SSR1 "dimer" **(Figure 2a)**. SDS PAGE analysis of the major peak indicated that the "monomeric" peak consists mainly of the cleaved protein, whereas the "dimeric" peak represents the full length protein **(Figure 2b)**. Similar results were obtained following analysis of untagged SSR1 (data not shown).

3.3. Mass spectrometry analysis of SSR1 samples.

Mass spectrometry analysis on the "monomeric" sample of C-terminal His_6-tagged SSR1 showed that it was composed of peptides with MW 20578 Da and 2884 Da compared to the untagged SSR1 protein sample that was composed of peptides with MW 20578 Da and 1819 Da **(Figure 3a & 3b)**. This is consistent with the larger fragment representing the N-terminal portion of the protein up to Ser191 but with the removal of the N-terminal methionine and the shorter fragment representing residues 192 to the C-terminus.

3.4. Preliminary biochemical assay indicates that SSR1 is a glutamine deamidase toxin.

5 μl of 0.072 μM "monomeric" SSR1 was incubated at room temperature with 90 μl of 40 μM human eIF4A that had been produced and purified as previously described (Mobbs et al., 2018, submitted) to give a ratio of 1:10000 of SSR1:eIF4A

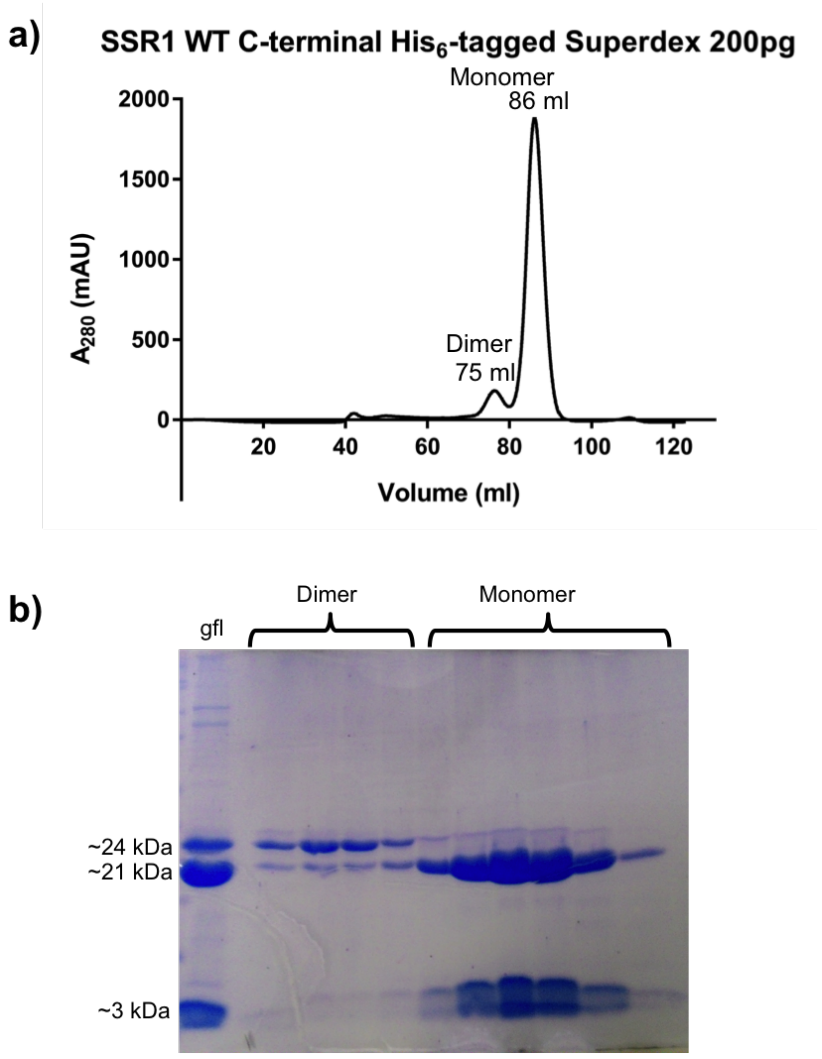


Fig. 2. **Purification of SSR1 proteins.** (a) Gel filtration chromatogram of the C-terminal His_6-tagged SSR1 showing the UV absorption at 280 nm (A_{280}) against elution volume. The elution volumes corresponding to the monomer and the dimer form of SSR1 are shown above the curve. (b) SDS-PAGE gel showing the dimer and monomer fractions from the gel filtration run of the C-terminal His6 tagged SSR1. The gel filtration load (gfl) after Nickel affinity column showed the presence of the full-length and the clipped SSR1.

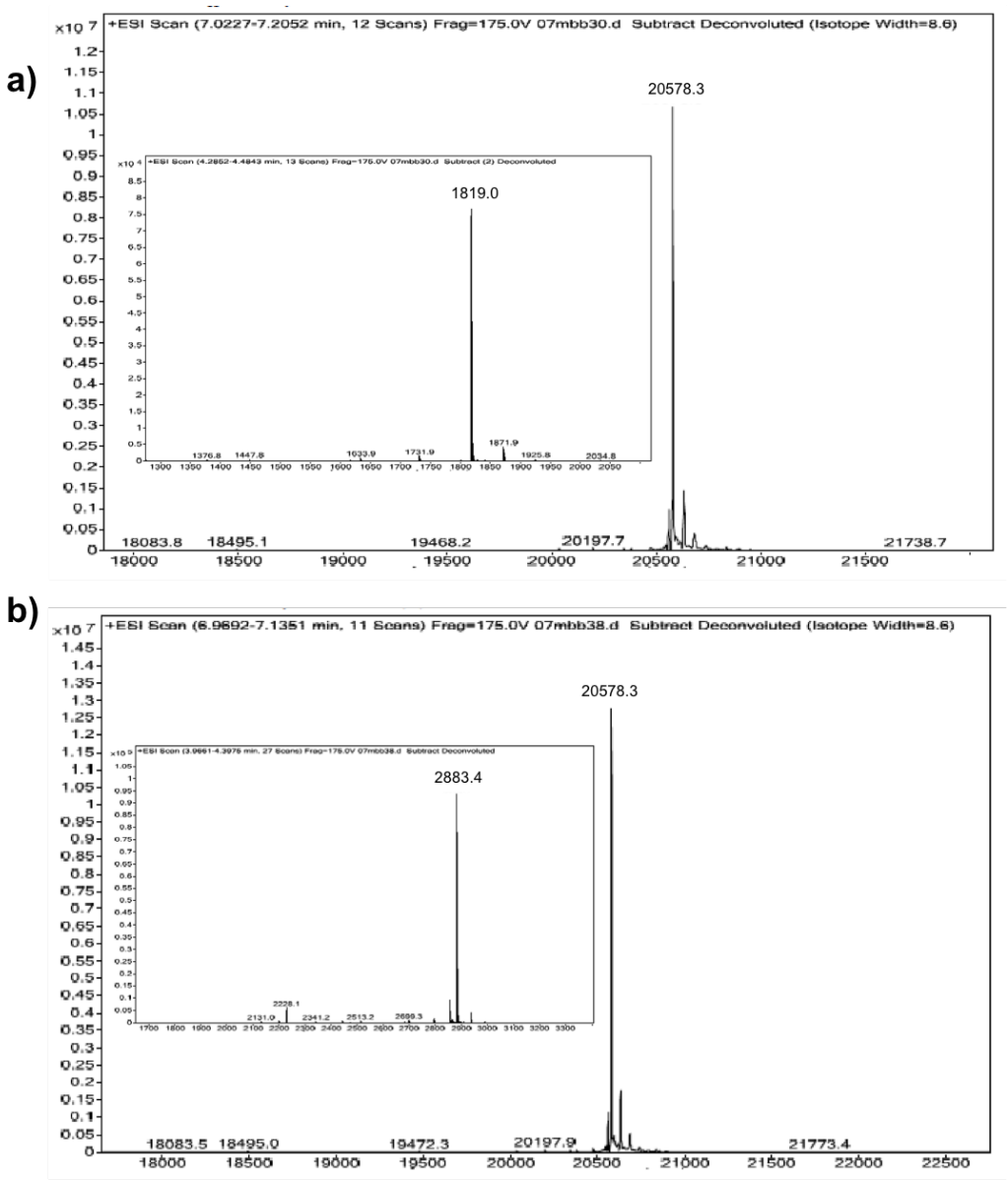


Fig. 3. **Mass spectrometry analysis of purified samples of untagged and tagged SSR1 monomer.** Mass Spectrometry analysis of (a) the untagged monomeric SSR1 showing peaks of 20578 and 1819 (inset) Daltons and (b) the C-terminal His_6-tagged SSR1 with peaks of 20578 and 2883 (inset) Daltons.

and analysed at different time-points with LC-ESI-MS as previously described (Cruz-Migoni *et al.*, 2011). Analysis of the eIF4A peptides revealed a significant level of Gln339 deamidation after 5 minutes incubation (60 % complete) and with complete deamidation after 30 minutes **(Figure 4)**.

3.5. Monomeric SSR1 crystals and diffraction data analysis.

Crystallisation trials on the pooled samples of the untagged SSR1 corresponding to the "monomeric" peak yielded two morphologically distinct crystal forms. These were optimised to give final conditions of 0.25 M $ZnSO_4$, and 41 % MPD (Form 1) and 0.1M HEPES pH 7.5, 65 % MPD (Form 2), producing crystals with maximum dimensions of up to 0.1 mm **(Figure 5a & 5b)**. Crystals belonging to both form 1 and form 2 were then subjected to X-ray data collection at DLS, resulting in diffraction to 1.2 Å and 1.6 Å respectively. The autoprocessing pipeline provided in the ISPyB SynchWeb database (Winter & McAuley, 2011; Fisher *et al.*, 2015) showed that both crystal forms belong to one of the two enantiomorphic pair of space groups $P3_121$ or $P3_221$ with different cell dimensions but in each case with one molecule in the asymmetric unit with V_m values of 3.26 and 3.12 for form 1 and form 2, respectively **(Table 3)**. Mass Spectrometry analysis of old crystals from both crystal forms showed that they both contain an identical C-terminal peptide as seen earlier but suggested that the N-terminal peptide has been modified probably due to oxidation of the cysteine residue **(Figure 6a & 6b)**.

3.6. Dimeric SSR1 crystals data analysis.

Crystals resulting from the trials on the pooled samples of the "dimeric" form of SSR1 were found in PACT F6 (0.2 M sodium formate, 0.1 M Bis tris propane pH 6.5, 20 % PEG 3350) **(Figure 5c)**. For data collection, these crystals were cryoprotected

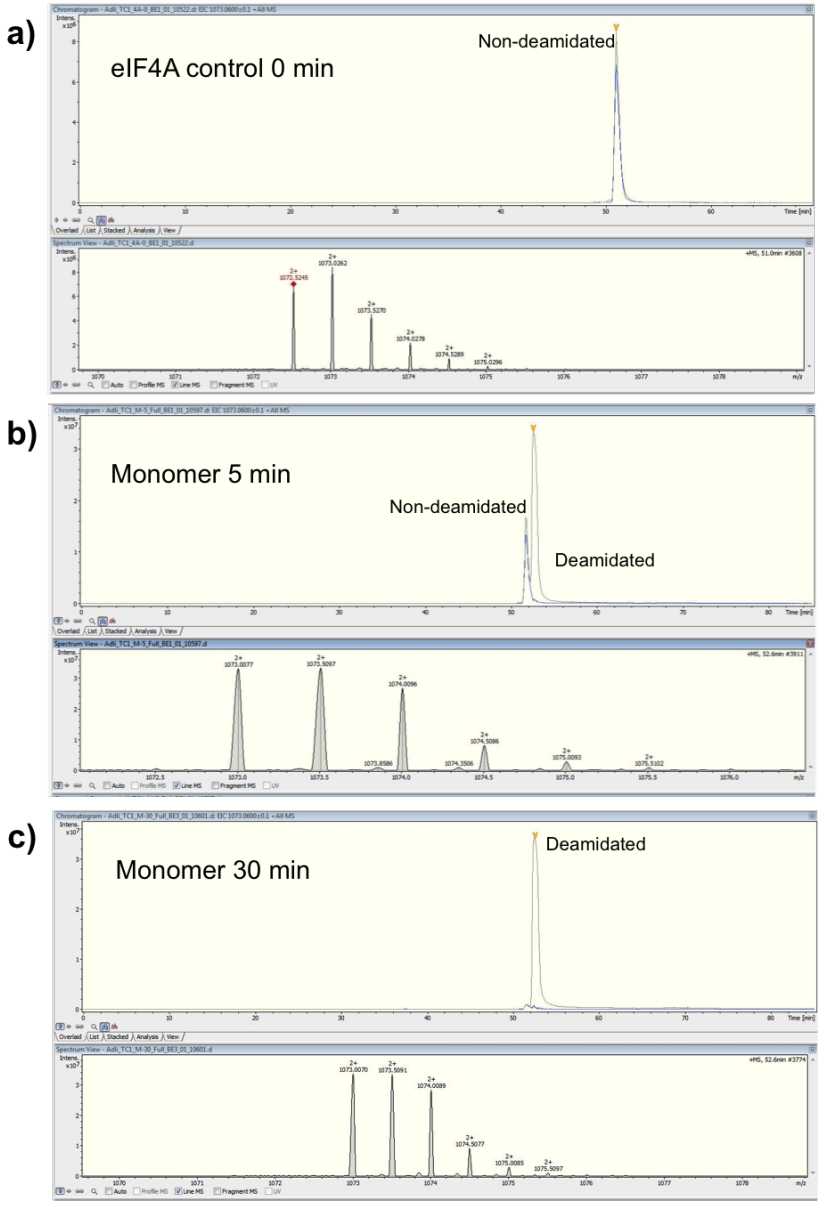


Fig. 4. **LC MS analysis of the Gln339 containing tryptic peptide from eIF4A.** Extracted ion chromatogram of the tryptic peptide GIDVEQVSLVINY-DLPTNR $[M+3H]^{3+}$ (top) from (a) eIF4A control at time 0 min, (b) SSR1 monomer+eIF4A following 5 mins incubation, revealing 60 % deamidation, and (c) SSR1 monomer+eIF4A following 30 mins incubation, revealing complete deamidation. MS1 spectra are shown underneath.

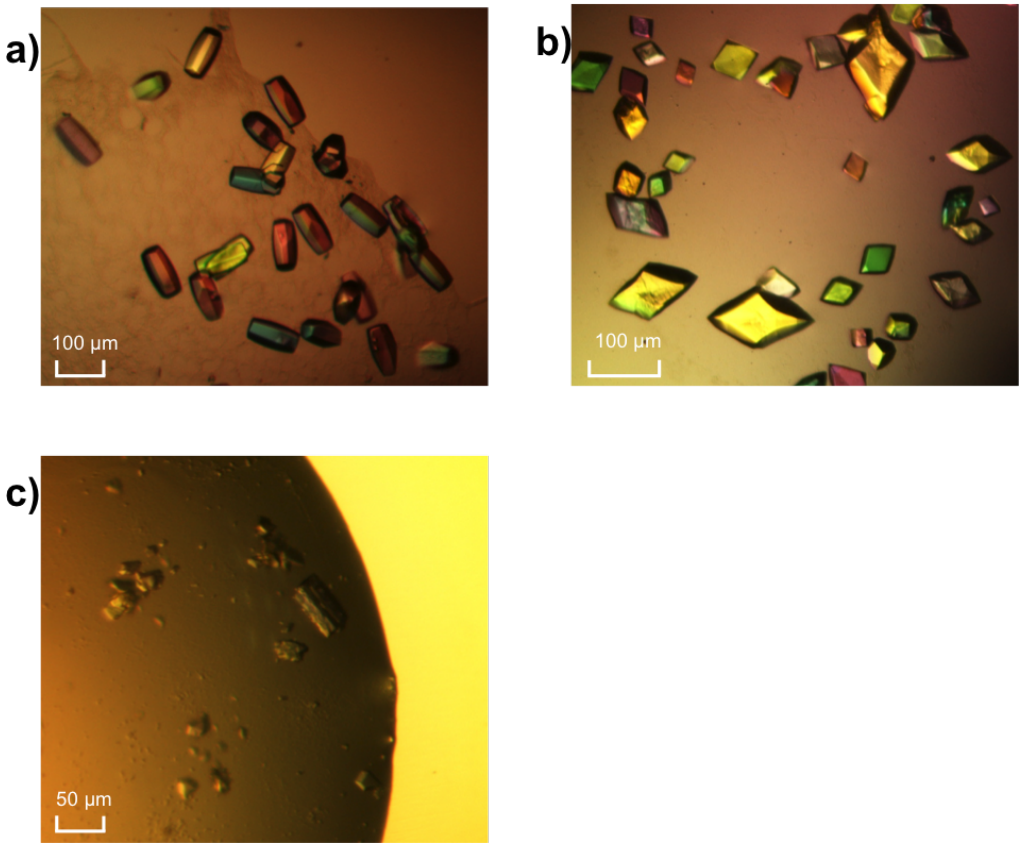


Fig. 5. **Crystals of SSR1 proteins.** Crystals of monomeric untagged SSR1 were grown in (a) 0.25 M zinc sulfate and 41 % (v/v) MPD and (b) grown in 0.1 M HEPES pH 7.5 and 65 % (v/v) MPD. (c) Crystals of dimeric C-terminal His_6-tagged SSR1 were grown in 0.2 M sodium formate, 0.1 M bis-tris propane pH 6.5 and 20 % (w/v) PEG 3350.

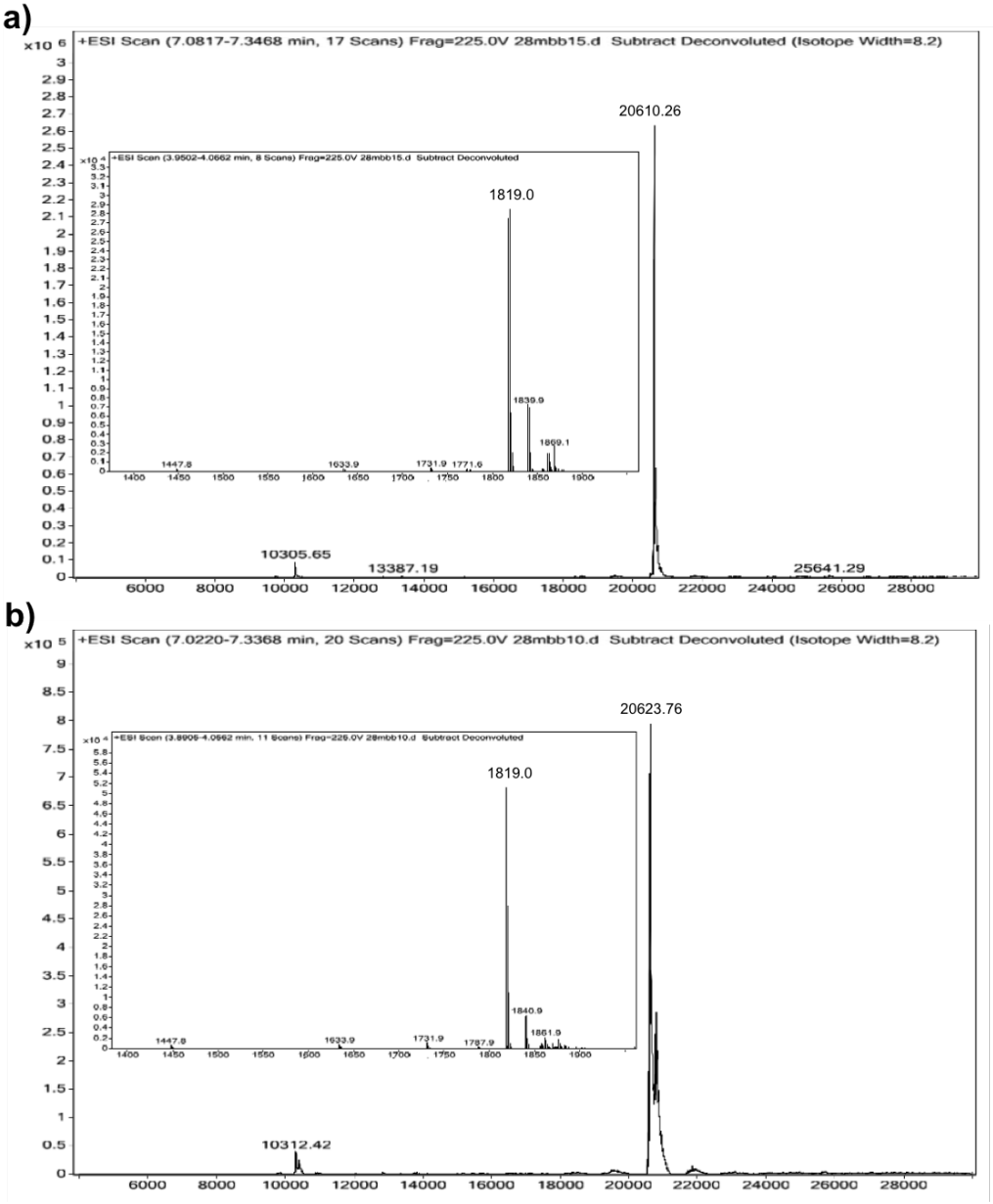


Fig. 6. **Mass spectrometry analysis on SSR1 WT crystals.** Mass spectrometry analysis showing the MW of the polypeptides in the crystals of the untagged SSR1 of crystal form 1 (a) and crystal form 2 (b). The deviation of the MW of the larger fragment in both crystal forms could have originated from oxidation of a cysteine residue either its doubly oxidised (+32 Da) or triply oxidised (+48) in crystal form 1 and 2 respectively. The corresponding shorter fragment for both crystal forms are shown inset.

in the same precipitant with an addition of 25 % (v/v) ethylene glycol and flash-cooled at 100 K. The crystals of the "dimeric" form of SSR1 diffracted to 2.2 Å and autoprocessing with xia2-dials showed that they belong to the orthorhombic space group $P2_12_12_1$ with cell dimension 39.2 x 83.3 x 115.8 Å. Consideration of possible values of the Matthews coefficient indicated that the crystals most likely contain 2 polypeptide chains in the asymmetric unit with a solvent content of 43 % and a V_m of 2.15.

4. Conclusion.

We have identified that a homolog of BLF1 in the genome of *S. stibiiresistens*, SSR1, and shown that it acts as a glutamine deamidase toxin targeting the site-specific demidation of Gln339 of human eIF4A. Given the similarities in the architecture of BLF1 and SSR1, namely their lack of the N-terminal region that is responsible for attachment and transport across host cell membrane compared to that seen in CNF1, we postulate that *S. stibiiresistens*, like *B. pseudomallei*, may be an intracellular eukaryotic pathogen. We further note that the overexpression of SSR1 results in the production of clipped and full length variants of the protein that behave quite differently on gel filtration, suggesting that they represent the "monomeric" and the "dimeric" forms of the protein that crystallize in distinct crystal forms. Structure determination of crystal forms 1 and 2 of the "monomeric" SSR1 as well as the "dimeric" variant is underway to identify the consequences of the cleavage and its implication for the folding and function of SSR1 and hence to explore the potential pathogenicity of the microbe.

5. Acknowledgements

A.A.A would like to thank Majlis Amanah Rakyat (MARA) Malaysia for the PhD scholarship.

References

Buetow, L., Flatau, G., Chiu, K., Boquet, P. & Ghosh, P. (2001). *Nature structural biology*, **8**(7), 584–588.

Caprioli, A., Falbo, V., Roda, L. G., Ruggeri, F. M. & Zona, C. (1983). *Infect. Immun.* **39**(3), 1300–1306.

Cruz-Migoni, a., Hautbergue, G. M., Artymiuk, P. J., Baker, P. J., Bokori-Brown, M., Chang, C.-T., Dickman, M. J., Essex-Lopresti, A., Harding, S. V., Mahadi, N. M., Marshall, L. E., Mobbs, G. W., Mohamed, R., Nathan, S., Ngugi, S. a., Ong, C., Ooi, W. F., Partridge, L. J., Phillips, H. L., Raih, M. F., Ruzheinikov, S., Sarkar-Tyson, M., Sedelnikova, S. E., Smither, S. J., Tan, P., Titball, R. W., Wilson, S. a. & Rice, D. W. (2011). *Science*, **334**(6057), 821–824.

Currie, B. J. (2008). *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **102**(3), 225–227.

Fabbri, A., Gauthier, M. & Boquet, P. (1999). *Molecular Microbiology*, **33**(1), 108–118.

Falbo, V., Pace, T., Picci, L., Pizzi, E. & Caprioli, A. (1993). *Infection and immunity*, **61**(11), 4909–14.

Fisher, S. J., Levik, K. E., Williams, M. A., Ashton, A. W. & McAuley, K. E. (2015). *Journal of applied crystallography*, **48**(Pt 3), 927–932.

Flatau, G., Lemichez, E., Gauthier, M., Chardin, P., Paris, S., Fiorentini, C. & Boquet, P. (1997). *Nature*, **387**(6634), 729–733.

Lemichez, E., Flatau, G., Bruzzone, M., Boquet, P. & Gauthier, M. (1997). *Molecular microbiology*, **24**(5), 1061–70.

Lerm, M., Selzer, J., Hoffmeyer, A., Rapp, U. R., Aktories, K. & Schmidt, G. (1999). *Infection and immunity*, **67**(2), 496–503.

Limmathurotsakul, D., Golding, N., Dance, D. A. B., Messina, J. P., Pigott, D. M., Moyes, C. L., Rolim, D. B., Bertherat, E., Day, N. P. J., Peacock, S. J. & Hay, S. I. (2016). *Nature Microbiology*, **1**(1), 15008.

Luo, G., Shi, Z., Wang, H. & Wang, G. (2012). *International Journal of Systematic and Evolutionary Microbiology*, **62**(6), 1271–1276.

Schmidt, G., Sehr, P., Wilm, M., Selzer, J., Mann, M. & Aktories, K. (1997). *Nature*, **387**(6634), 725–729.

Washington, E. J., Banfield, M. J. & Dangl, J. L. (2013). *Microbiology and molecular biology reviews : MMBR*, **77**(3), 527–39.

Wiersinga, W. J., Virk, H. S., Torres, A. G., Currie, B. J., Peacock, S. J., Dance, D. A. B. & Limmathurotsakul, D. (2018). *Nature Reviews Disease Primers*, **4**, 17107.

Winter, G. & McAuley, K. E. (2011). *Methods*, **55**(1), 81–93.

Zhu, W., Huang, J., Li, M., Li, X. & Wang, G. (2014). *Standards in Genomic Sciences*, **9**(3), 1211–1220.

4.4 Appendix

The extended method section encompassing the work that are not described in Paper 2 will be discussed in detail in Appendix B (page 171)

Chapter 5

Paper 3

5.1 Summary

This manuscript describes the structure determination of the monomeric and dimeric SSR1 WT, revealing the clip site on the former between Ser191-Ser192 located on a loop, but an unclipped domain-swapped dimer in the latter with those residues lying in a linear conformation, suggestive of a strain during folding. The paper then describes the investigations carried out to understand the reasons underlying the mutually-exclusive events of peptide clipping and domain-swapping which include crystallisation and structure determination of various mutants of SSR1. In light of the similarity between glutamine deamidation and peptide cleavage, a C94S mutation which inactivates the glutamine deamidase activity still resulted in clipping suggesting that the toxin activity plays no role in clipping. Subsequent analysis on the behaviour and crystal structures of a series of clip site mutants revealed that the origin of the clipping and domain-swapping lie in a strain between Ser192 and Pro193 in the monomeric fold, which would be resolved by either domain swapping or clipping. The clipping, however, requires a nucleophile at position 192 (serine/cysteine/threonine) which seems to be related to the mechanism of self-splicing inteins.

5.2 Author contributions

D.W.R., conceived the investigations; A.A.A., designed the experiments; A.A.A., and S.E.S., conducted the experiments; A.A.A., S.E.S., and D.W.R. interpreted the data and A.A.A., P.J.B, and D.W.R wrote the manuscript with additional help from F.R. and G.W. Experimentally, I carried out the site-directed mutagenesis, expression, purification and crystallisation of various SSR1 constructs, which were followed by data collection, structure determination, model building and refinement including every aspects of the analysis.

5.3 Paper 3

Paper 3 manuscript will start on the next page.

Strain-induced clipping: a novel form of post-translational modification

Adli Aziz[1], Firdaus Raih[4], Svetlana Sedelnikova[1], Patrick Baker[1], Gejiao Wang[3], and David Rice[1,*]

[1]Department of Molecular Biology and Biotechnology, The University of Sheffield, Western Bank, S10 2TN, Sheffield, England.

[3]Institute of Agricultural Research, 3425246 Wuhan, Hubei, China.

[4]School of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

[*]d.rice@sheffield.ac.uk

ABSTRACT

We have determined crystal structures of a glutamine deamidase toxin from *Skermanella stiibiresistens*, SSR1, that is homologous to the Burkholderia Lethal Factor 1 (BLF1). The structure reveals that it either folds into a monomer, clipped between residues serine-191 and serine-192 where the two fragments remain associated as a single domain, or folds into an unclipped domain-swapped dimer. Investigations on the origins of the clipping/domain-swapping suggest that the peptide bond between the two serine residues is strained during folding of the monomeric SSR1 facilitating clipping which also demands the presence a Ser/Cys/Thr residue at position 192, a manner possibly related to self-splicing inteins. A further consequence of the strain is to stall the folding allowing the formation of unstrained domain-swapped dimer.

Introduction

Glutamine deamidase toxins (GDTs) are effectors employed by pathogenic bacteria to cause damage in host cells by deamidating critical glutamine residues in host target proteins thereby affecting their normal function which ultimately results in disease. GDTs can be divided into two families based on their tertiary fold, one of which is like papain and the other being similar to the Cytotoxic Necrotizing Factor 1 (CNF1) (Washington et al., 2013)[1]. The papain-like family of GDTs fold into a globular domain consisting a single core β-sheet surrounded by α-helices and include toxins such as Osp1 from *Shigella flexineri* (Sanada et al., 2012)[2], C3 domain of Pmt from *Pasteurella multocida* (Kitadokoro et al., 2007)[3], and Cif from a number of pathogens including *Escherichia coli*, *Burkholderia pseudomallei*, *Photorhabdus luminiscence*, and *Yersinia pseudotuberculosis* (Yao et al., 2012, Crow et al., 2012)[4,5]. The CNF1-like family of GDTs however consists of two β-sheets surrounded by α-helices, but the family is currently limited to CNF1 and its close homologs in *E. coli* and *Y. pseudotuberculosis* (Lockman et al., 2002, Buetow et al., 2001)[6,7], and to the Burkholderia Lethal Factor 1 (BLF1) from *B. pseudomallei* (Cruz-Migoni et al., 2011)[8]. Catalysis by both of these families appears to involve a thiolate-imidazolium ion pair in a mechanism that, despite the radical difference in their folds, has been suggested to be related to the cysteine proteases (Washington et al., 2013, Mobbs et al., 2018, submitted)[1].

BLF1 was identified as a potent toxin in the pathogenicity of *B. pseudomallei*, a gram-negative, soil-dwelling, multiple drug resistant bacterium that is the causative agent of melioidosis, a disease which is endemic in northern Australia and southeast Asia (Wiersinga et al., 2018)[9]. Recently, melioidosis has been suggested to be under-reported in endemic regions and it has been estimated that there may be over 150000 cases worldwide each year, nearly half of which may well be fatal (Limmathurotsakul et al., 2016)[10]. Furthermore, the disease is now thought to be more widespread with cases reported in the Indian subcontinent, Europe,

Africa and South America (Currie et., al 2008)[11]. Concerns surrounding the difficulties in diagnosis, the variety of clinical manifestations and the lack of effective treatments have led to a call for new methods to control infections caused by this pathogen (Wiersinga et al., 2018)[9]. Structural and biochemical studies on BLF1 have identified that its molecular target is the deamidation of a critical glutamine residue, Gln339, in human eukaryotic initiation factor 4A (eIF4A) (Cruz-Migoni et al., 2011)[8]. This has recently led to an investigation of the molecular basis of substrate specificity and the mechanism of action of BLF1 on eIF4A which has been shown to be closely related to papain utilising a thiolate-imidazolium ion pair (Mobbs et al., 2018, submitted).

Bioinformatic studies have identified a homolog of BLF1, referred to here as SSR1, in the genome of *S. stibiiresistens*, an antimony-resistant, rod-shaped, gram-negative bacterium isolated from a coal mine in Jixi, China (Luo et al., 2012, Aziz et al., 2018, submitted)[12]. SSR1 shares 55 % sequence identity to BLF1, including the conservation of the residues which, in BLF1, form the Cys-His thiolate imidazolium ion pair, and has been shown to target the deamidation of Gln339 on eIF4A like BLF1 (Aziz et al., 2018, submitted). However, unlike BLF1, which appears to be strictly monomeric, purification of recombinant SSR1 indicated a mixture of a full length protein and a clipped variant that behave differently on gel filtration possibly representing a dimer and a monomer in solution, respectively (Aziz et al., 2018, submitted). Mass spectrometry analysis has suggested that the clip site in SSR1 is the peptide bond between residues Ser191 and Ser192 (Aziz et al., 2018, submitted). In this paper, we report the crystal structures of the full-length SSR1 dimer and the clipped monomer to understand the molecular basis of the catalytic properties of the SSR1 toxin. This study has provided insights into the features of the sequence that direct the folding of the protein into a domain-swapped dimer or a monomer and that control the clipping.

Results and Discussion

Structure determination of monomeric SSR1 in two crystal forms. SSR1 was purified and crystallised in two distinct crystal forms (form 1 and form 2) as previously described (Aziz et., al 2018 submitted). Crystals of form 1 of SSR1, which were grown in the presence of Zn^{2+}, belong to space group $P3_221$ and data were collected to 1.1 Å. The crystals contain one molecule in the asymmetric unit and the structure was solved by zinc single anomalous dispersion (Zn-SAD) revealing 11 Zn^{2+} sites. The structure of crystal form 2 of SSR1, which belongs to space group $P3_121$, was then determined by Molecular Replacement using coordinates of crystal form 1 as a search model and refined to 1.6 Å. The form 2 crystals were grown from conditions without zinc and, consistent with this, no metal ions were seen in the electron density map supporting the view that the zinc ions play no part in the deamidase activity of this toxin. Full details of data collection and processing are highlighted in **Table 1**.

Overall structure of SSR1 monomer. In both crystal forms 1 and 2, SSR1 is a monomer and the two structures are essentially identical (rmsd 0.40 Å for 202 equivalenced C_α carbons). The architecture of SSR1 is a single globular domain consisting of two stacks of antiparallel β-sheets with a total of 13 β-strands surrounded by 6 α-helices **(Figure 1)**. Given the high sequence similarity to BLF1 (55%), superposition of the coordinates of SSR1 with BLF1 (PDB:3TU8) indicates that their folds are closely related (rmsd 0.92 Å over 198 equivalenced C_α carbons) **(Figure 1)**. This similarity is particularly strong at their respective active sites including the LSGC motif containing the nucleophilic cysteine (Cys94) and the associated histidine (His106) which form the catalytic machinery of the toxin, and the flanking residues that form important elements of the specificity pocket which targets the deamidation of Gln339 in eIF4A **(Figure 2)** (Mobbs et al., 2018 submitted, Aziz et al., 2018, submitted).

Consistent with mass spectrometry analysis (Aziz et al., 2018, submitted), the crystal structure of the monomeric SSR1 in both crystal forms clearly showed that the polypeptide

Table 1. Data collection and refinement statistics

Construct	SSR1 WT monomer (Crystal form 1)	SSR1 WT monomer (Crystal form 2)	SSR1 WT dimer	SSR1 C94S monomer	SSR1 C94S dimer	SSR1 S191A-S192G monomer	SSR1 S191A-S192A monomer	SSR1 S192A monomer	SSR1 S192C monomer
Data Collection§									
Beamline	I03	I03	I03	I04	I04	I03	I03	I03	I03
Wavelength (Å)	0.97629	0.9763	0.9763	0.9795	0.9795	0.9763	0.9718	0.9718	0.9763
Data-processing package	xia2-3dii	xia2-dials	xia2-3dii	xia2-dials	xia2-3dii	xia2-3dii	xia2-3dii	Multixia2-3dii	xia2-3dii
Space group	$P3_221$	$P3_121$	$P2_12_12_1$	$P4_32_12$	$I4_122$	C121	$P3_221$	$P3_121$	$P12_11$
a, b, c (Å)	69.6, 69.6, 104.7	71.2, 71.2, 96.4	39.2, 83.3, 115.8	70.2, 70.2, 143.2	109.2, 109.2, 189.1	94.4, 68.5, 129.5	76.1, 76.1, 155.9	70.3, 70.3, 97.5	47.2, 96.4, 87.2
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 106.3, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 98.36, 90.0
Resolution range (Å)	52.33 - 1.09 (1.12 - 1.09)	61.68 - 1.58 (1.62 - 1.58)	83.30 - 2.20 (2.26 - 2.20)	63.04 - 2.24 (2.30 - 2.24)	94.58 - 1.93 (1.98 - 1.93)	54.62 - 1.20 (1.23 - 1.20)	50.31 - 1.79 (1.84 - 1.79)	60.86 - 2.03 (2.08 -2.03)	86.31 - 1.16 (1.19 - 1.19)
R_{merge}♩	0.114 (1.367)	0.072 (1.612)	0.143 (0.637)	0.103 (1.602)	0.058 (2.512)	0.052 (0.849)	0.068 (2.425)	0.083 (1.521)	0.057 (0.639)
$\langle I/\sigma(I)\rangle$	15.6 (1.8)	11.3 (1.3)	7.1 (2.2)	14.7 (1.3)	29.8 (1.6)	10.2 (1.2)	16.4 (1.2)	16.9 (1.3)	9.0 (1.7)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	99.1 (99.2)	100.0 (100.0)	99.9 (99.8)	97.5 (88.3)	99.7 (99.5)	99.9 (99.9)	97.7 (93.6)
Multiplicity	18.3 (12.2)	10.3 (9.9)	5.6 (5.7)	22.8 (13.6)	25.9 (23.6)	3.4 (2.9)	10.8 (11.1)	10.0 (5.5)	3.6 (3.1)
Unique reflections	122575	39356	19787	18015	43220	240221	49930	18553	259830
Refinement									
Molecules in the asymmetric unit.	1	1	2 (domain swapped)	1	2 (domain swapped)	4 independent subunits	2 independent subunits	1	4 independent subunits
No. of non-H atoms	1905	1706	3353	1605	3302	7131	3402	1628	6977
R/R_{free} (%)	15 / 16	20 / 21	22 / 29	21 / 26	23 / 26	18 / 19	21 / 24	22 / 28	17 / 18
RMSD bond angle (°)	1.64	1.21	1.95	1.79	1.58	1.75	1.48	1.39	1.79
RMSD bond length (Å)	0.012	0.009	0.019	0.011	0.008	0.014	0.011	0.010	0.015
Ramachandran plot	98.6 % favoured 1.4 % allowed	98.5 % favoured 1.5 % allowed	93.8 % favoured 6.2 % allowed	97.5 % favoured 2.5 % allowed	97.1 % favoured 2.9 % allowed	98.2 % favoured 1.8 % allowed	97.0 % favoured 2.5 % allowed	95.6 % favoured 3.9 % allowed	98.3 % favoured 1.7 % allowed
Molprobity score¶	1.56 (61st)	1.38 (94th)	2.38 (61st)	2.15 (77th)	2.07 (63rd)	1.45 (77th)	1.61 (89th)	2.23 (56th)	1.69 (53rd)

♩ $R_{merge} = \Sigma_{hkl}\ \Sigma_i\ |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\ \Sigma_i\ I_{i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed intensities and the mean intensity of related reflections, respectively. §For the data collection table, values in parentheses are for data in the high resolution shell. ¶ Values in parentheses are percentiles normalised to be on the same scale as structures of comparable X-ray resolution (Chen et al., 2010).

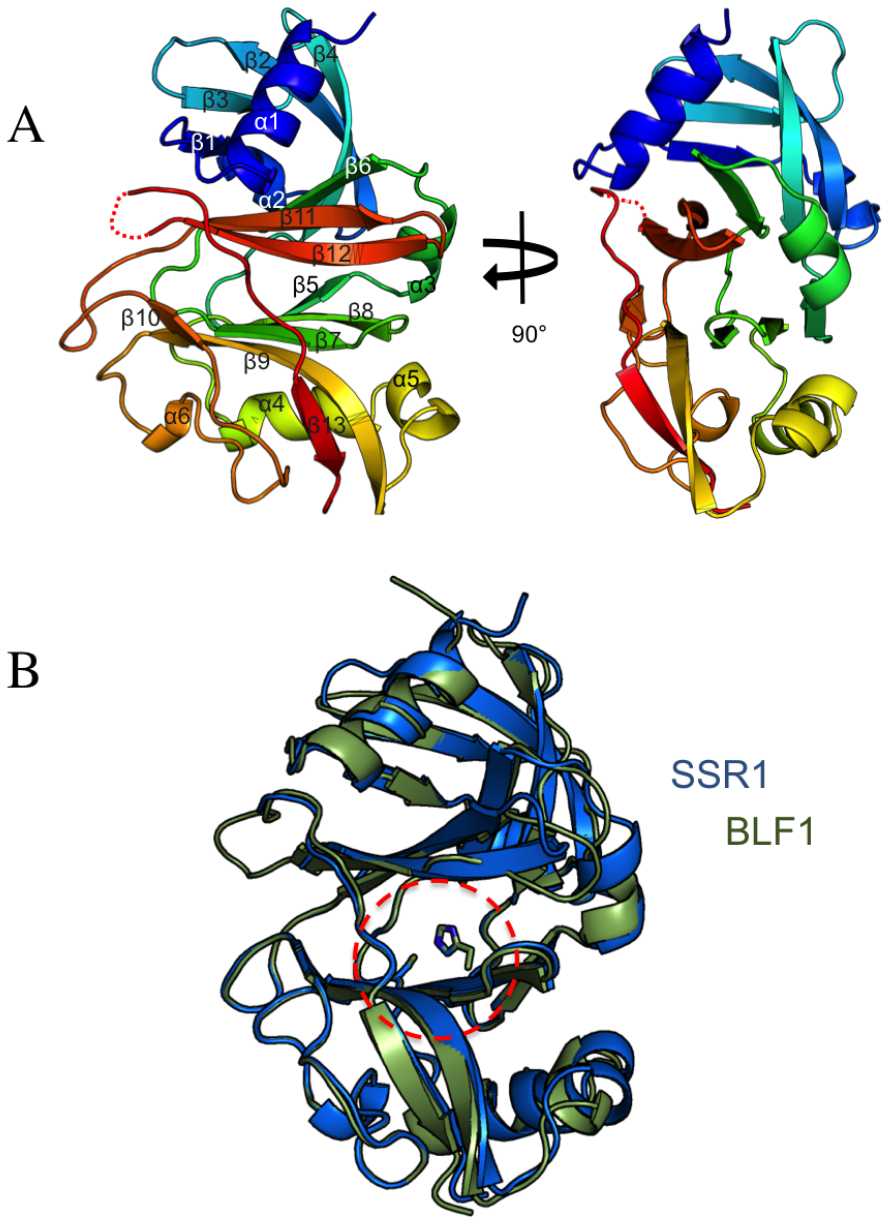


Figure 1. Crystal structure of SSR1. A) Overall structure of SSR1 in rainbow colours with the elements of secondary structure labelled. The position of the clip between Ser191 and Ser192 is shown as a dotted red line. B) SSR1 (blue) superimposed on BLF1 (PDB:3TU8, light green) showing their identical fold and the Cys-His pair in the active site (red circle).

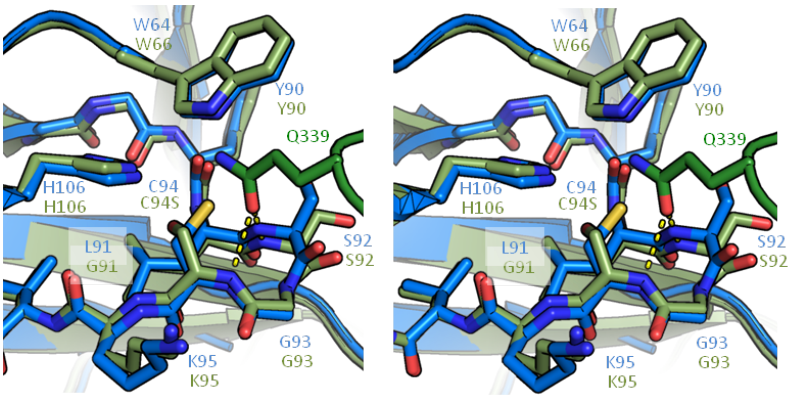


Figure 2. Stereo-diagram of the comparison between the active site of SSR1 and BLF1 C94S in complex with eIF4A (PDB:XXXX). SSR1 is in blue whilst BLF1C94S and eIF4A are in light and dark green, respectively. The torsion angle change at position 94 between SSR1 and BLF1C94S is a consequence of the mutation.

chain had been clipped following Ser191, which lies on the loop connecting β12 and β13 **(Figure 3)**. In the structure of both crystal forms, the electron density for the new C-terminus at Ser191 is clear but the map of crystal form 1 at 1.1 Å also showed that a low occupancy zinc ion can be seen to be bound close to Ser192. Comparisons between SSR1 and BLF1 at the clip site reveals no sequence conservation between them with the connecting loop in BLF1 being longer **(Figure 4)**.

Structure determination of dimeric SSR1 reveals a 3D domain-swapped dimer. Examination of the gel filtration profile of recombinant SSR1 revealed that, in addition to the major peak with an apparent MW of 24 kDa corresponding to the clipped monomer, a second smaller peak with an apparent MW of 45 kDa could be identified (Aziz et al., 2018 submitted). Crystallisation trials on protein from the second peak yielded a new crystal form in space group $P2_12_12_1$. The structure was determined by Molecular Replacement using the SSR1 monomer as a search model and refined to 2.2 Å revealing two molecules in the asymmetric unit arranged as a domain-swapped dimer, whereby a central β-hairpin structure consisting of residues from Tyr165 to Ser192 is swapped between the two symmetry-related subunits **(Figure 5)**. The 14 residues C-terminal to the swapped region then fold back into the original polypeptide chain forming the remainder of the chain to the C-terminus as in the structure of the monomer. Critically, whilst the peptide bond between Ser191 and Ser192 in both crystal forms of the monomer is clipped, in the dimer, no such clipping is observed and continuous electron density can be seen with the polypeptide chain between Ser191 and Ser192 lying in an extended linear conformation **(Figure S1)**.

Following the convention adapted in the review of domain-swapping by Liu and Eisenberg (Liu and Eisenberg et al., 2002)[13], it is convenient to compare the structure of the monomer and the domain-swapped dimer by regarding the subunit structure in the former as being "closed" whereas in the latter it is "open". A closed interface then describes those interactions that exist in both the monomer and the domain-swapped dimer. The open interface represents

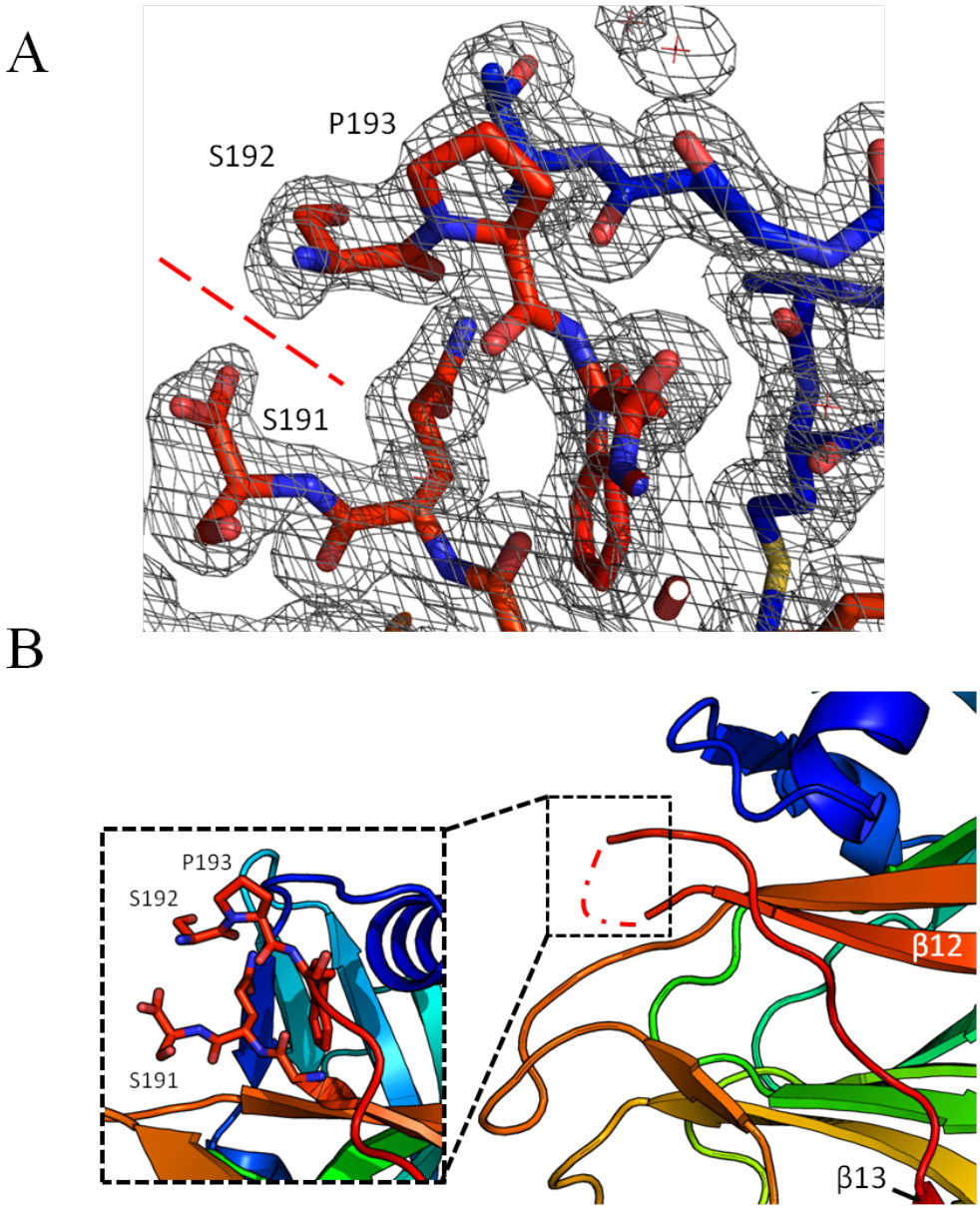


Figure 3. Clipping between Ser191 and Ser192 in monomeric SSR1. A) The electron density map ($2F_{obs}$-F_{calc}) of the structure of crystal form 2 of SSR1 at 1.6 Å, contoured at 1.0 σ level showing clear separation between Ser191 and Ser192 revealing the C-terminal carboxyl group of Ser191. B) A cartoon representation showing the position of the clip site in the loop between β12 and β13 with the inset box showing the position of residues near the clip site.

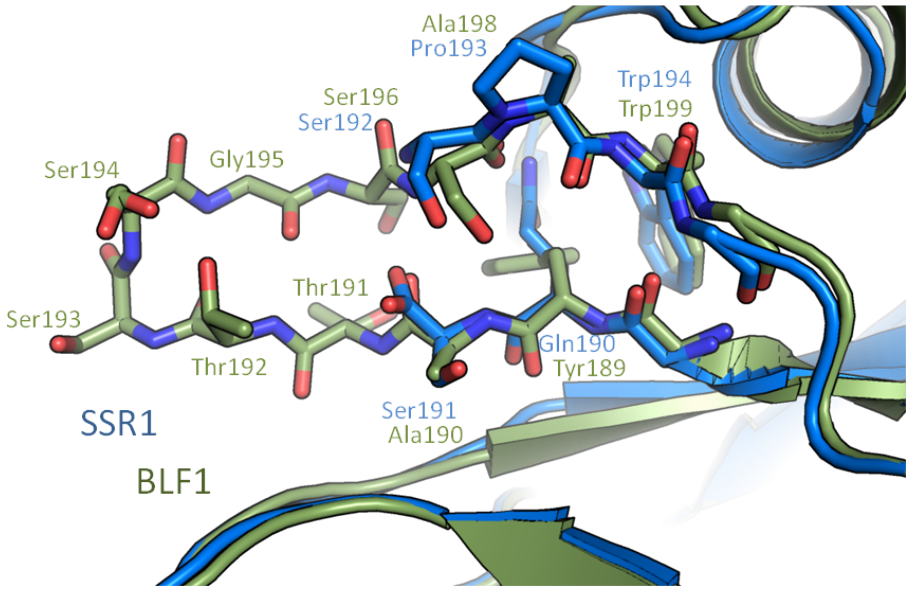


Figure 4. The corresponding loop in BLF1 is different than that in SSR1. Comparison between the loop at the clip site in SSR1 (blue) and the corresponding region in BLF1 (light green) revealing a longer loop and different amino acid sequence in the latter.

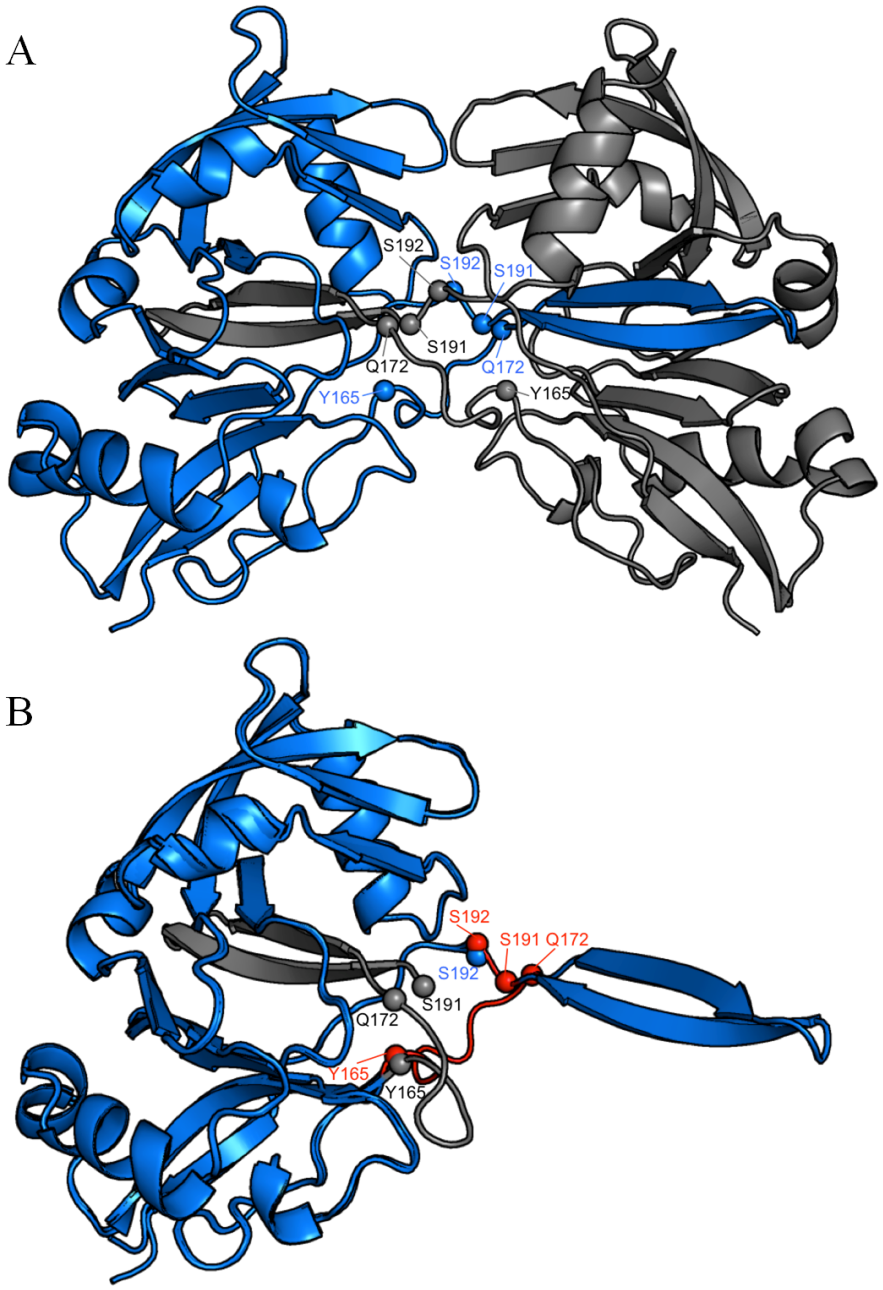


Figure 5. The structure of SSR1 WT dimer revealed a domain swapping event. A) Cartoon representation of the structure of the SSR1 WT dimer with each subunit coloured blue and gray, respectively. B) A subunit from the dimer, the so called "open monomer" superposed on the clipped "closed monomer". The presumed "hinge" regions are coloured red and the residues at both ends of the hinges are shown as spheres. The region in the closed monomer that would be replaced by that from the other subunit is coloured gray.

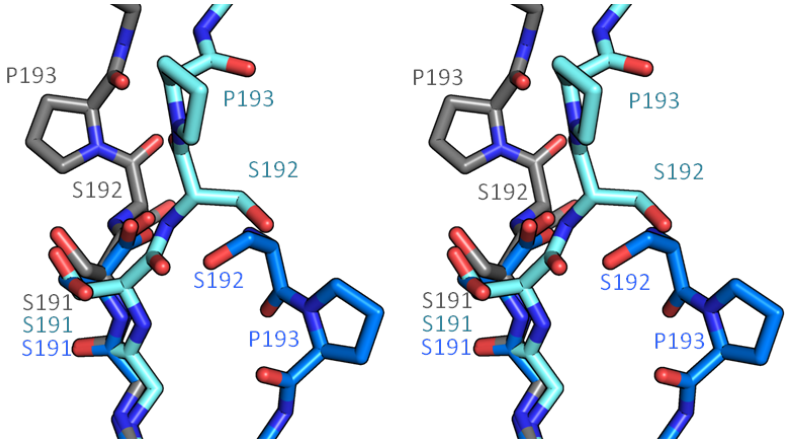


Figure S1. The conformation of the serine residues 191 and 192 in the clipped monomer and the two distinct domain-swapped dimers. A stereo-diagram showing the conformation of the Ser191 and Ser192 in SSR1 WT (gray) and C94S mutant (cyan) domain-swapped dimers compared to those of the clipped monomer (blue) revealing an extended linear conformation of the two residues in the former two structures.

new interactions not seen in the monomer with a flexible hinge describing those residues whose conformations change substantially to allow domain swapping. Comparison of the monomer of SSR1 and the domain-swapped dimer shows that 18 residues are involved in the closed interface, 4 residues in the open interface with the hinges being formed of 10 residues from two regions of the polypeptide chain between Tyr165 and Gln172 and Ser191-Ser192 **(Figure 5, Table 2)**.

Investigation of the origin of peptide cleavage. Analysis of the behaviour of SSR1 samples of the dimer on storage indicate that it is not in exchange with the monomer form implying that the formation of the monomer and the dimer represent the outcome of different events in the folding pathway that are not in equilibrium. We therefore wondered if the conformation of the open monomer in the domain-swapped dimer of SSR1 represented a transient species in the folding pathway that stalls long enough to interact with another subunit of SSR1 in an equivalent conformation to swap domains, whereas in the alternative pathway, the protein folds into a monomeric state that is subsequently clipped **(Figure 6)**.

The observation of the mutually exclusive clipping or domain swapping in the SSR1 monomer and dimer, respectively, suggest three possible mechanisms for the clipping. These involve either an extrinsic protease or a side reaction related to the deamidase activity of the toxin or being a property of the sequence at or near the clip site. Since, an *E. coli* protease deficient strain was used for production of SSR1, a role for an extrinsic protease seemed unlikely. Moreover, the clipping was not reduced by the inclusion of a cocktail of protease inhibitors during isolation, and, to our knowledge, no proteases have yet been identified that cleave a Ser-Ser peptide bond. This left open the latter two options which were investigated by the analysis of a series of mutants.

1) Clipping is not linked to the activity of the toxin. To test the possibility that clipping might be due to the deamidase activity of the toxin, we created an inactive C94S mutant of SSR1 and examined its structure. Analysis of SDS-PAGE gel of SSR1 C94S purified

Table 2. Residues involved in domain swapping and their significance

Construct	SSR1 WT domain-swapped dimer	SSR1 C94S domain-swapped dimer
Swapped residues in the closed interface	A173, I174, V175, G176, G177, V178, V179, N180, A181, S182, K183, Q184, I185, D186, F187, Y188, F189, Q190	P193, W194, A195, T196, L197, P198, Y199, T200, R201, Q202, T203, R204, K205, N206.
Residues in the open interface	L42, G43, T44, S45	T18, G43, T44, S45, Y90, S92, V170, G171, Q172
Hinge residues	Y165, G166, S167, G168, N169, V170, G171, Q172, S191, S192	S191, S192

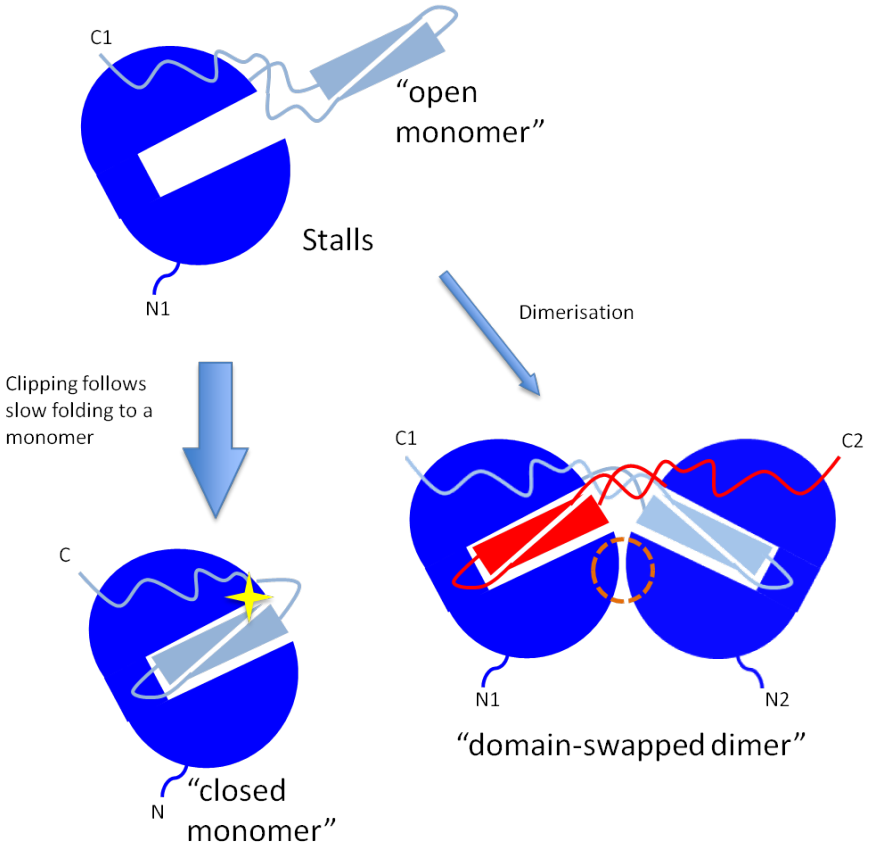


Figure 6. A possible model for the domain swapping. During folding of the SSR1 WT protein, an "open monomer" stalls, allowing time for capture of a second "open monomer" permitting domain-swapping to occur. In a parallel event, the "open monomer" folds into a subunit which then becomes clipped. The swapped element in each subunit is coloured light blue and red, respectively, and the rest of the protein coloured blue. Whilst the "closed interface" involve interactions similar to that in the monomer, the newly formed "open interface" is circled orange. Clipping is indicated by a yellow star.

by gel filtration showed a similar profile to that of the WT indicating that both dimeric and monomeric forms were produced with the monomer still being clipped and the dimer full-length. This indicates that the clipping is not related to the deamidation chemistry of the toxin. The structure of the SSR1 C94S monomer was solved at 2.2 Å from crystals grown in space group $P4_32_12$ with one subunit in the asymmetric unit. Superposition of the monomeric SSR1 C94S structure on SSR1 WT showed that they are very similar (rmsd of 0.44 Å for 202 equivalenced C_α carbons) with an identical clipping pattern. The structure of the dimeric form of SSR1 C94S was determined at 1.9 Å from crystals belonging to space group $I4_122$ with a dimer in the asymmetric unit. Intriguingly, analysis of this structure showed that, whilst SSR1 C94S dimer is not clipped, it still exhibits domain swapping, albeit with the pattern of domain swapping observed being quite different to that seen in the WT toxin. In this instance, residues Tyr165 to Gln190, that form the flexible hinge and the swapped element in the SSR1 WT dimer, fold in a similar fashion to the closed monomer but residues following Ser191 to the C-terminus are swapped between the two subunits, with the hinge only involving Ser191 and Ser192 rather than the more extensive 10 residues hinge seen in the domain-swapped dimer of SSR1 WT **(Figure 7, Table 2)**. In contrast, the open interface in the SSR1 C94S dimer is much larger compared to that of the SSR1 WT dimer, involving 9 residues **(Table 2)**. This distinct pattern in domain-swapping suggests a presence of an alternative folding pathway where the outcome is dependent upon whether the C-terminus is associated with its own subunit or free in solution that also stalls during folding **(Figure 8)**. Interestingly, despite this different pattern of domain swapping in the C94S mutant compared to the WT SSR1, an extended linear conformation of residues Ser191-Ser192 is maintained in both domain-swapped dimers. **(Figure S1)**.

2) The cleavage of SSR1 is sequence-dependent. Analysis of proteins in the adult human lens, a tissue devoid of proteases, has shown that aging can be associated with autocleavage involving the hydroxyl group of serine residues attacking the peptide bond on the N-terminal

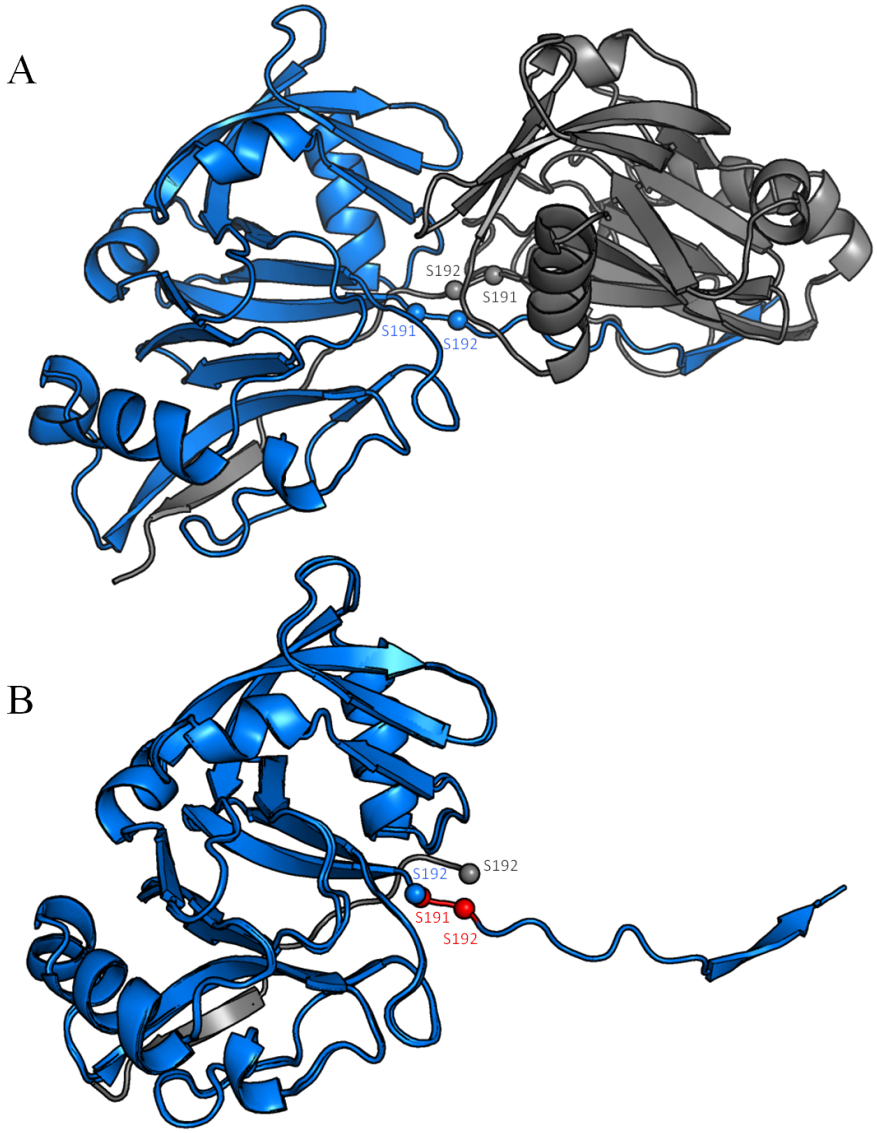


Figure 7. The structure of SSR1 C94S dimer revealed a distinct pattern of domain swapping. A) Cartoon representation of the structure of the SSR1 C94S dimer with each subunit coloured blue and gray, respectively. B) The "open monomer" superposed on the clipped "closed monomer". The presumed "hinge" region is coloured red and the residues at both ends of the hinges are shown as spheres. The region in the closed monomer that would be replaced by that from the other subunit is coloured gray.

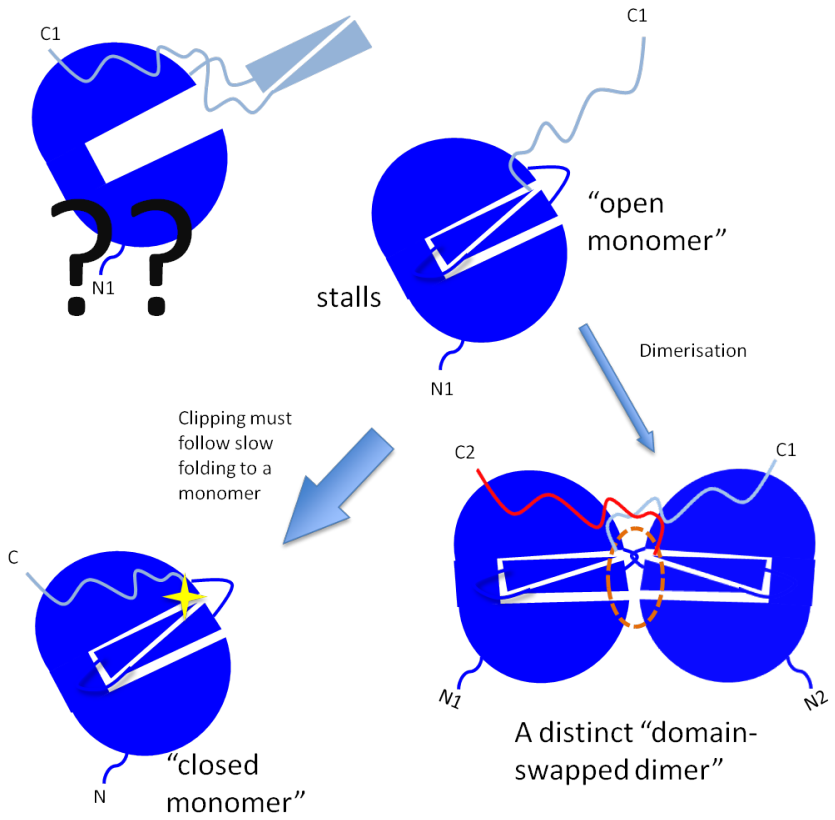


Figure 8. A possible model for the domain swapping of SSR1 C94S mutant. During folding of the SSR1 C94S protein, a distinct "open monomer" stalls, permitting a different domain-swapping pattern. The majority of the "open monomers" then fold into single subunits which then become clipped. The swapped element in each subunit is coloured light blue and red, respectively. The newly formed "open interface" is circled orange and clipping is indicated by a yellow star.

side of the sidechain (Su et al., 2012)[14]. Furthermore, serine residues have been identified as being important in the postulated N-O acyl shift in the first step of self-splicing inteins whereby the nucleophilic hydroxyl group attacks the peptide bond of the residue immediately upstream of it, with some inteins having a threonine or cysteine at the equivalent position and showing the same activity (Noren et al., 2000, Shah & Muir, 2014)[15,16]. This raises the question as to whether either of the serine residues in SSR1 play a direct role in cleavage and whether other factors, such as the influence of residues close to the loop of the toxin or possible strain in the structure, could enhance the propensity of domain swapping and/or the tendency of the peptide bond between the two residues to be clipped.

In an initial study to investigate the roles of residues around the cleavage site, it was decided to create a series of mutants replacing Ser191 and Ser192 with residues lacking the hydroxyl group (Ala/Gly/Val) or residues with potentially equivalent chemistry (Cys/Thr). In addition, given the postulated role of proline residues in influencing the propensity for domain swapping in proteins by introducing strains in the backbone during folding (Rousseau et al., 2001)[17], we further mutated Pro193 to Gly. These mutants were then characterised and, where crystals could be obtained the structures were determined, and analysed to identify the adoption of unusual Ramachandran angles, deviations from standard bond length/angle and/or unusual pattern of thermal parameters which might indicate instability or the presence of close contacts.

SDS-PAGE analysis of an S191A mutant of SSR1 showed a similar pattern of cleavage to WT SSR1 and therefore the sidechain of this residue is clearly not involved in the cleavage reaction. In contrast, analysis of the S192A mutant, and the S191A-S192A and S191A-S192G double mutants of SSR1 revealed no sign of clipping suggesting that Ser192 was essential. However, despite the absence of clipping, the gel filtration run for each of these three mutants clearly revealed the presence of both monomeric and dimeric forms of the full length proteins but with the proportion of the dimer being significantly reduced in the S191A-S192G mutant

(data not shown).

Crystals of monomeric samples of SSR1 S192A, S191A-S192A and S191A-S192G mutants diffracted to 2.0, 1.8, and 1.2 Å, respectively, enabling their structures to be solved and confirming the absence of clipping between residues 191 and 192 **(Figure 9, Table 1)**. Superposition of all three mutants on the coordinates of the SSR1 WT monomer showed their almost identical fold indicating the effects from the mutations were minor (rmsd of 0.35, 0.44 and 0.45 Å for 202 equivalenced C_α carbons for SSR1 S192A, S191A-S192A and S191A-S192G mutants, respectively). Comparison of these structures revealed that the β12-β13 loops in the S192A and the S191A-S192A mutant are very similar albeit with a slight shift of approximately 1 Å **(Figure 10)**. Analysis of the Ramachandran angles for residues at position 192 in both the S192A and the S191A-S192A mutants shows that they are located in or close to an outlier region **(Figure 11)** with the C_β of Ala192 in both structures lying adjacent to the C_δ of Pro193 (approximately 3 Å) giving rise to an adverse steric interaction **(Figure 10)**. Further analysis of the omega angle of the peptide bond between residues 191 and 192 indicates that they are distorted from planarity by approximately 24^o and 16^o (The average for the two subunits in the dimer) in the SSR1 S192A and S191A-S192A mutants of SSR1, respectively, suggesting that this loop in both mutants is in a "strained" state. Consistent with this, the loop in the S191A-S192G mutant adopts a different conformation with Gly192 located in a banned region of the Ramachandran plot for residues other than glycine **(Figure 10)**. Thus, replacing alanine at position 192 with the less sterically hindered glycine, as in the S191A-S192G mutant, allows a different pattern of Ramachandran angles to be adopted in which the strain in this region is reduced, with the omega angle deviating from ideality only by 6^o (the average for the four subunits in the asymmetric unit). This is also consistent with the reduced proportion of the dimer for this mutant suggesting that it folds more easily. Analysis of a mutation of Pro193 to glycine completely abolishes clipping even with the two serine residues at position 191 and 192 (data not shown), consistent with the suggestion that the

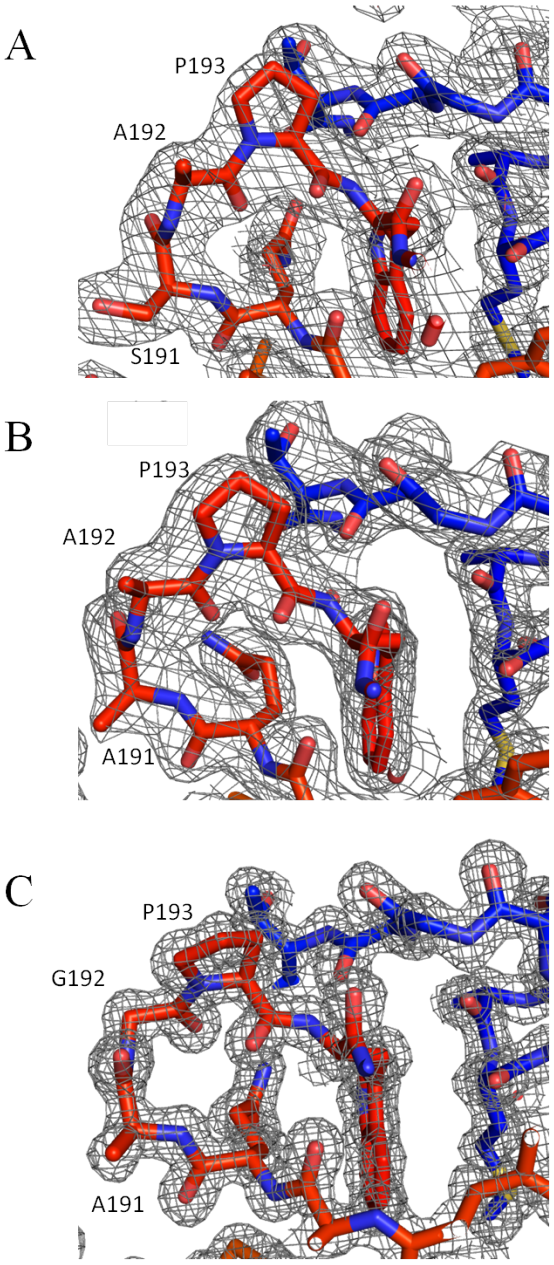


Figure 9. The absence of clipping in clip-site mutant structures. Sticks representation in rainbows with the corresponding electron density map (contoured at 1.0 σ level) surrounding the residues at the clip site of the A) S192A, B) S191A-S192A and C) S191A-S192G mutants, respectively, revealing an intact loop.

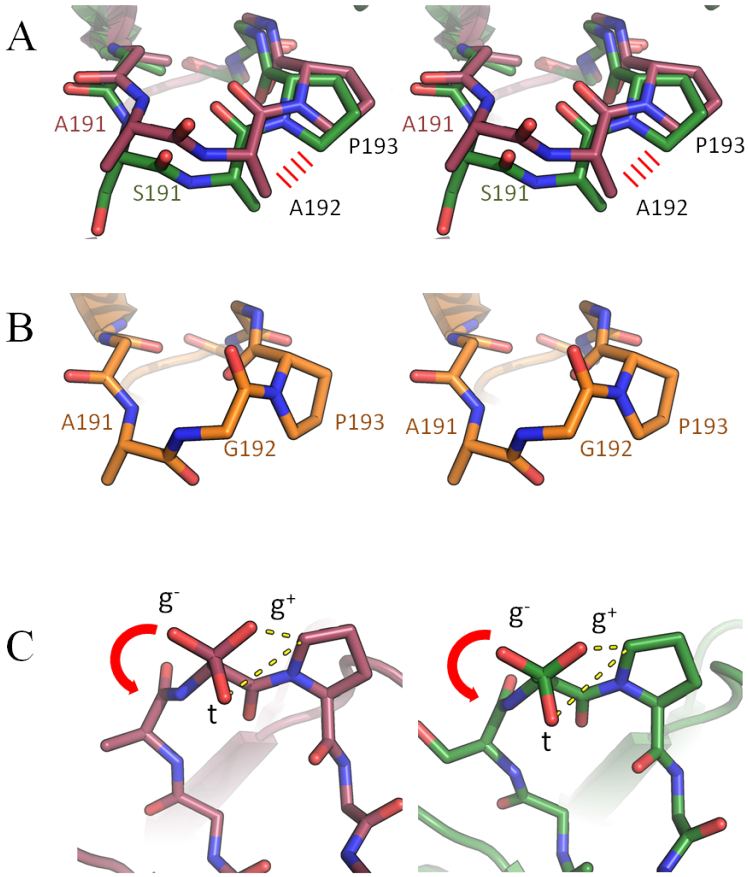


Figure 10. Analysis of the region surrounding the clip site in S191A-S192A, S192A and S191A-S192G mutants of SSR1. A) Stereo-diagram of the sticks representation of the loop in the SSR1 S191A-S192A mutant (raspberry) superposed on that of the S192A mutant (green) revealing similar conformation but with a slight shift. The C_β carbon of Ala192 in both mutants lie close to the C_δ of Pro193 indicating a strain (red lines). B) Stereo-diagram of the corresponding loop in S191A-S192G mutant showing a different conformation. C) Modelling the residue in position 192 as serine in the S191A-S192A mutant (raspberry) (left) and the S192A mutant (green) (right) showed that the rotamer with the least steric clash (yellow dashes) with Pro193, g-, points towards the peptide bond upstream of it.

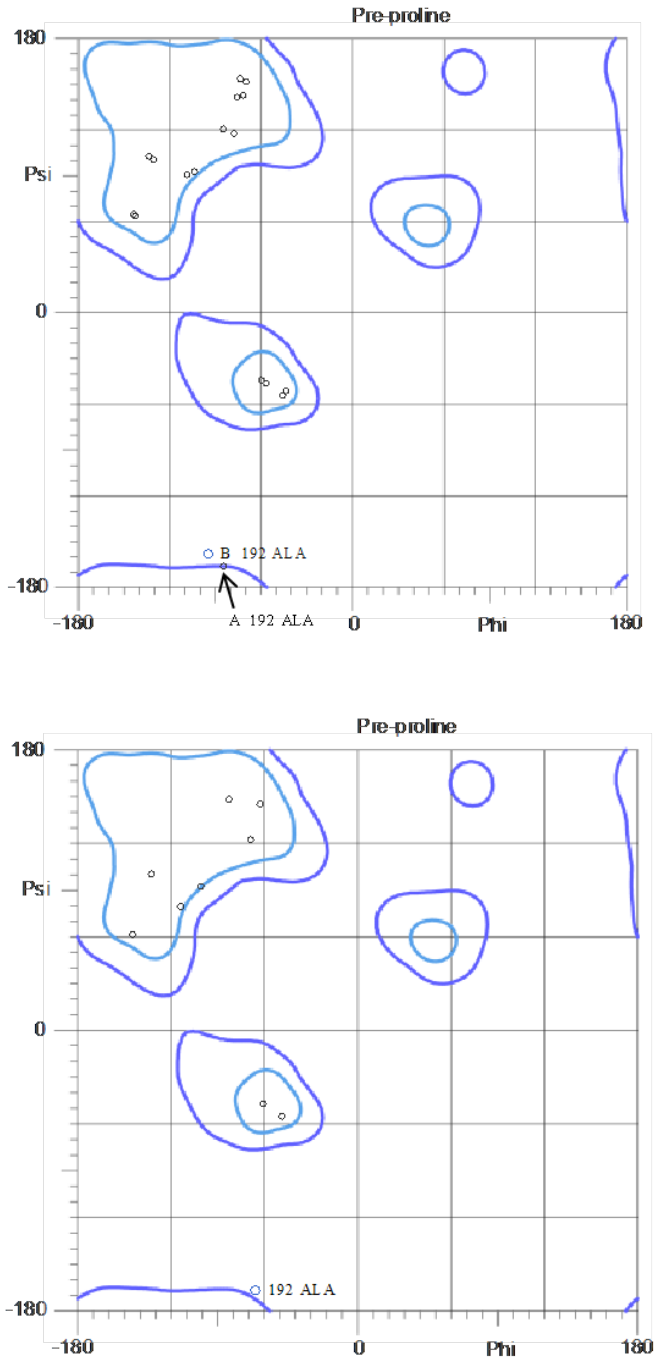


Figure 11. Ramachandran angles of Ala192 in S191A-S192A and S192A SSR1 mutants. Ala192 in the two subunits of SSR1 S191A-S192A mutant (top) and the S192A mutant (bottom) lie in or close to the outlier region in the Ramachandran plot.

origin of the strain lies in the clash between the side chain of Pro193 with the C_β carbon at position 192.

The remaining question was to analyse how the side chain at position 192 influences the clipping. Modeling Ala192 by serine in the structures of SSR1 S192A and S191A-S192A mutants indicates that the clash with Pro193 restricts the *chi1* torsion angle for the serine C_α-C_β bond to the g- conformation, which places the hydroxyl group facing the peptide bond on the N-terminal side possibly assisting nucleophilic attack. In order to further test the possible involvement of the hydroxyl group at position 192 in the clipping, the behavior of S192C and S192T mutants of SSR1, which preserve a possible nucleophile at the γ position of the side chain, was analysed. SDS-PAGE analysis of these mutants following gel-filtration reveals that both monomer and dimer can be observed, and that clipping can be seen to occur in the monomer, albeit at approximately 2-fold or 5-fold reduced rates in the S192C and S192T mutants, respectively. However, unlike the pattern of clipping seen in the WT SSR1 and the other mutants that are clipped, whereby the monomeric peak consists of completely clipped molecules whilst the unclipped variants run as dimers, SDS-PAGE gel analysis of S192C showed a mixture of clipped and unclipped molecules in both the monomeric and dimeric peaks **(Figure S2)**. Analysis of the structure of crystals obtained from the monomer peak of S192C at 1.16 Å resolution revealed that each of the four SSR1 molecules in the asymmetric unit were completely clipped between residues 191-192 but also showed that the cysteine residue at position 192 formed a disulphide bridge with the equivalent residue in another SSR1 molecule to form a quite different dimer **(Figure S3)**. Given the reducing conditions in the cell, we presume that the formation of this disulphide bond is an artefact. This behaviour would explain the different pattern of cleavage that is observed in solution. In contrast to the above, an S192V mutant which removes the functional group at the γ position of the side chain of residue 192, showed no clipping confirming the importance of this atom (Sγ and Oγ of Cys, and Ser/Thr, respectively) in the cleavage reaction in SSR1.

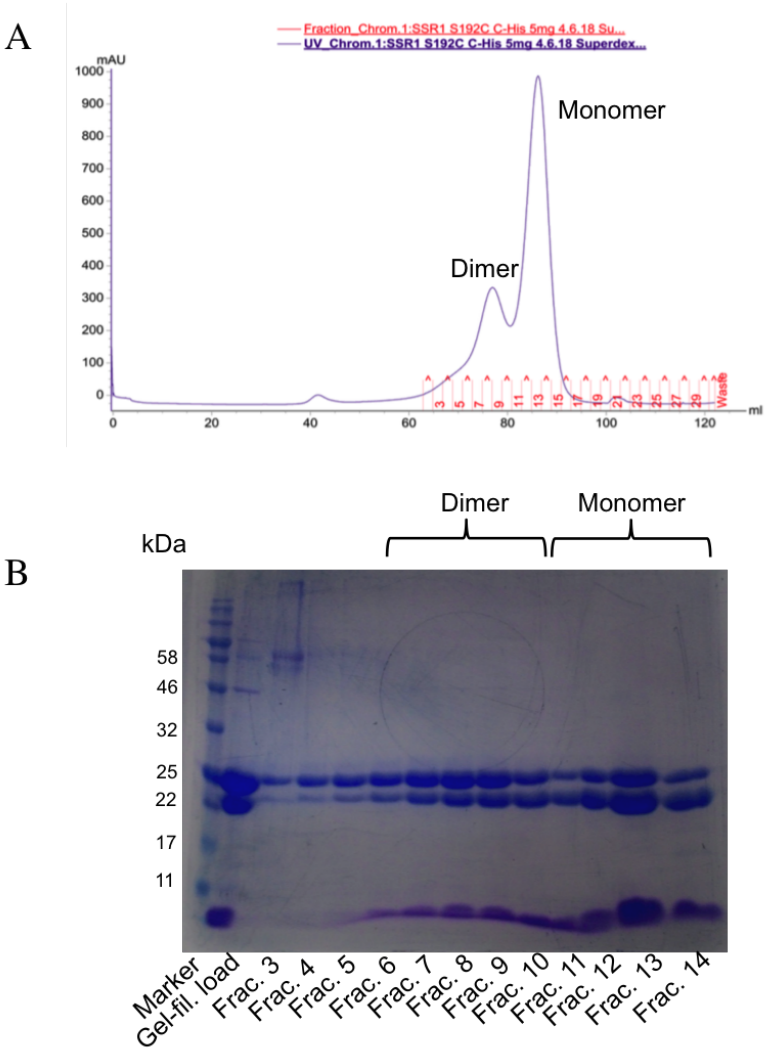


Figure S2. Purification of SSR1 S192C mutant by gel filtration. A) The chromatogram trace for the gel filtration of SSR1 S192C mutant B) The corresponding SDS-PAGE analysis showed a significant amount of unclipped proteins in the monomeric fractions and clipped proteins in the dimeric fractions unlike the situation seen in other SSR1 proteins with S192 (Aziz et al., 2018, submitted).

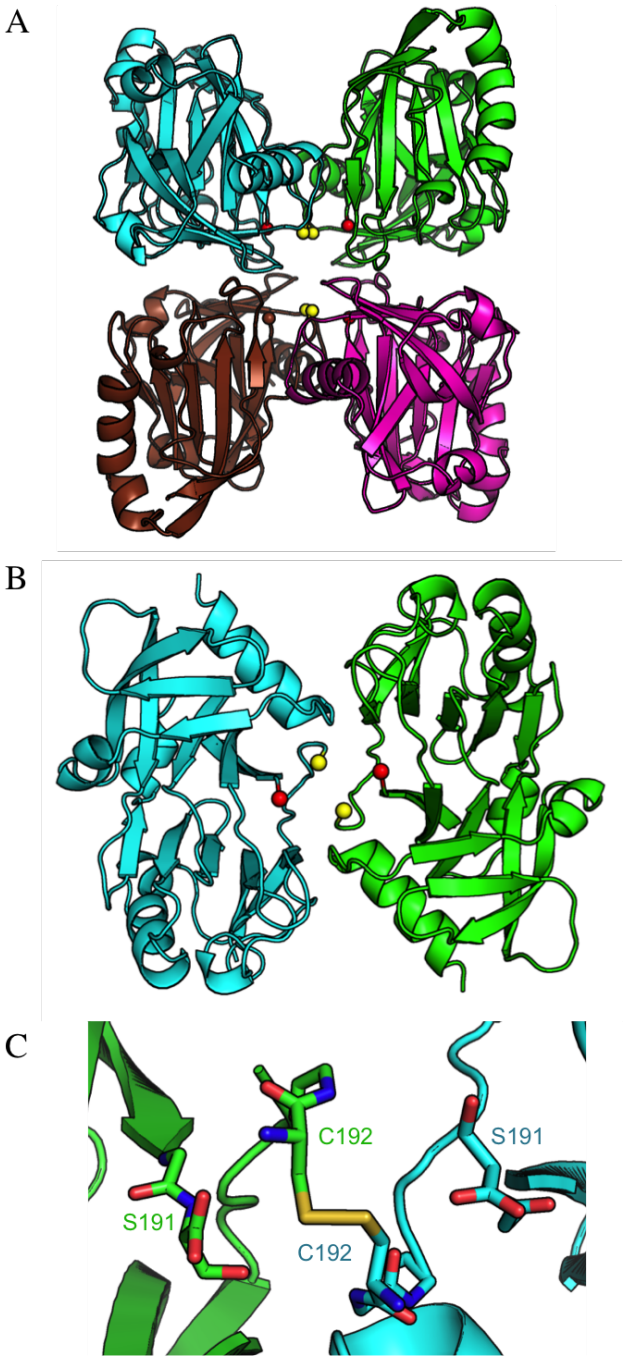


Figure S3. The crystal structure of SSR1 S192C mutant revealing two dimers in the asymmetric unit. A) Four subunits of the SSR1 S192C mutants arranged as two disulphide-linked dimers. B) The disulphide-linked dimer rotated to show residues at positions 191 and 192 as red and yellow spheres, respectively. C) Sticks representation revealing complete clipping and the disulphide bond formed between equivalent C192 residues from the two subunits.

Taken together, these observations suggest that the root of the clipping in SSR1 lies in the strain arising from a clash due to the presence of a C_β carbon at position 192 combined with the presence of a proline residue at position 193. We suggest that the strain results in a stall in the folding leaving the chain in an open, extended “more relaxed" state, in which the residues at position 191-192 lie in a linear extended conformation, which persists for long enough to find partners resulting in the formation of a domain-swapped dimer. Subunits that succeed in folding into a monomeric form are however strained, leading to the distortion of the peptide omega angle and, presumably, a partial loss of some resonance stabilisation of the peptide bond. This would then facilitate nucleophilic attack on the peptide by the hydroxyl group of Ser192 (or equivalent functionality with Cys and Thr). This pattern of clipping, which is therefore Ser/Cys/Thr-dependent resembles that seen in peptide splicing in the intein family of proteins although it clearly has different origins, with the splicing in the latter being the result of enzyme-like catalysis (Noren et al., 2000, Shah & Muir, 2014)[15,16].

The observation in SSR1 of domain-swapping as an alternative folding pathway to overcome the effects of strain serves to illustrate the plasticity of the structure of proteins which enables them to maintain the essential elements of their fold, and by implication, their functions, in the face of potentially destabilising random mutations. This is further illustrated by the different pattern of domain swapping seen in the C94S mutant of SSR1 and other multiple domain-swapping pattern seen elsewhere, for instance in RNAse A (Liu et al., 1998, Liu et al., 2001, Liu et al., 2002)[18–20]. This suggests that the energy difference in the folding landscape between all these different forms is very small thereby contributing to the ability of proteins to adopt a stable fold during evolution.

Methods

0.1 SSR1 production, crystallisation and structure determination

All WT and mutants of SSR1 were produced in recombinant *E. coli* strains and purified in a similar fashion to that described previously (Aziz et al., 2018, submitted). Sitting-drop crystallisation trials were carried out using either Matrix Hydra II (Thermo Fisher Scientific, USA) or mosquito® LCP (TPP Labtech) and a variety of 96-well format screens. Whilst some diffraction data were collected from crystals obtained via the sitting-drop technique, some crystals were optimised via hanging-drop and details on crystallisation conditions are summarised in **Table S1**. The structure of SSR1 WT crystal form 1 was determined by zinc single anomalous dispersion (Zn-SAD) using the ShelXC/D/E suite implemented in the hkl2map GUI (Pape & Schneider, 2004)[21,22]. The initial electron density map revealed 11 Zn^{2+} ions arising from the precipitant, one of which lies in the active site, whilst some are located at the protein surface making crystal contacts. Iterative model building and refinement was carried out using *Coot* and REFMAC in the CCP4 suite (Elmsley et al., 2010, Murshudov et al., 2011, Winn et al., 2011)[23–25].

0.2 Structure determination of dimeric SSR1 and SSR1 mutants

The refined coordinates of the structure of SSR1 from crystal form 1 were used as a search model in molecular replacement by Phaser in the CCP4 suite to solve the structures of all the other crystal forms (McCoy et al., 2007, Winn et al., 2011)[25,26]. Data processing and refinement statistics are summarised in **Table 1**. All structures were validated via the stand-alone wwPDB validation server (Gore et al., 2017)[27] and Molprobity (Chen et al., 2010)[28]. Structural analysis was carried out on *Coot* (Elmsley et al., 2010)[23] and figures were made using PyMOL (Schrodinger).

Table S1. Crystallisation conditions

Construct	SSR1 WT monomer (Crystal form 1)	SSR1 WT monomer (Crystal form 2)	SSR1 WT dimer	SSR1 C94S monomer	SSR1 C94S dimer	SSR1 S191AG monomer	SSR1 S191A-S192A monomer	SSR1 S192A monomer	SSR1 S192C monomer
Crystallisation condition	(Aziz et al., 2018)	(Aziz et al., 2018)	(Aziz et al., 2018)	0.2 M Sodium malonate, 0.1 M Bis Tris propane pH 6.5, 20 % (w/v) PEG 3350.	0.1 M MES pH 6.0, 12 % (w/v) PEG 6000.	0.2 M Calcium chloride, 0.1 M HEPES pH 7.0, 20 % (w/v) PEG 6000.	0.2 M Calcium chloride, 40 % (w/v) MPD.	0.1 M Bicine pH 9.0, 65 % (w/v) MPD.	0.1 M Magnesium chloride, 0.1 M Sodium citrate pH 6.0, 20 % (w/v) PEG 4000.
Sample buffer	(Aziz et al., 2018)	(Aziz et al., 2018)	(Aziz et al., 2018)	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 8	10 mM Tris-HCl pH 7.5	10 mM Tris-HCl pH 8, 0.1M Sodium chloride.	10 mM Tris-HCl pH 8, 0.1M Sodium chloride.	10 mM Tris-HCl pH 8, 0.1M Sodium chloride.
Concentration (mg ml^{-1})	22	22	10	22	11	20	20	20	20

References

1. Washington, E. J., Banfield, M. J. & Dangl, J. L. What a difference a Dalton makes: bacterial virulence factors modulate eukaryotic host cell signaling systems via deamidation. *Microbiol. molecular biology reviews : MMBR* **77**, 527–39 (2013).

2. Sanada, T. *et al.* The Shigella flexneri effector OspI deamidates UBC13 to dampen the inflammatory response. *Nature* **483**, 623–6 (2012).

3. Kitadokoro, K. *et al.* Crystal structures reveal a thiol protease-like catalytic triad in the C-terminal region of Pasteurella multocida toxin. *Proc. Natl. Acad. Sci. United States Am.* **104**, 5139–44 (2007).

4. Yao, Q. *et al.* Structural mechanism of ubiquitin and NEDD8 deamidation catalyzed by bacterial effectors that induce macrophage-specific apoptosis. *Proc. Natl. Acad. Sci. United States Am.* **109**, 20395–400 (2012).

5. Crow, A., Hughes, R. K., Taieb, F., Oswald, E. & Banfield, M. J. The molecular basis of ubiquitin-like protein NEDD8 deamidation by the bacterial effector protein Cif. *Proc. Natl. Acad. Sci. United States Am.* **109**, E1830–8 (2012).

6. Lockman, H. A., Gillespie, R. A., Baker, B. D. & Shakhnovich, E. Yersinia pseudotuberculosis produces a cytotoxic necrotizing factor. *Infect. immunity* **70**, 2708–14 (2002).

7. Buetow, L., Flatau, G., Chiu, K., Boquet, P. & Ghosh, P. Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1. *Nat. structural biology* **8**, 584–588 (2001).

8. Cruz-Migoni, a. *et al.* A Burkholderia pseudomallei Toxin Inhibits Helicase Activity of Translation Factor eIF4A. *Science* **334**, 821–824 (2011).

9. Wiersinga, W. J. *et al.* Melioidosis. *Nat. Rev. Dis. Primers* **4**, 17107 (2018).

10. Limmathurotsakul, D. *et al.* Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. *Nat. Microbiol.* **1**, 15008 (2016).

11. Currie, B. J. Advances and remaining uncertainties in the epidemiology of Burkholderia pseudomallei and melioidosis. *Transactions Royal Soc. Trop. Medicine Hyg.* **102**, 225–227 (2008).

12. Luo, G., Shi, Z., Wang, H. & Wang, G. Skermanella stibiiresistens sp. nov., a highly antimony-resistant bacterium isolated from coal-mining soil, and emended description of the genus Skermanella. *Int. J. Syst. Evol. Microbiol.* **62**, 1271–1276 (2012).

13. Liu, Y. & Eisenberg, D. 3D domain swapping: As domains continue to swap. *Protein Sci.* **11**, 1285–1299 (2002).

14. Su, S.-P. *et al.* Molecular signatures of long-lived proteins: autolytic cleavage adjacent to serine residues. *Aging Cell* **11**, 1125–1127 (2012).

15. Noren, C. J., Wang, J. & Perler, F. B. Dissecting the Chemistry of Protein Splicing and Its Applications. *Angewandte Chemie Int. Ed.* **39**, 450–466 (2000).

16. Shah, N. H. & Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem. science* **5**, 446–461 (2014).

17. Rousseau, F., Schymkowitz, J. W., Wilkinson, H. R. & Itzhaki, L. S. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc. Natl. Acad. Sci. United States Am.* **98**, 5596–601 (2001).

18. Liu, Y., Hart, P. J., Schlunegger, M. P. & Eisenberg, D. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution. *Proc. Natl. Acad. Sci. United States Am.* **95**, 3437–42 (1998).

19. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. A domain-swapped RNase A dimer with implications for amyloid formation. *Nat. Struct. Biol.* **8**, 211–214 (2001).

20. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. Structures of the two 3D domain-swapped RNase A trimers. *Protein science : a publication Protein Soc.* **11**, 371–80 (2002).

21. Pape, T. & Schneider, T. R. *HKL2MAP* : a graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).

22. Sheldrick, G. M. A short history of *SHELX*. *Acta Crystallogr. Sect. A Foundations Crystallogr.* **64**, 112–122 (2008).

23. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Sect. D, Biol. crystallography* **66**, 486–501 (2010).

24. Murshudov, G. N. *et al.* REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

25. Winn, M. D. *et al.* Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. applied crystallography* **40**, 658–674 (2007).

27. Gore, S. *et al.* Validation of Structures in the Protein Data Bank. *Struct. (London, Engl. : 1993)* **25**, 1916–1927 (2017).

28. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Sect. D, Biol. crystallography* **66**, 12–21 (2010).

Appendix B

Extended Methodology - SSR1

Several protein constructs of SSR1 were generated as part of this study. This section serves as an appendix to Paper 2 and Paper 3 to provide a detailed overview on the work that was carried out from cloning to protein production, purification and crystallisation, and the subsequent structure determination of all SSR1 constructs.

B.1 Construct design

B.1.1 *S. stiibiresistense* genomic DNA

The gDNA of *S. stiibiresistense* SB22^{T} was kindly provided by Prof. Gejiao Wang from State Key Laboratory of Agricultural Microbiology, College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan, China. The gDNA was then used to clone N825_12680, which will be referred to hereinafter as SSR1, DNA into vectors suitable for tunable expression with the *E.coli* recombinant system. The vectors containing *ssr1* gene would then be used as the template for the generation of different mutants of SSR1 employing site-directed mutagenesis (SDM).

B.1.2 Generation of SSR1 WT constructs

The primers for cloning *ssr1* (GenBank:EWY38661.1) were designed using the NEB primer design tool (https://tmcalculator.neb.com/) to be used with the Q5® High-Fidelity 2X Master Mix by NEB. The primers used are detailed in Paper 2, Table 1, to produce an untagged and a C-terminal His_6-tagged SSR1 WT constructs and PCR reaction was carried out as in Section 2.3. The amplified gene products were then purified by DNA Purification Kit (Qiagen) producing 47 and 40 ng μl^{-1} DNA amplicon for the untagged and the C-terminal His_6-tagged SSR1 WT constructs, respectively.

The purified PCR products and pET24a vector conferring selectable kanamycin resistance marker and T7 expression system, were treated separately with the corresponding restriction enzymes (Paper 2, Table 1) to create sticky ends (Section 2.3.2). The digested pET24a vector was then subjected to 1% DNA agarose gel electrophoresis to be purified using DNA Gel Extraction Kit (Section 2.3.3). The digested PCR products were then purified from traces of restriction enzymes using DNA purification Kit (Qiagen) prior to ligation with T4 DNA Ligase in the ratio of 5:1 (gene:vector), see Section 2.3.5 for more details. The ligated constructs were then transformed into chemically-competent *E. coli* DH5α silver cells (Section 2.5) and incubated on LB-agar plates treated with 50 μg ml^{-1} kanamycin.

The successful ligation of *ssr1* gene into pET24a vector was initially tested via colony PCR (Section 2.6) by screening multiple colonies to increase the success rate of obtaining genes of the correct size. Agarose gel analysis of these PCR products showed potential candidates (Figure B.1).

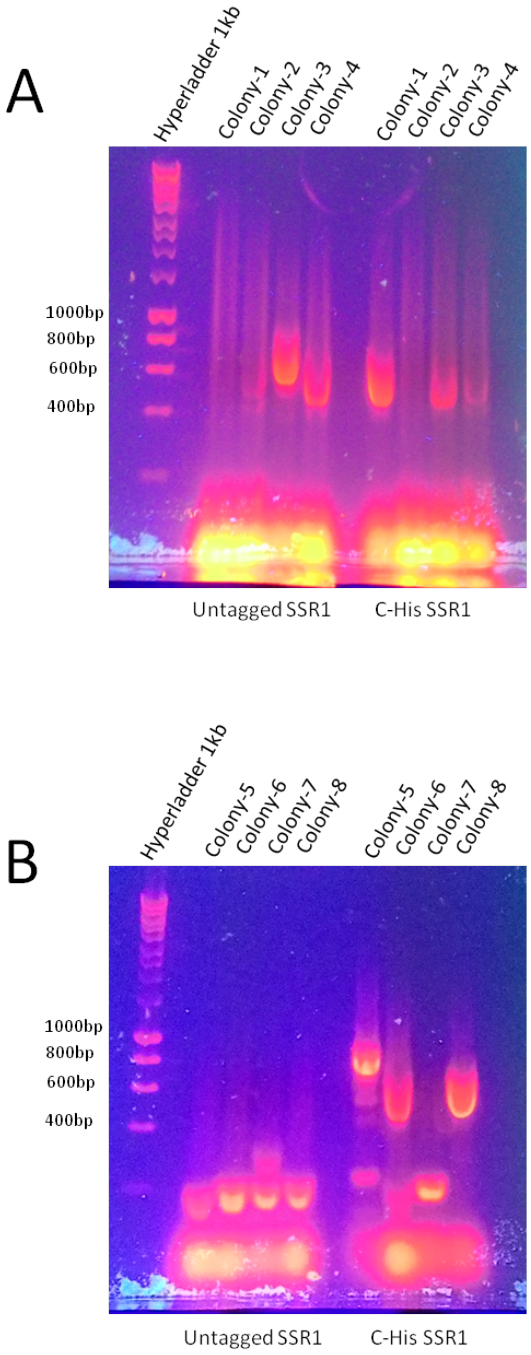


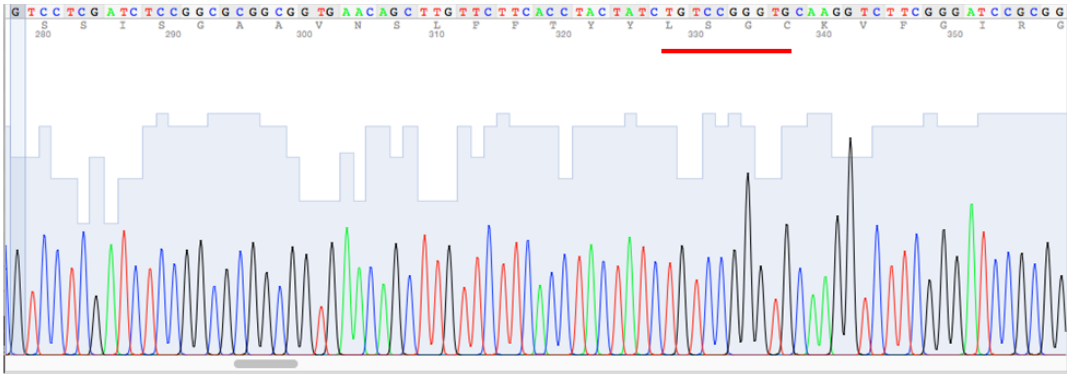
Fig. B.1 **An agarose gel for the colony PCR of SSR1 Untagged and C-His constructs.** Agarose gels showing colonies (A) 1-4 and (B) 5-8 of the untagged and the C-His constructs, respectively. Potential candidates include colonies 3 and 4 of the untagged constructs, and colonies 1,3,6 and 8 for the C-His variant.

Chosen colonies were then grown for plasmid propagation and purification. The purified plasmids were subsequently sent for DNA sequencing (GATC Biotech, Germany) to confirm the correct gene sequence. As an example, the DNA sequence chromatogram obtained from colony 4 of the untagged SSR1 WT construct was analysed by ExPASy Translate (SIB) and BLASTp (NCBI) revealing a full length SSR1 protein (Figure B.2). Analysis of the T7 region of plasmid also revealed similar sequences critical for ensuring successful T7 expression. Similar analysis was done on the other constructs to ensure the accuracy of the DNA sequence.

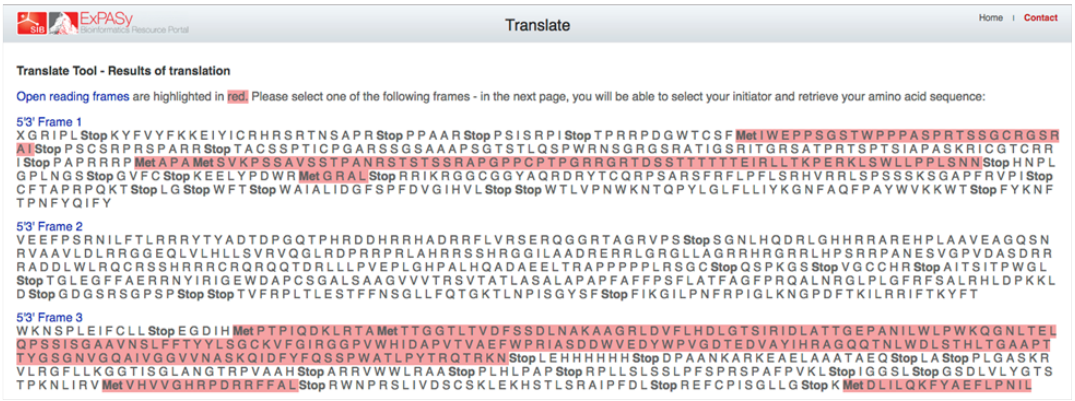
B.1.3 Generation of a variety of SSR1 mutant constructs

The work done for the investigations described in Paper 3 involved the generation of a series of SSR1 mutants via SDM using Q5® Site-Directed Mutagenesis Kit by NEB (protocols are included with the kit or obtainable online). The primers for the different mutants of SSR1 were designed using an online tool, NEBaseChanger™ provided by NEB (http://nebasechanger.neb.com/) (Figure B.3), and are listed in Table B.1. For single base mutations, purified plasmids harbouring WT SSR1 sequence were used as a template, whereas, for double mutation, related SSR1 mutant plasmids were used as templates instead. Following PCR runs and KLD incubation (Q5 SDM Kit), the plasmids were transformed into *E. coli* NEB5α cells and grown on LB-agar plates supplemented with kanamycin. Colonies were selected, and their plasmids purified for sequencing in triplicates. The correct mutations were checked by analysing the quality of the chromatogram traces and the corresponding translated products using ExPASy Translate and BLASTp online tool as described above.

A



B



C

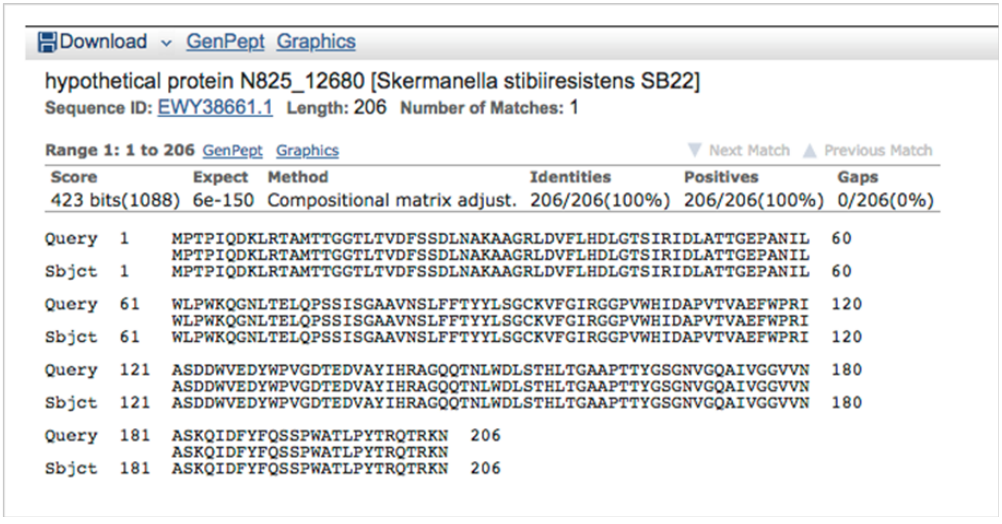


Fig. B.2 **Analysis of DNA sequencing result of Colony 4 of the untagged SSR1 WT from the Colony PCR.** A) The chromatogram peaks were analysed for quality to ensure that there were no overlapping peaks. The DNA sequence corresponding to the LSGC motif is underlined. B) The DNA sequence was analysed by ExPASy Translate to reveal that Frame 3 is the correct reading frame. C) The translated protein sequence was then compared to the SSR1 sequence in the database by NCBI BLASTp.

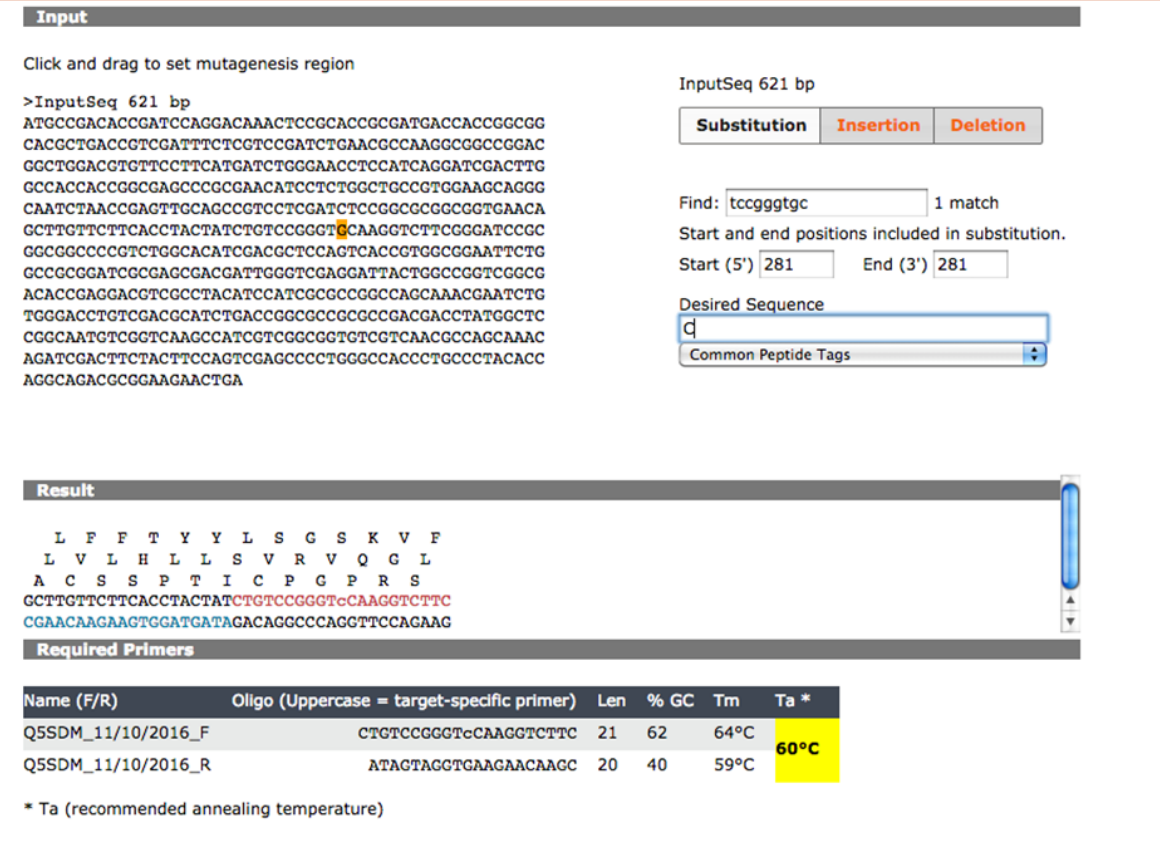


Fig. B.3 **The NEBaseChanger online tool allows designing primers for use with the Q5 SDM Kit.** Primers for the C94S mutant of SSR1 were designed by inserting SSR1 DNA sequence and substituting the nucleotides with the desired mutation. The resulting mutant protein product is shown and the corresponding primers with their properties are highlighted. The design of the primers for the other constructs was done in a similar fashion.

Table B.1 Primers for all SSR1 constructs in this study. Bold texts indicate the substitution of nucleotides to generate the desired mutants

Construct	Forward Primer	Reverse Primer	Template
SSR1 C94S C-His	CTGTCCGGGT**C**CAAGGTCTTC	ATAGTAGGTGAAGAACAAGC	SSR1 WT C-His
SSR1 S191A C-His	CTACTTCCAG**G**CGAGCCCCTG	AAGTCGATCTGTTTGCTGGC	SSR1 WT C-His
SSR1 S191A-S192G C-His	CTACTTCCAG**G**CG**G**GCCGAGCCCCTG	AAGTCGATCTGTTTGCTG	SSR1 WT C-His
SSR1 S191A-S192A C-His	TTCCAGGCGG**C**CCCCTGGGCC	GTAGAAGTCGATCTGTTTGCTGGCGTTGAC	SSR1 S191A-S192G C-His
SSR1 S192A C-His	CTTCCAGTCG**GC**CCCCTGGGCC	TAGAAGTCGATCTGTTTGCTG	SSR1 WT C-His
SSR1 S192C C-His	CTTCCAGTCG**T**GCCCCTGGGC	TAGAAGTCGATCTGTTTGCTGGC	SSR1 WT C-His
SSR1 S192T C-His	TTCCAGTCGA**C**CCCCTGGGCC	GTAGAAGTCGATCTGTTTGCTGG	SSR1 WT C-His
SSR1 P193G C-His	CCAGTCGAGC**GG**CTGGGCCACC	AAGTAGAAGTCGATCTGTTTGC	SSR1 WT C-His

B.2 Protein production of a variety of SSR1 constructs in *E.coli* recombinant system

The incorporation of the *ssr1* gene into pET24a allows tightly-controlled expression by the T7 promoter region which is inducible by IPTG. Following transformation of the plasmids into *E. coli* BL21 (DE3) strains, test expressions were carried out to obtain an optimal condition for the production of soluble SSR1 at high expression level which is crucial for crystallisation studies.

B.2.1 Initial test expression of SSR1 WT in LB-kanamycin media

As an initial trial, BL21 (DE3) *E. coli* cells harbouring the plasmids of SSR1 WT were grown in LB media supplemented with kanamycin (50 μg ml^{-1}) and following induction with 1 mM IPTG, cell cultures were grown at various temperatures and incubation time. The level of expression and solubility of the proteins were then analysed via SDS-PAGE. Unlike BLF1 that expresses very well at 37 ^{o}C (Cruz-Migoni et al., 2011), initial expression trials indicated that SSR1 proteins were mainly insoluble over a wide range of temperatures and incubation times (Figure B.4).

B.2.2 Initial test expression of SSR1 WT in M9 minimal media

The growth medium for SSR1 expression was then switched to M9 minimal media and expression trials were done at different temperatures and incubation times. SDS-PAGE analysis showed that the optimum condition for SSR1 production is at 12 ^{o}C for 64 hours incubation period (Figure B.4). Subsequent overproduction of the SSR1 proteins were then fixed to 72 hours incubation time. See Paper 2 for more details on the protocol. In addition, the initial test expression suggested that the soluble SSR1 proteins were clipped, producing a mix of full-length SSR1 and another variant approximately 3 kDa shorter (discussed in detail

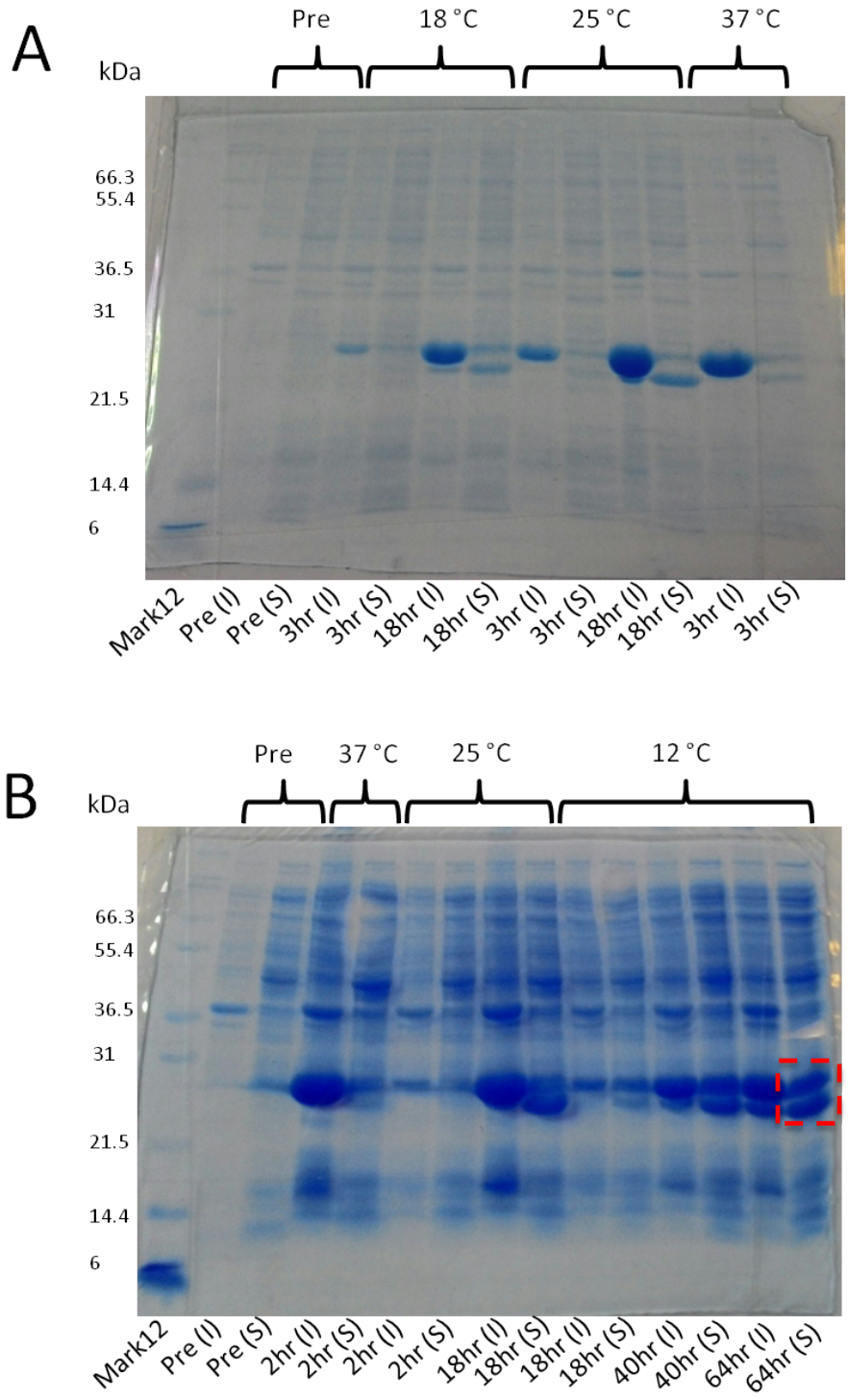


Fig. B.4 **SDS-PAGE gels of the test expression trials for the production of SSR1 proteins.** A) Expression trials using LB-kan media over a variety of temperatures mainly produced insoluble proteins. B) M9 minimal media expression produced more soluble SSR1 proteins but with a significant sign of clipping (red box). Pre: pre-induced, I: insoluble fraction, S: soluble fraction.

in Paper 2 & 3). Subsequent analysis including a cocktail of protease inhibitors (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Merck) in sonication buffer did not affect the level of cleavage (Figure B.5) which implies that the clipping is not likely to be the result of proteases supporting the argument in Paper 3.

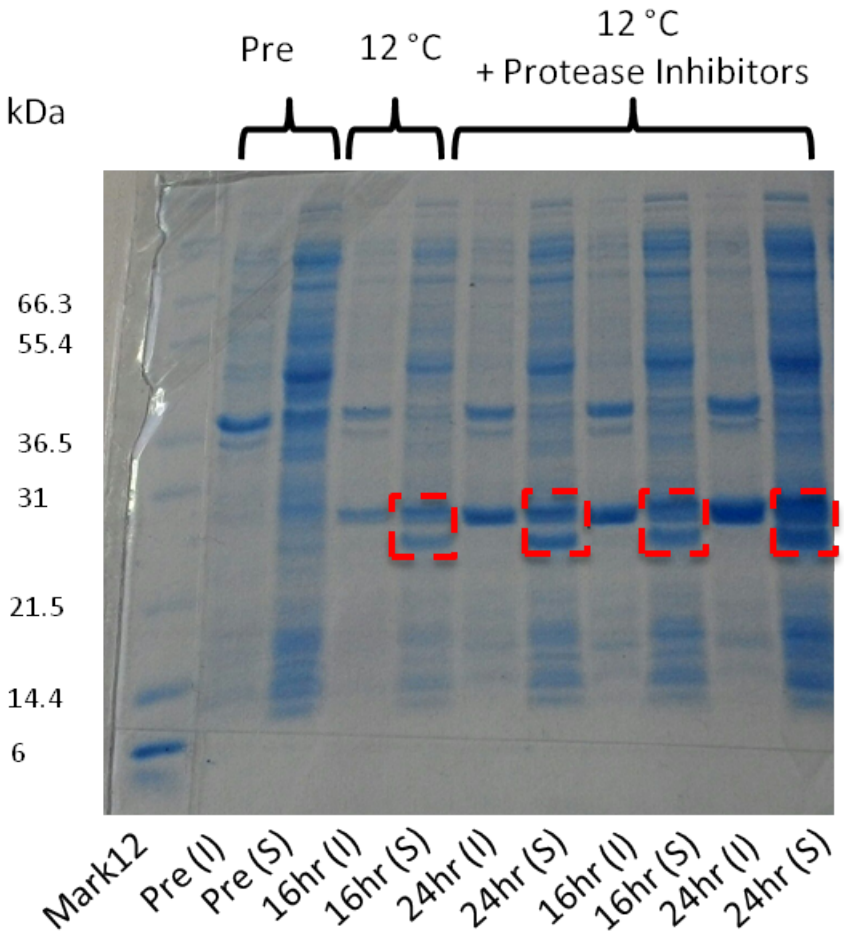


Fig. B.5 **SDS-PAGE gels of the test expression trials including the use of protease inhibitors.** Clipping persisted (red boxes) even with the inclusion of protease inhibitors in the buffer compared to the control. Pre: pre-induced, I: insoluble fraction, S: soluble fraction.

B.2.3 Large-scale production of SSR1 proteins for structural and biochemical studies

Large-scale production of both SSR1 WT constructs was done in M9 minimal media supplemented with kanamycin and cells were grown under the optimum condition described above in 4 x 500 ml cultures. The results of the over production of the untagged and the C-terminal His_6-tagged constructs of SSR1 WT are shown in Figure B.6. The optimum condition for the production of the different mutants of SSR1 was shown to be similar in the initial test expression, hence the subsequent expression of all SSR1 constructs utilised the same approach producing identical level of soluble proteins. However, as discussed in Paper 3, the pattern of clipping is distinct and could be observed on the SDS-PAGE gels as illustrated in Figure B.7 & Figure B.8.

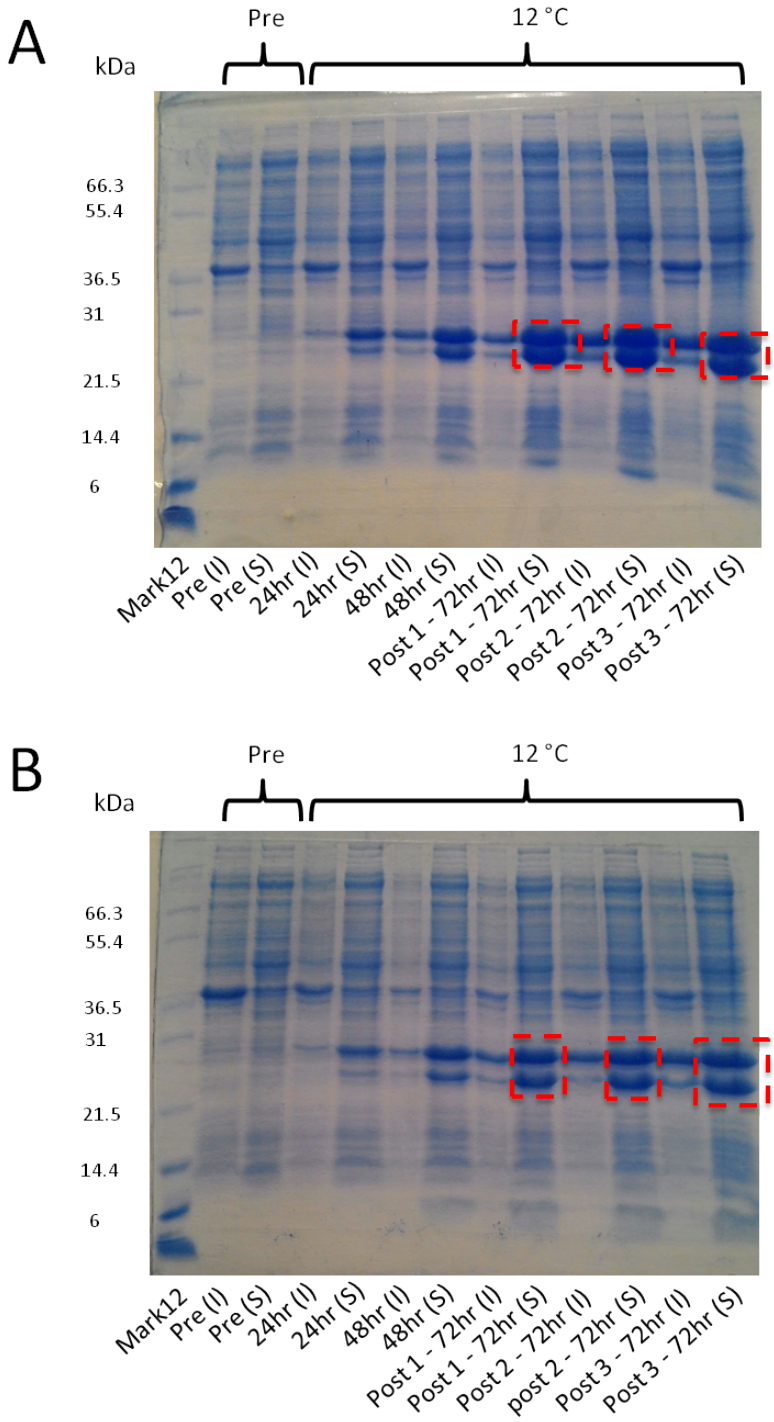


Fig. B.6 **SDS-PAGE gels of the large-scale expression of SSR1 proteins.** A) SDS-PAGE gel of the untagged SSR1 WT proteins showing a high level of soluble protein while still being clipped (red boxes). B) Similar pattern (red boxes) can be seen in the C-His SSR1 WT gel. Pre: pre-induced, Post 1,2: post-induced flask 1,2 etc, I: insoluble fraction, S: soluble fraction.

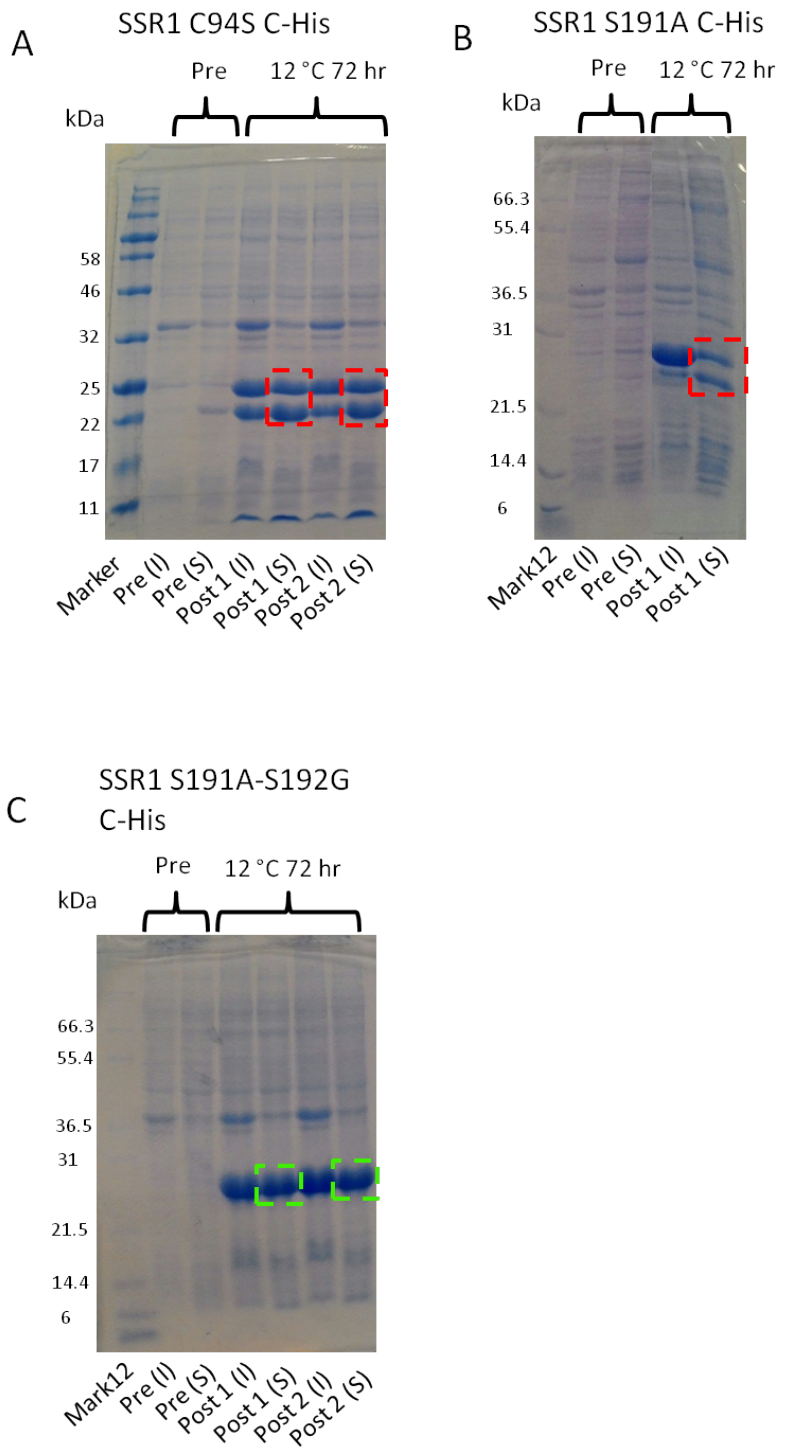


Fig. B.7 **SDS-PAGE gels of the large-scale expression of SSR1 proteins.** SDS-PAGE gels of A) SSR1 C94S C-His, B) SSR1 S191A C-His and C) SSR1 S191A-S192G C-His mutants, respectively. The presence of clipping is shown in red box whilst the absence is shown in green box. Pre: pre-induced, Post 1,2: post-induced flask 1,2 etc, I: insoluble fraction, S: soluble fraction.

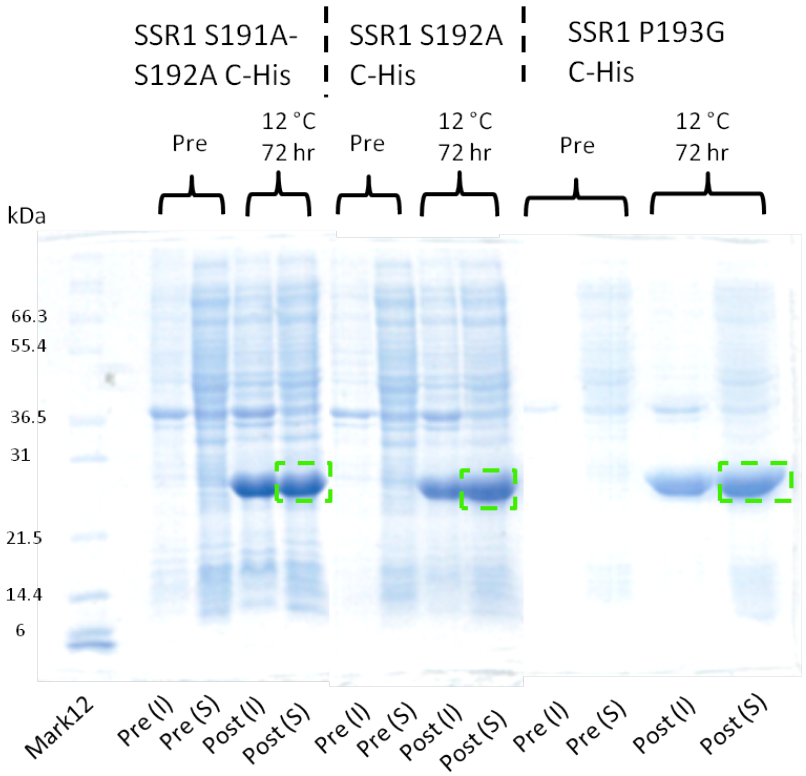


Fig. B.8 **SDS-PAGE gels of the large-scale expression of SSR1 proteins.** SDS-PAGE gels of, from left to right, SSR1 S191A-S192A C-His, SSR1 S192A C-His and SSR1 P193G C-His mutants, respectively. Clipping appeared to be absent in these mutants (green boxes). Pre: pre-induced, Post 1,2: post-induced flask 1,2 etc, I: insoluble fraction, S: soluble fraction.

B.3 Protein purification of the various SSR1 constructs

B.3.1 Purification of the untagged and the C-terminal His_6-tagged SSR1 WT proteins

For the purification of untagged SSR1 WT, ion exchange (DEAE Sepharose) followed by gel-filtration chromatography were carried out as described in Paper 2. The chromatogram traces and the associated SDS-PAGE are shown in Figure B.9. The SDS-PAGE gel showed that the peak corresponding to the MW of monomeric SSR1 contained clipped proteins with >95% purity which were subsequently crystallised. In contrast, the peak corresponding to the dimer appeared to contain both the unclipped and clipped variants of SSR1, as a result of contamination from the monomer peak and significant other impurities Figure B.9. Therefore, given that the sample was only 50 % pure, no crystallisation trials were attempted with this sample. The C-terminal His_6-tagged SSR1 WT was purified as described in Paper 2. Since there was no need to crystallise the C-terminal His_6-tagged monomer, as the structure of the untagged WT SSR1 had been determined, only the purification and crystallisation of the tagged dimer was pursued. The chromatogram of the Nickel-affinity purification and the corresponding SDS-PAGE gel are shown in Figure B.10, whilst details of the gel-filtration run are discussed in Paper 2. The gel-filtration peak running at the dimeric MW for the C-terminal His_6-tagged SSR1 WT contained proteins with higher levels of purity and therefore were subjected for crystallisation trials.

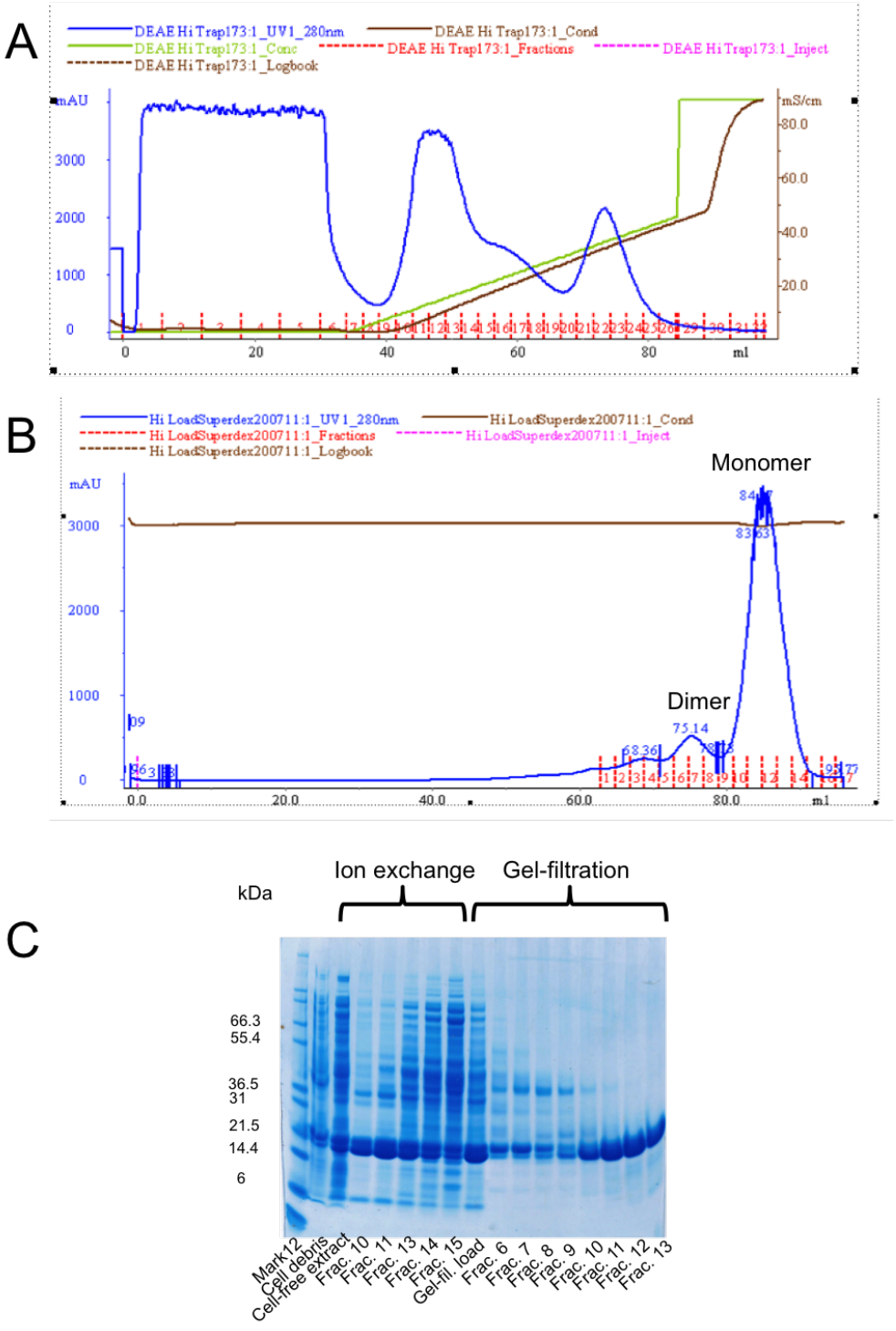


Fig. B.9 **Purification of the untagged SSR1 WT protein.** A) The chromatogram trace of the ion exchange column. B) The chromatogram trace of the following gel-filtration column purification C) SDS-PAGE gel analysis of fractions from A) and B).

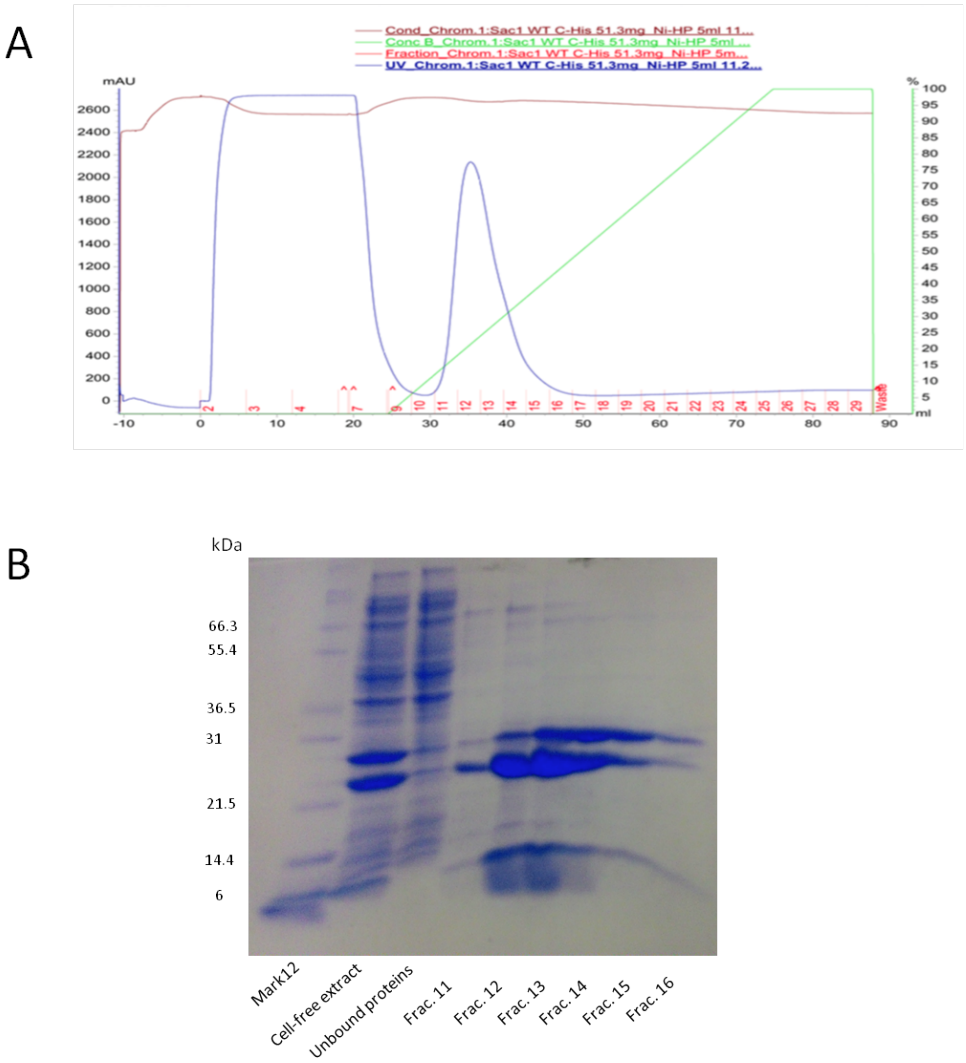


Fig. B.10 **Purification of the SSR1 WT C-His protein via Nickel-affinity column.** A) The chromatogram trace of the Nickel-affinity column. SSR1 eluted at approximately 0.13 M Imidazole. B) SDS-PAGE gel analysis of the resulting fractions. Since the 3 kDa C-terminal clipped fragment co-elutes with the larger 21 kDa fragment, this indicates that despite clipping, both fragments remain associated.

B.3.2 General information regarding the purification of other SSR1 constructs

Comparison of the homogeneity level of the dimeric form of SSR1 WT in the untagged and the C-terminal His_6-tagged variant has led to a decision to use only the C-terminal His_6-tagged construct for further mutagenesis studies enabling the production and purification of homogeneous samples of the dimeric form in addition to the monomeric form of SSR1 proteins. Since all the following constructs are C-terminal His_6-tagged, similar purification approach was used as that used in purifying C-terminal His_6-tagged SSR1 WT proteins.

B.3.3 Purification of the C-terminal His_6-tagged SSR1 C94S mutant protein

C-terminal His_6-tagged SSR1 C94S proteins were purified in a similar fashion to the C-terminal His_6-tagged SSR1 WT proteins and the results are shown in Figure B.11 & Figure B.12. From 2 g cell pellet, in 34 ml re-suspension buffer, 3 mg ml^{-1} CFE was obtained, and following Nickel-affinity purification, the combined fractions containing the proteins were concentrated and applied to gel-filtration column. The fractions corresponding to the monomeric and the dimeric form of the protein were concentrated separately to 22 and 11 mg ml^{-1} in 10 mM Tris-HCl pH 7.5 and 10 mM Tris-HCl pH 8 buffer, for the monomeric and dimeric forms, respectively. As discussed in Paper 3, similar clipping profile can be seen from the SDS-PAGE analysis of the gel-filtration fractions (Figure B.12). The two samples were then concentrated and subjected to crystallisation trials.

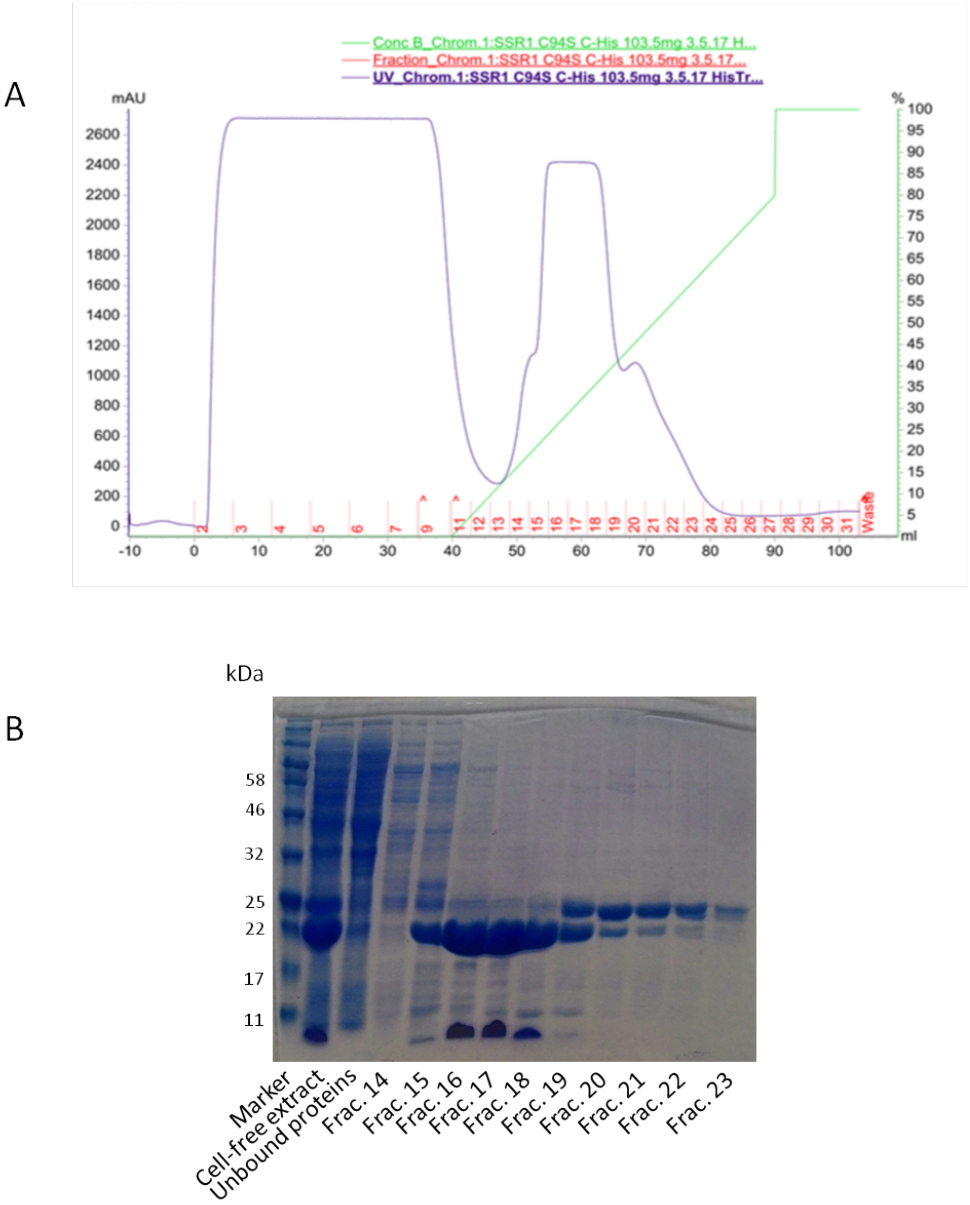


Fig. B.11 **Purification of the SSR1 C94S C-His proteins via Nickel-affinity column.** A) The chromatogram trace of the Nickel-affinity column. B) SDS-PAGE gel analysis of resulting fractions. Interestingly, the unclipped variant (assumed to be in the dimeric form) eluted at higher imidazole concentration.

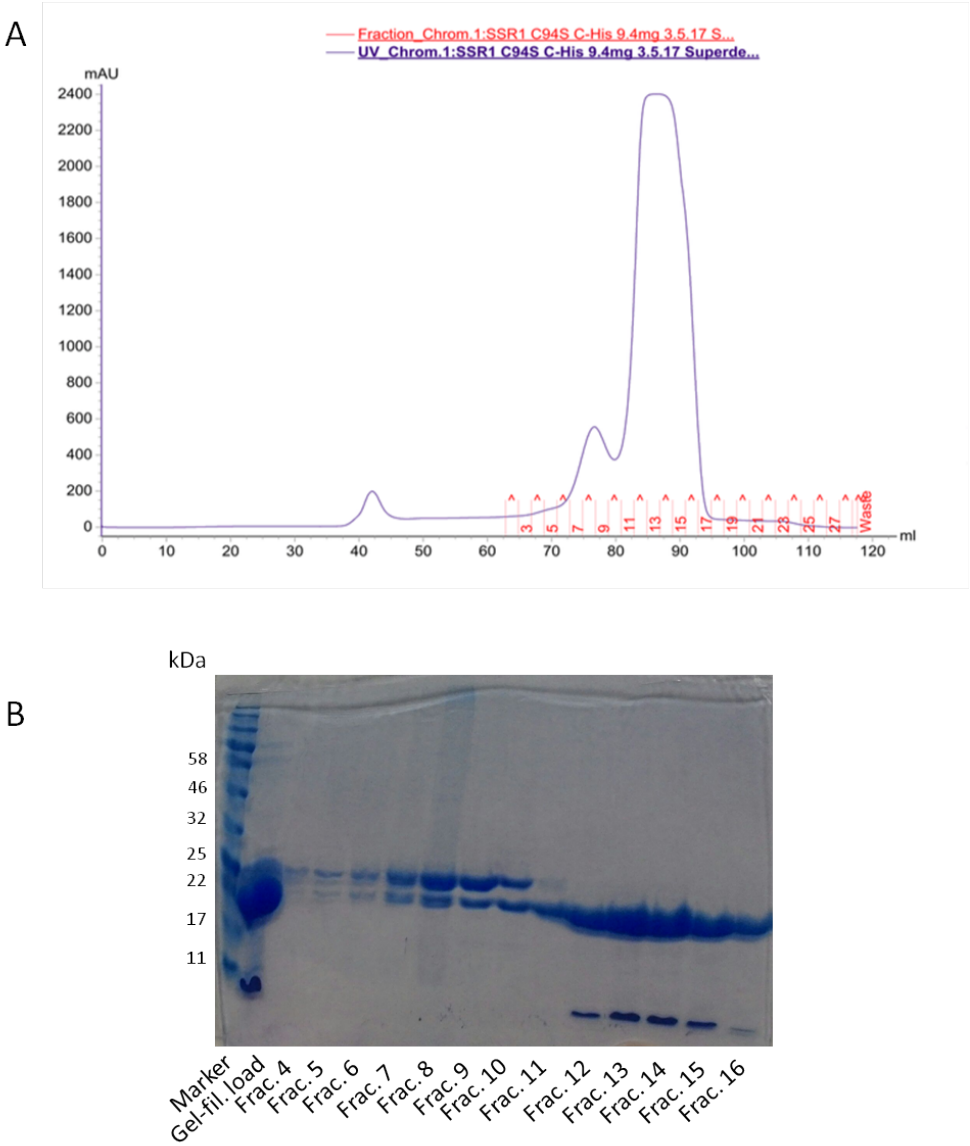


Fig. B.12 **Purification of the SSR1 C94S C-His proteins via gel-filtration column.** A) The chromatogram trace of the gel-filtration chromatography revealing the minor dimer peak and the major monomer peak. B) SDS-PAGE gel analysis of the resulting fractions.

B.3.4 Purification of the C-terminal His_6-tagged SSR1 S191A-S192G mutant protein

2 g of cells were lysed in 23 ml re-suspension buffer, and separated from the cell debris, yielding 2.3 mg ml^{-1} CFE which was subjected to Nickel-affinity chromatography. The eluted proteins were then concentrated before being applied to the gel-filtration column. The chromatograms of both purification steps and their corresponding SDS-PAGE results are shown in Figure B.13 & Figure B.14. Interestingly, although both the dimeric and the monomeric peaks could still be seen, the dimeric peak was reduced significantly compared to that of the SSR1 WT and C94S mutant. As discussed in Paper 3, analysis of an unclipped, intact monomer, would shed light on the reasons behind the clipping problem, therefore only the monomeric form of this mutant was used for crystallisation trials. The final sample was buffer exchanged into 10 mM Tris-HCl pH 7.5 and concentrated to 20 mg ml^{-1}.

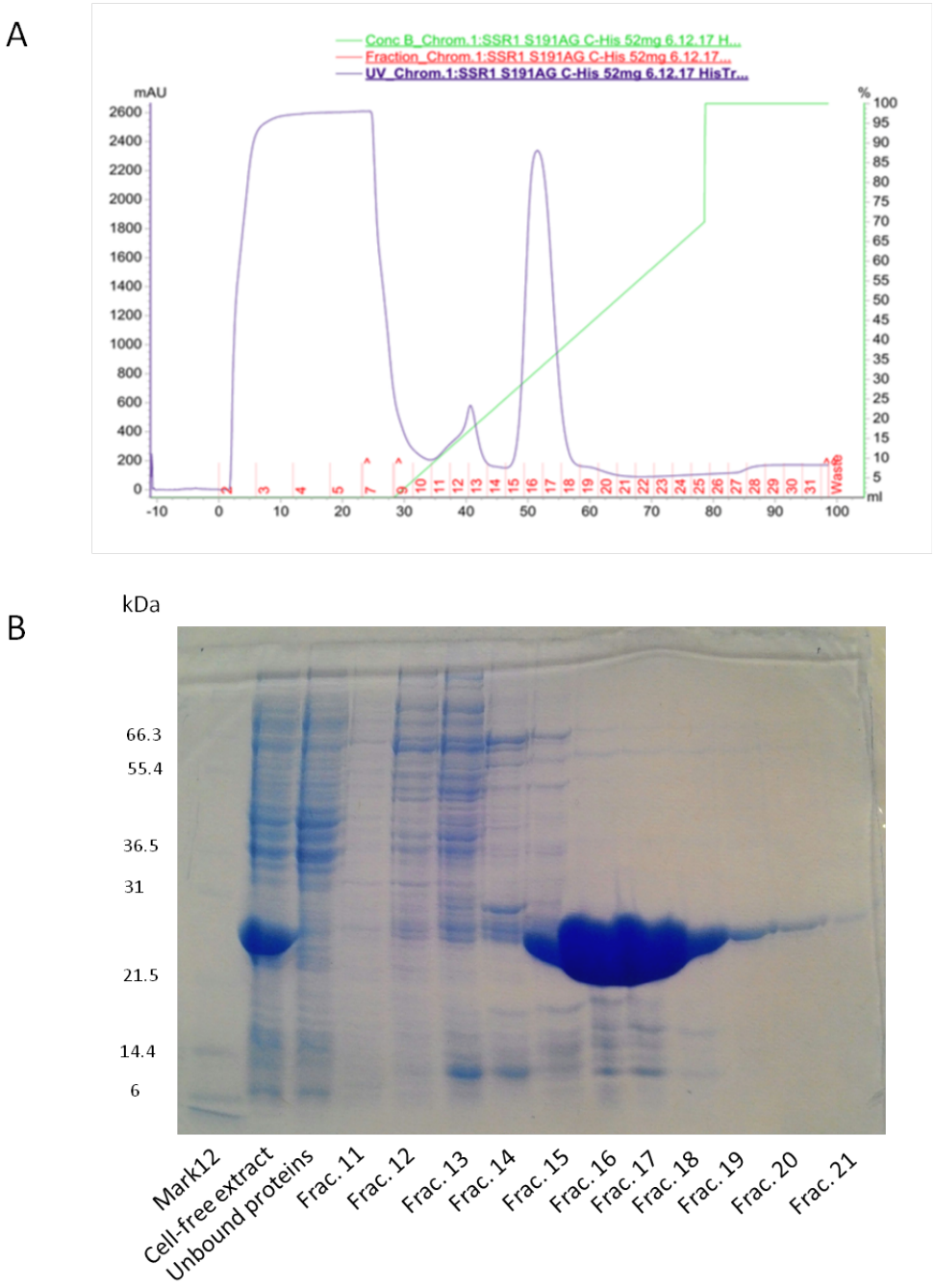


Fig. B.13 **Purification of the SSR1 S191A-S192G C-His proteins via Nickel-affinity column.** A) The chromatogram trace of the Nickel-affinity column. B) SDS-PAGE gel analysis of the resulting fractions.

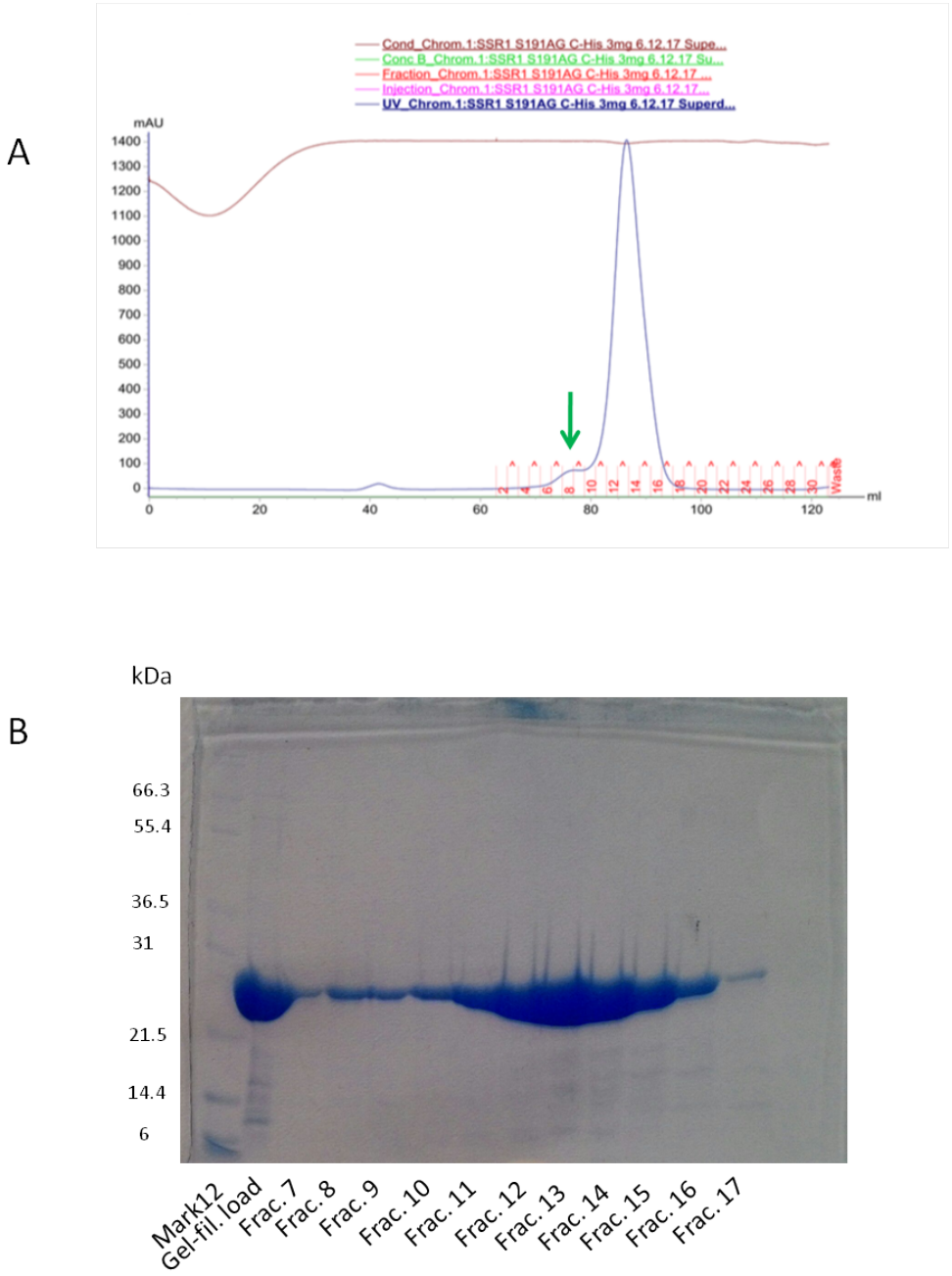


Fig. B.14 **Purification of the SSR1 S191A-S192G C-His proteins via gel-filtration column.** A) The chromatogram trace of the gel-filtration column. A reduced dimeric peak can be seen (green arrow). B) SDS-PAGE gel analysis of the resulting fractions.

B.3.5 Purification of the C-terminal His_6-tagged SSR1 S192A mutant protein

2 g of cells were lysed in 23 ml re-suspension buffer, and clarified to obtain 2.7 mg ml^{-1} CFE before applying to Nickel-affinity column. The eluted proteins were combined and concentrated to 2 ml before being applied to gel-filtration column and results are shown in Figure B.15. The SDS-PAGE analysis clearly showed a complete absence of clipping in both the dimeric and the monomeric form of the C-terminal His_6-tagged SSR1 S192A mutant. The final sample was concentrated to 20 mg ml^{-1} in 10 mM Tris pH8, 0.1M NaCl buffer and subjected to crystallisation trials.

B.3.6 Purification of the C-terminal His_6-tagged SSR1 S191A-S192A mutant protein

2 g of cells were lysed in 23 ml re-suspension buffer yielding 2.5 mg ml^{-1} CFE which was applied to the Nickel-affinity column. The fractions containing the protein were combined, concentrated, and applied to the gel-filtration column. The chromatograms with the accompanying SDS-PAGE gels are shown in Figure B.16. Similar to the SSR1 S192A mutant, clipping was almost certainly absent, and the sample containing the monomeric form was concentrated to 20 mg ml^{-1} in 10 mM Tris pH8, 0.1M NaCl buffer for crystallisation trials.

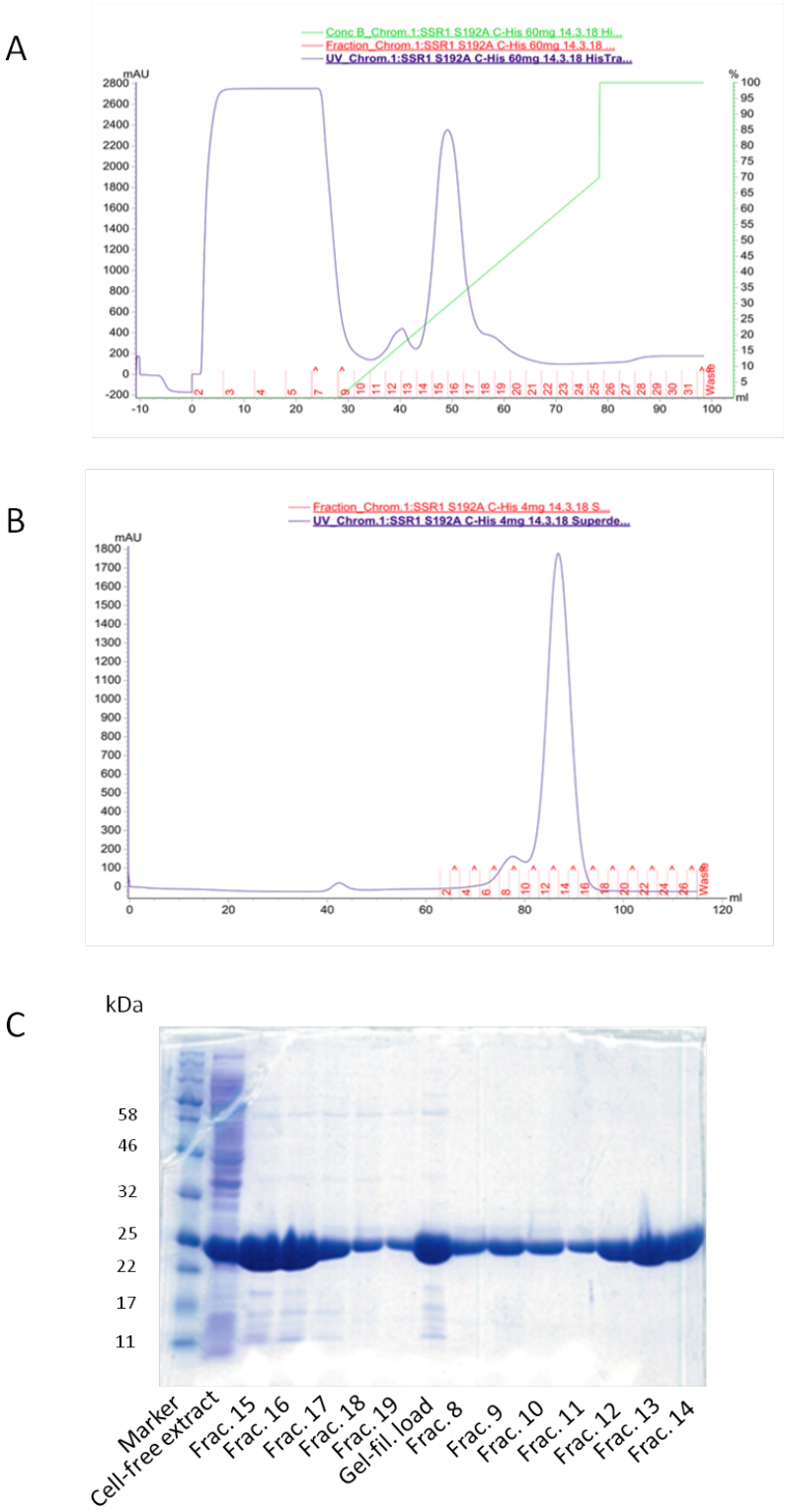


Fig. B.15 **Purification of the SSR1 S192A C-His proteins via Nickel-affinity followed by a gel-filtration column.** A) The chromatogram trace of the Nickel-affinity column. B) The chromatogram trace of the gel-filtration column revealing the dimer and monomer peaks. C) SDS-PAGE gel analysis of the resulting fractions.

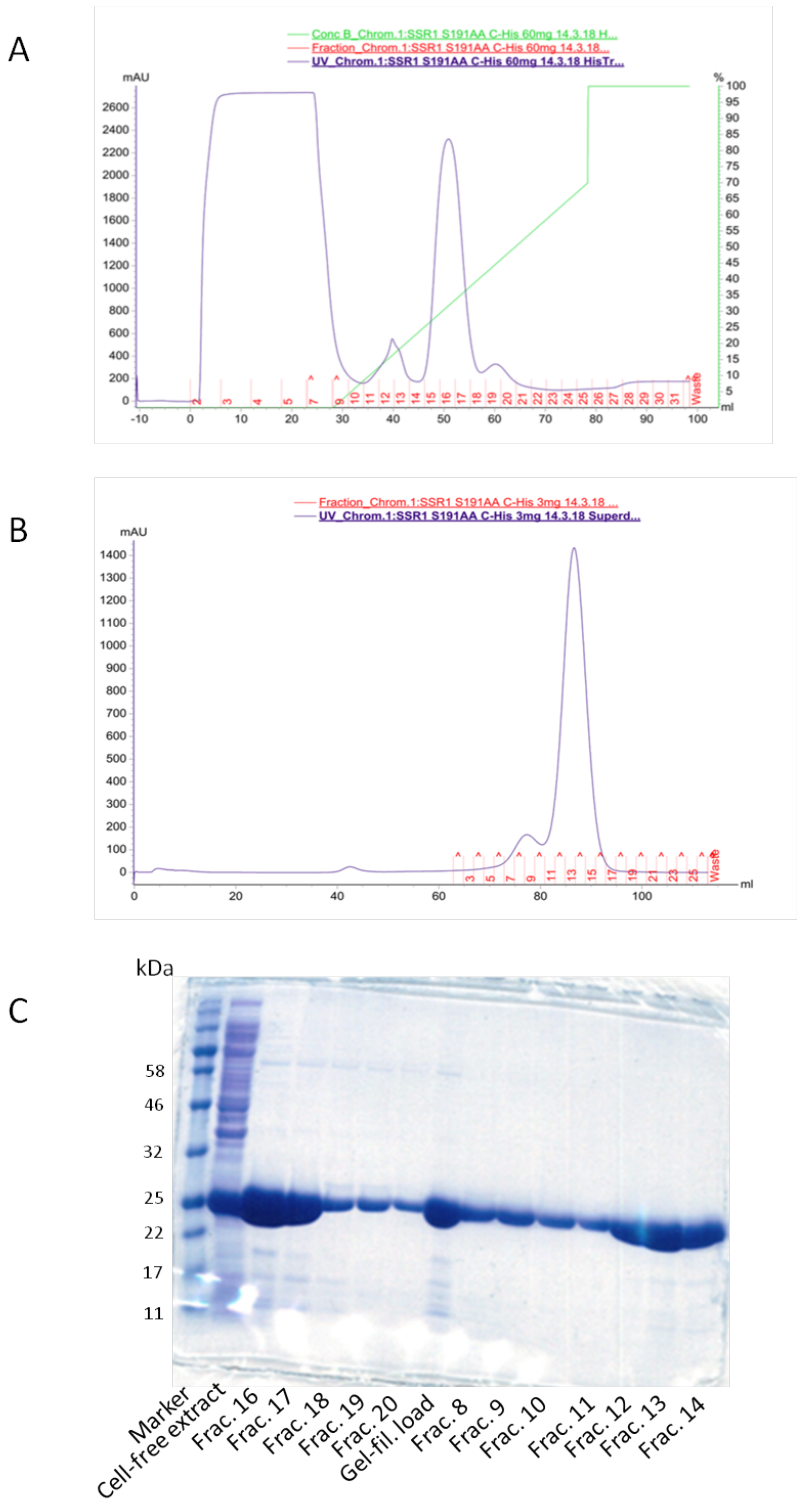


Fig. B.16 **Purification of the SSR1 S191A-S192A C-His proteins via Nickel-affinity followed by a gel-filtration column.** A) The chromatogram trace of the Nickel-affinity column. B) The chromatogram trace of the gel-filtration column. C) SDS-PAGE gel analysis of the resulting fractions.

B.3.7 Purification of the C-terminal His_6-tagged SSR1 S192C mutant protein

Initial test expression and the following large scale production of C-terminal His_6-tagged SSR1 S192C mutant proteins indicated a trace level of clipping which raised an interesting question about the possible role of the cysteine residue in the clipping. This was then investigated by purifying the protein to analyze clipping and dimerisation patterns. 2 g of cell paste were lysed in 20 ml re-suspension buffer, and following centrifugation, 3 mg ml^{-1} of CFE was obtained and applied to the Nickel-affinity column. The eluted proteins were then concentrated and applied to the gel-filtration column. The chromatograms with the resultant SDS-PAGE gel can be seen in Figure B.17 revealing a bigger peak corresponding to the MW of the dimer. Interestingly, through the course of purification, approximately 50 % of the proteins had been clipped, but the clipped proteins were resolved in both the monomeric and dimeric forms as can be seen in the SDS-PAGE gel of the gel-filtration column (Figure B.17). Fractions containing the monomeric proteins were then combined, and concentrated to 20 mg ml^{-1} in 10 mM Tris pH8, 0.1M NaCl buffer for crystallisation trials. The subsequent structure determination revealed the presence of a disulphide bond linking two clipped monomers suggesting that the dimer peak in the purification is a mixture of domain-swapped dimer of intact SSR1 and non-swapped clipped dimer where the subunits are linked by a S-S bond explaining the pattern of clipping seen during purification. This was not evident at the time.

B.3.8 Purification of the C-terminal His_6-tagged SSR1 S192T mutant protein

Similar to SSR1 S192C mutant, a small level of cleavage was observed in the initial expression trials of SSR1 S192T mutant proteins worthy of analysis by gel-filtration. SDS-PAGE analysis following purification indicated that the clipping persisted in both monomeric and dimeric form of the S192T mutant but with only approximately 10-20 % clipping (Figure B.18). The monomeric fractions were then combined and concentrated to 20 mg ml^{-1} for crystallisation trials. However, as of yet, no crystals have appeared which prompts further trials and optimisation.

B.3.9 Purification of the C-terminal His_6-tagged SSR1 S192V mutant protein

SSR1 S192V mutant was generated to test the possible involvement of the hydroxyl functional group at the γ position (in comparison with S192T) in clipping between residues 191 and 192. SDS-PAGE analysis following purification indicated that the clipping was completely abolished supporting the argument that the hydroxyl functional group at the γ position in the S192T mutant is crucial for cleaving the peptide bond (Figure B.19). Unfortunately, crystallisation trials on this mutant has not been carried out due to time constraint.

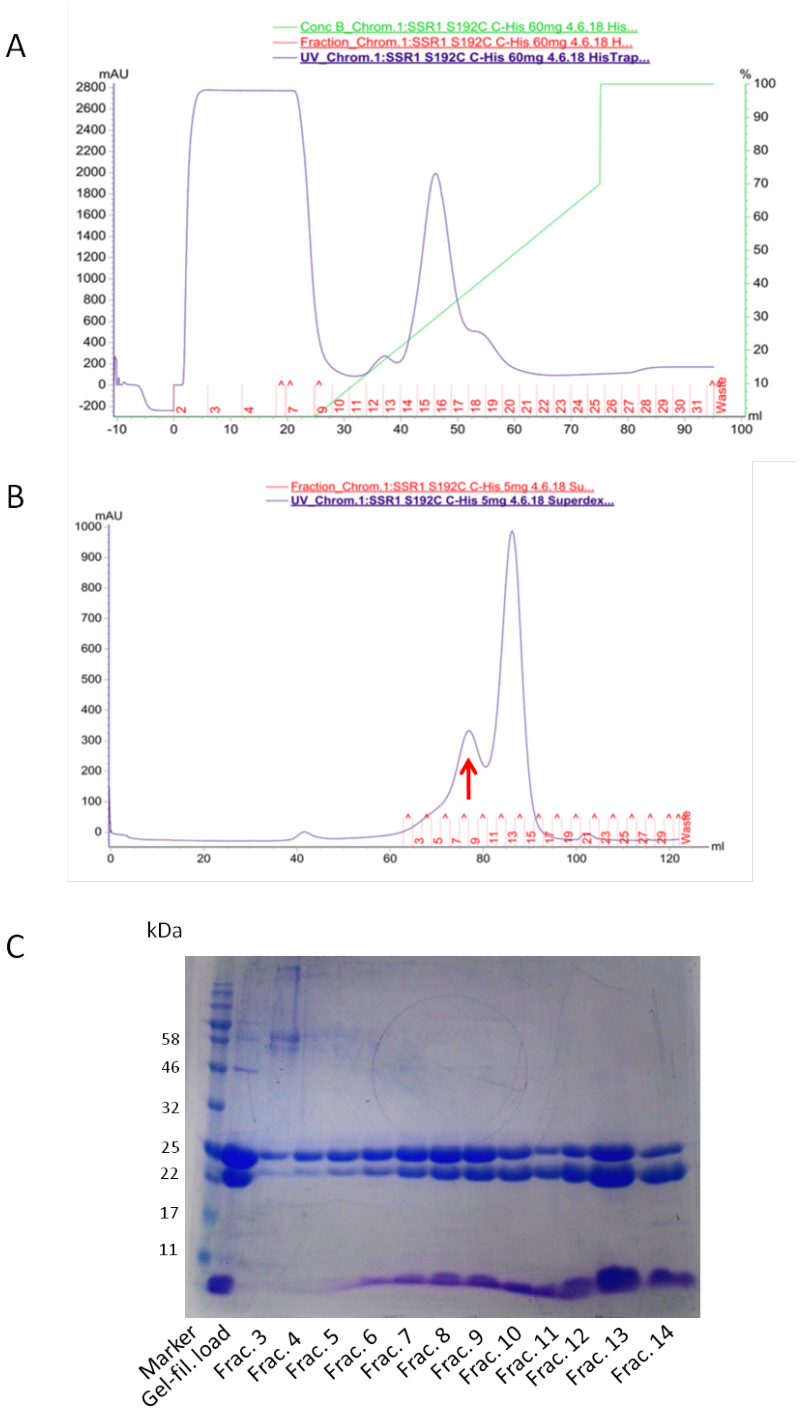


Fig. B.17 **Purification of the SSR1 S192C C-His proteins via Nickel-affinity followed by a gel-filtration column.** A) The chromatogram trace of the Nickel-affinity column. B) The chromatogram trace of the gel-filtration column. Red arrow indicates an increased amount of proteins eluted at dimeric MW. C) SDS-PAGE gel analysis of the fractions of the gel-filtration run.

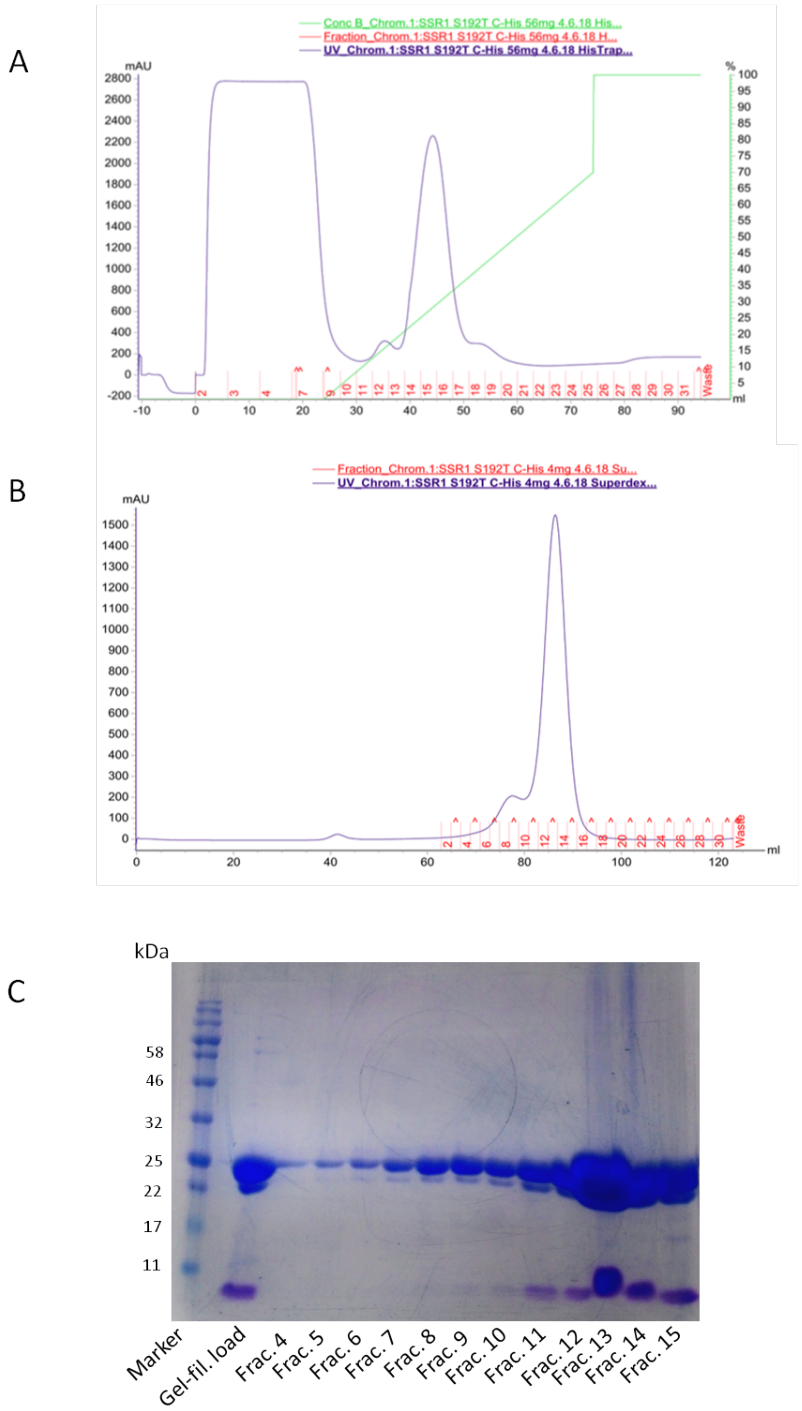


Fig. B.18 **Purification of the SSR1 S192T C-His proteins via Nickel-affinity followed by a gel-filtration column.** A) The chromatogram trace of the Nickel-affinity column. B) The chromatogram trace of the gel-filtration column. C) SDS-PAGE gel analysis of the fractions of the gel-filtration run.

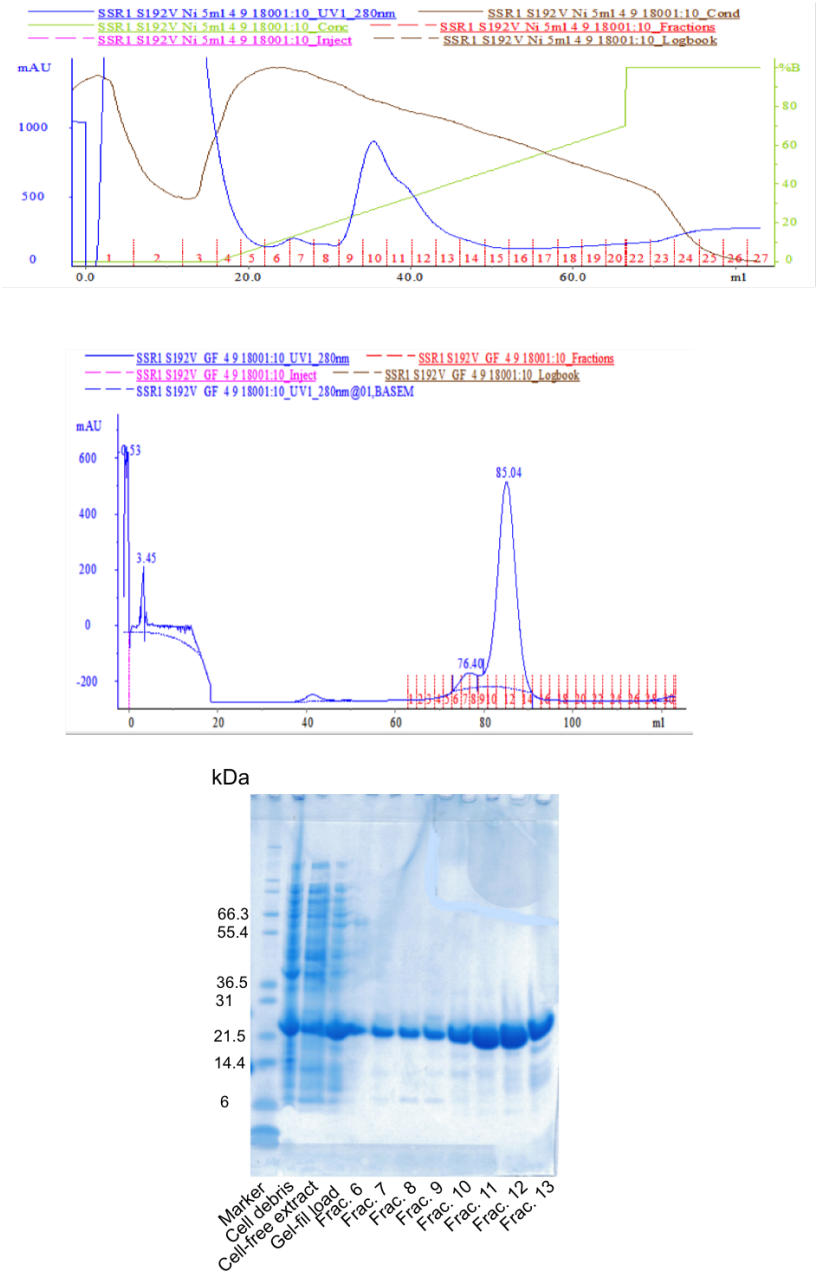


Fig. B.19 **Purification of the SSR1 S192V C-His proteins via Nickel-affinity followed by a gel-filtration column.** A) The chromatogram trace of the Nickel-affinity column. B) The chromatogram trace of the gel-filtration column. C) SDS-PAGE gel analysis of the fractions of the gel-filtration run showing no obvious signs of clipping.

B.4 Mass spectrometry analysis on the untagged and C-terminal His_6-tagged SSR1 WT protein samples

Prior to crystallisation, purified samples of the untagged and C-terminal His_6-tagged SSR1 WT protein were analysed by mass spectrometry (Simon Thorpe, Chemistry Mass Spectrometry Facility, Department of Chemistry) to identify the accurate MW of each polypeptide fragments to pinpoint the site of clipping. The resulting MW of the fragments from the monomeric sample of both variants were discussed in Paper 1. The mass spectrometry result of the dimeric C-terminal His_6-tagged SSR1 WT protein sample is shown in Figure B.20, revealing the MW of the intact polypeptide chain with some contamination of the clipped fragments. This had led to the identification of the clip site from MW calculations as discussed in Paper 1 and is demonstrated in Figure B.21.

B.5 Crystallisation of SSR1 proteins

The purified sample of the various SSR1 constructs were concentrated to the desired concentrations and subjected to crystallisation trials as described in Section 2.13. Initial crystals obtained from sitting drop plates were then analysed by X-ray diffraction, and depending on the quality of the diffraction, optimisation of the crystallisation conditions were carried out by hanging drop vapour diffusion. The crystallisation of SSR1 WT constructs were discussed in detail in Paper 2. Whilst, the details regarding the crystallisation of the SSR1 mutant constructs were omitted from Paper 3, and hence, will be described below.

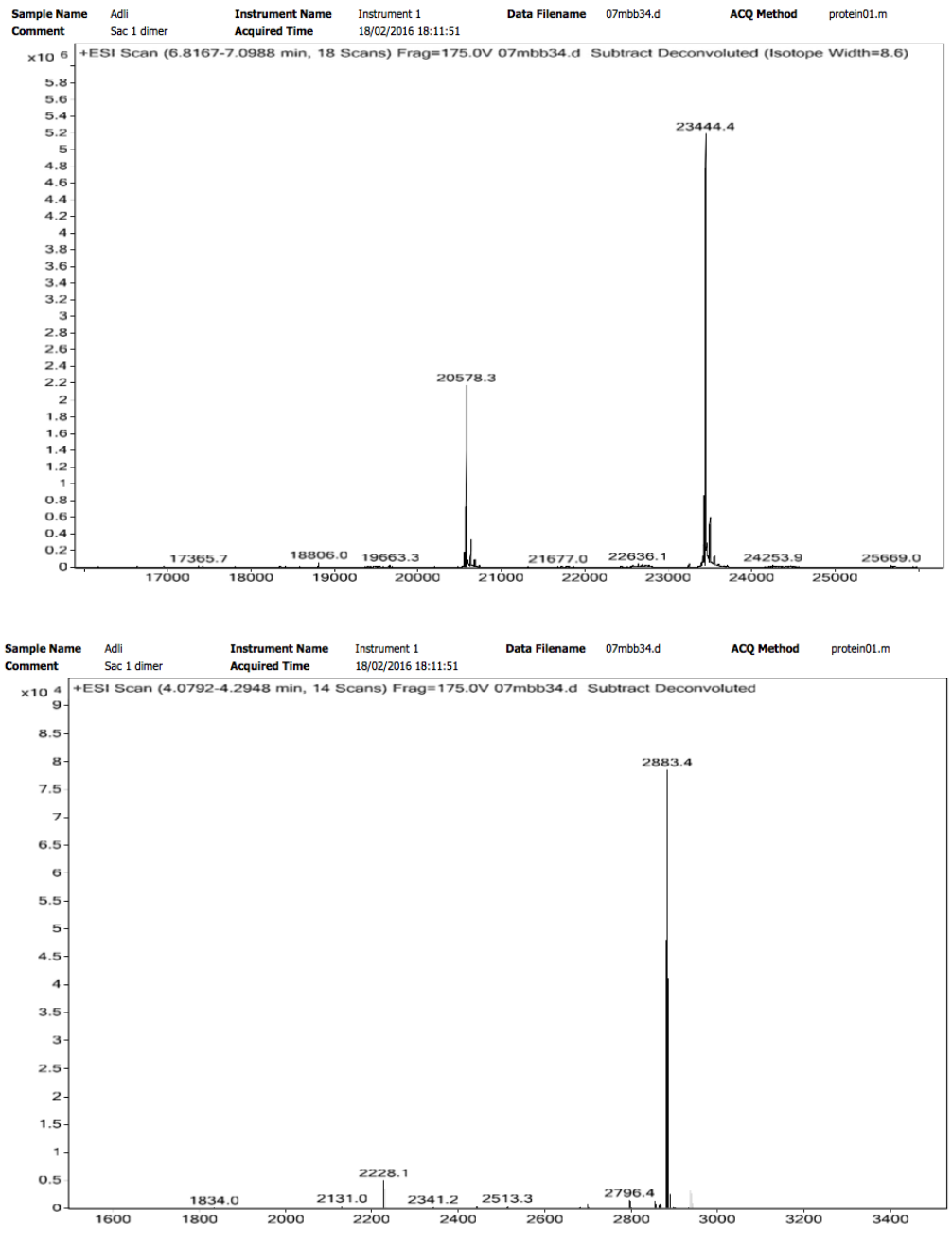


Fig. B.20 **Mass spectrometry analysis of the dimeric sample of C-terminal His_6-tagged SSR1 WT.** Top) The peaks corresponding to the unclipped SSR1 polypeptide (23444 Da) and the contaminating clipped fragment (20578 Da). Bottom) The shorter clipped fragment can also be identified (2883 Da).

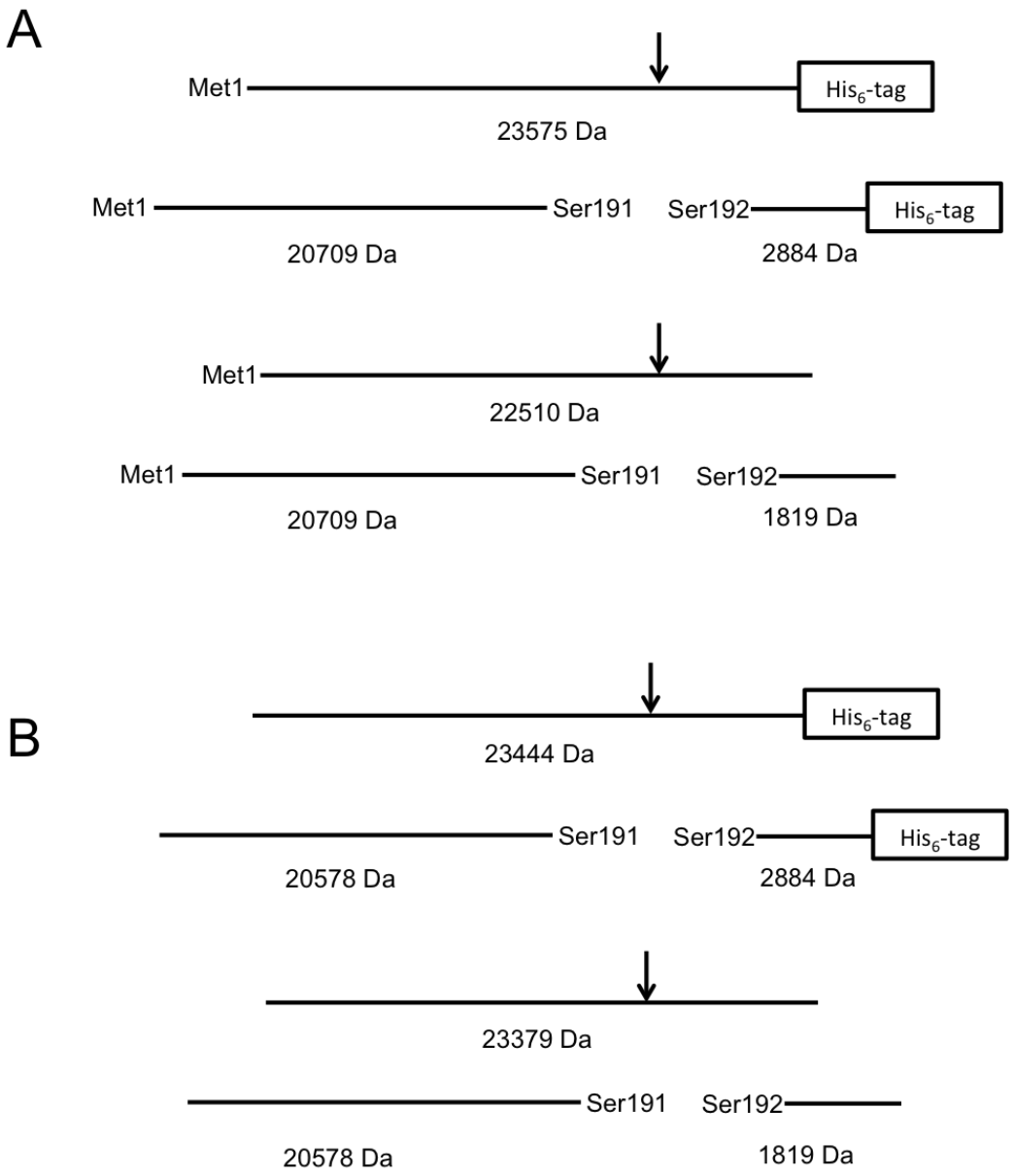


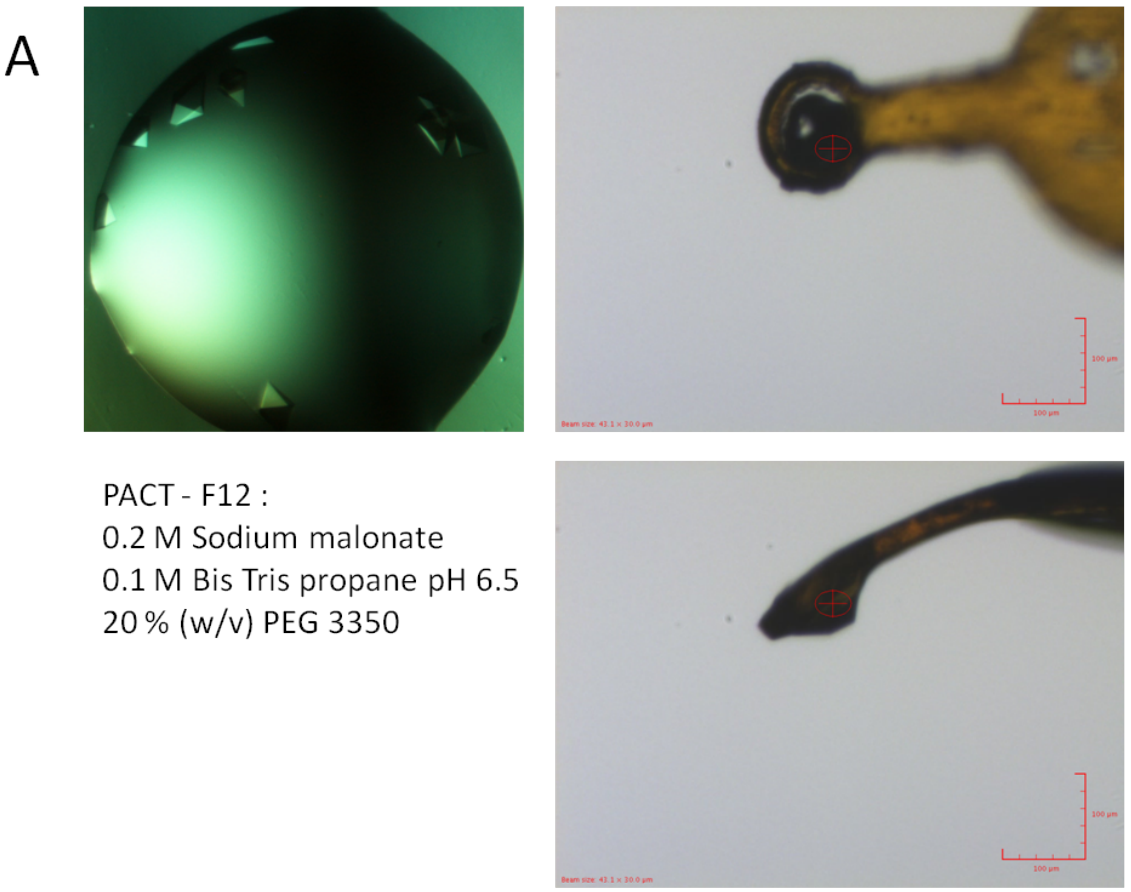
Fig. B.21 **MW calculations based on mass spectrometry analysis identified the clipping site.** A) Theoretical MW of SSR1 constructs including the N-terminal methionine residue and the resulting truncated products. B) Calculated MW following the removal of the N-terminal methionine residue. Clip site is indicated by arrows. MW of the fragment were calculated using ExPASy ProtParam online tool (https://web.expasy.org/cgi-bin/protparam/protparam).

B.5.1 Crystallisation of C-terminal His_6-tagged SSR1 C94S mutant

The monomeric sample of C-terminal His_6-tagged SSR1 C94S mutant crystallised in PACT - F12 (Figure B.22). The dimeric sample crystallised in pHClear - C9 and crystals were successfully optimised (Figure B.23). Crystals were then cryoprotected, mounted in cryo-condition and subjected to X-ray diffraction analysis as described in Section 2.14. The resulting diffraction pattern can be seen in Figure B.22 & Figure B.23.

B.5.2 Crystallisation of C-terminal His_6-tagged SSR1 clip site mutants

The sample containing the monomeric form of the C-terminal His_6-tagged SSR1 clip site mutants, namely the S191A-S192G, S192A, S191A-S192A, and S192C mutants were crystallised, and in some cases, optimised in the conditions described in Figure B.24, Figure B.25, Figure B.26, & Figure B.27, respectively. Crystals were then cryo-cooled in cryo-protectant, mounted and analysed by X-ray diffraction.



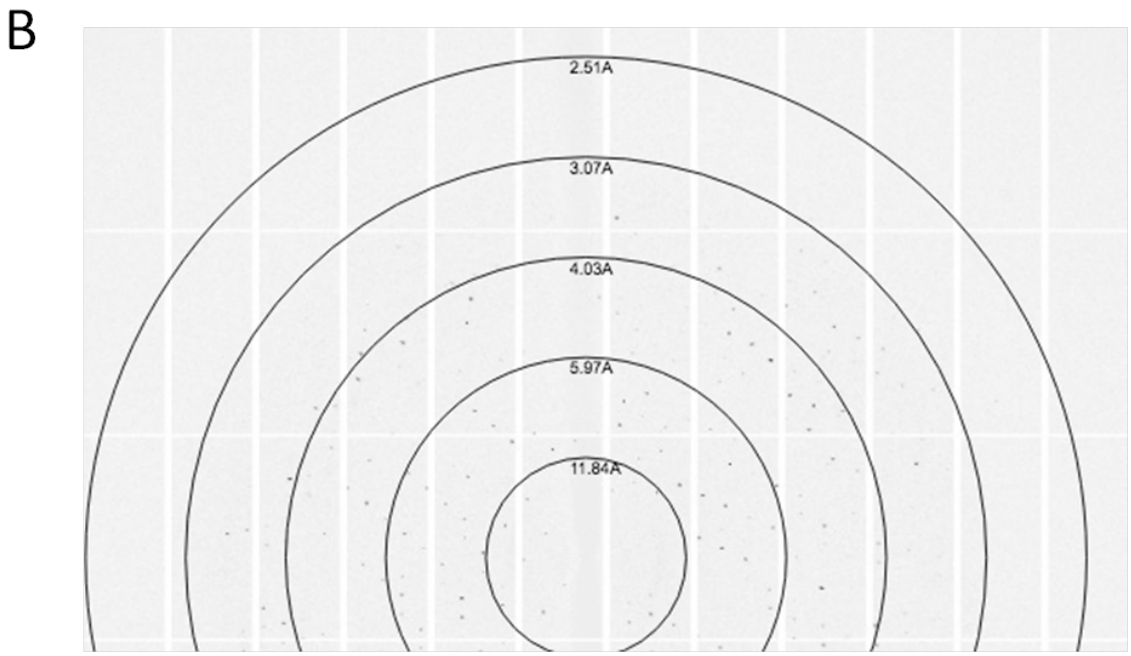
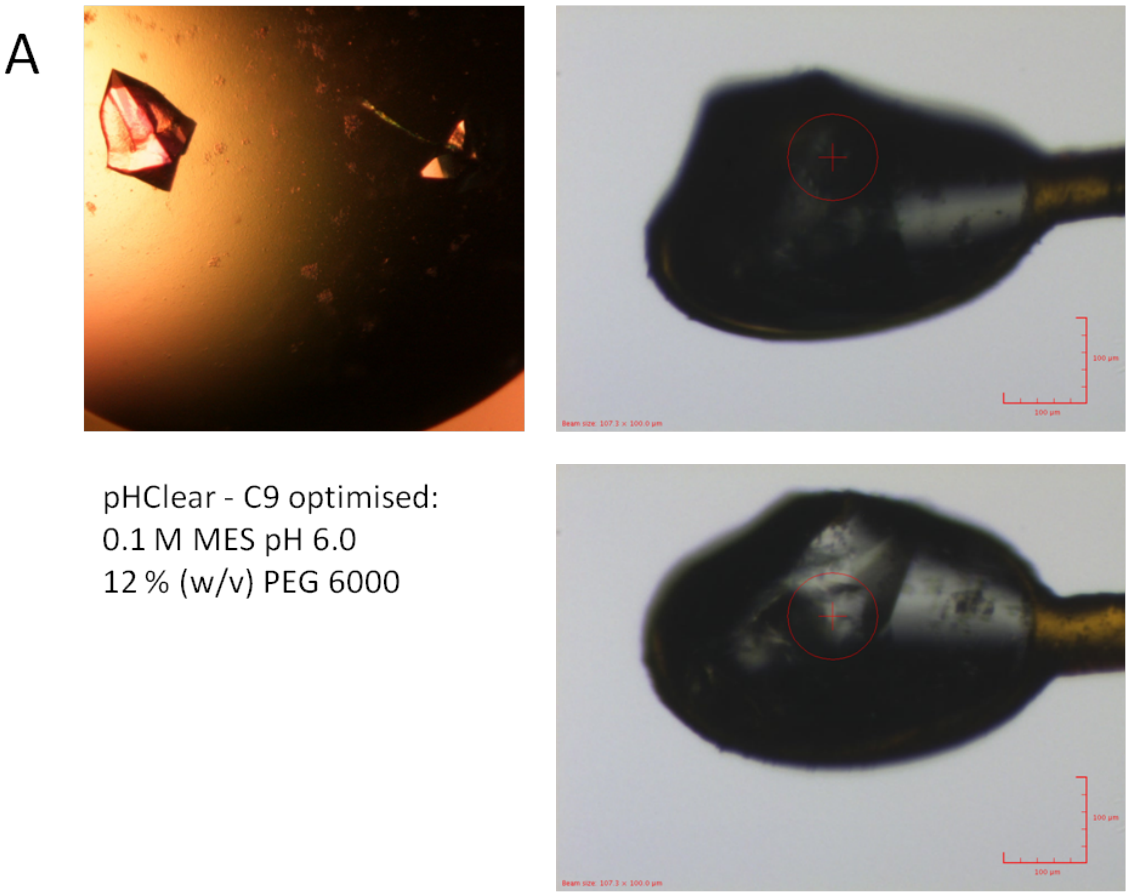


Fig. B.22 **Crystals of the monomeric form of SSR1 C94S C-His.** A) Crystals were obtained from sitting drop vapour diffusion in the corresponding crystallisation condition. Snapshots of the crystals mounted on a loop are also shown. B) A sample of diffraction image at 0.1^{o} rotation with resolution rings labeled.



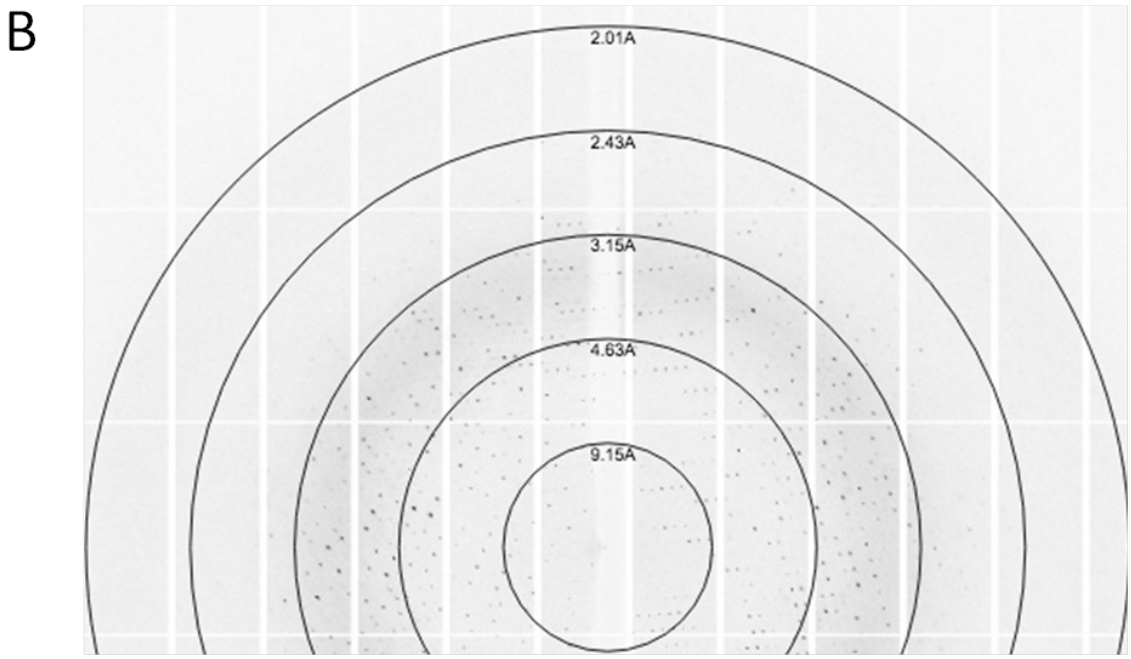


Fig. B.23 **Crystals of the dimeric form of SSR1 C94S C-His.** A) Crystals were obtained from an optimised crystallisation condition using hanging drop vapour diffusion. Snapshots of the crystals mounted on a loop are also shown. B) A sample of diffraction image at 0.1° rotation with resolution rings labelled.

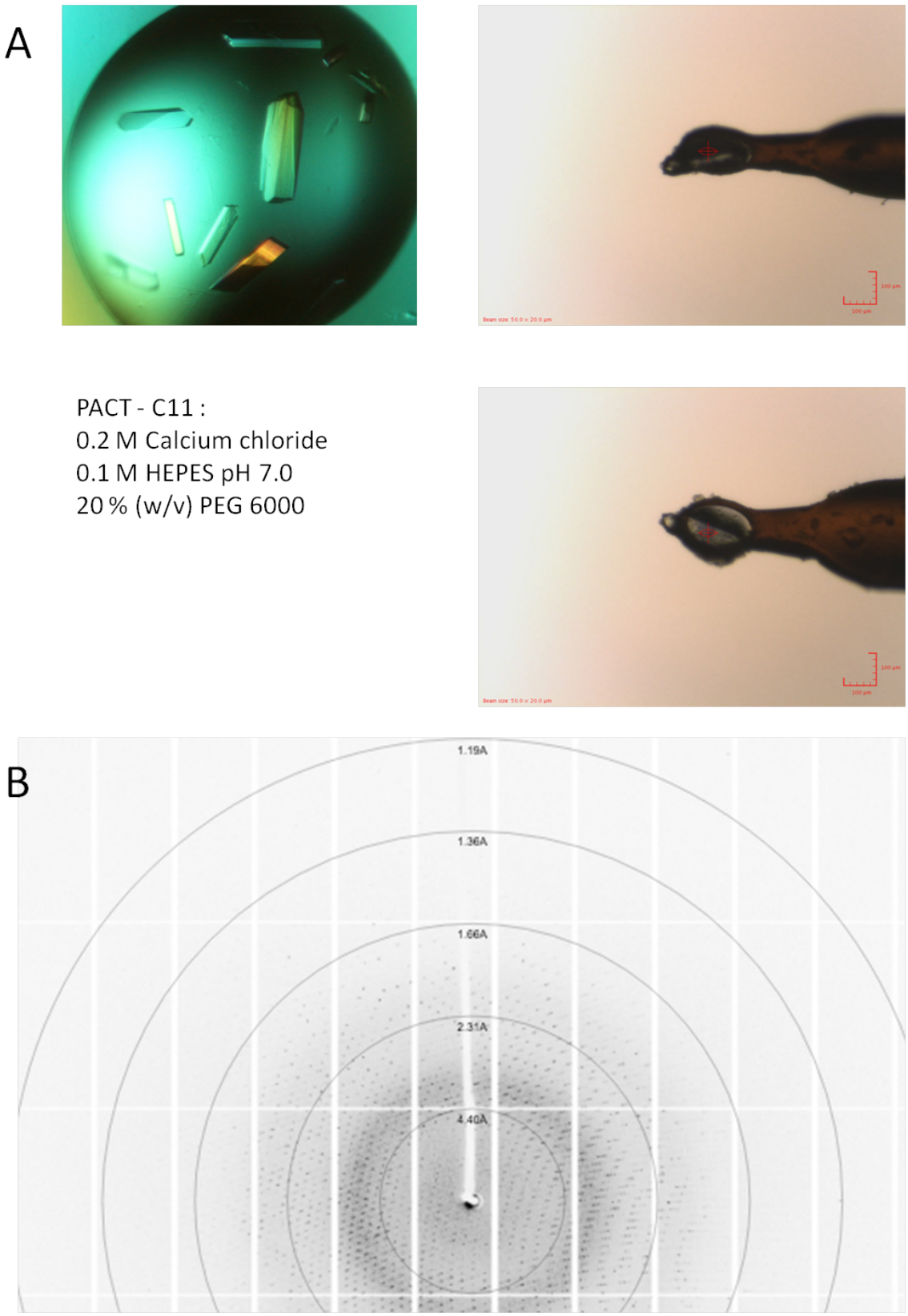


Fig. B.24 **Crystals of the monomeric form of SSR1 S191A-S192G C-His.** A) Crystals were obtained from a sitting drop plate with the crystallisation condition as shown. Snapshots of the crystals mounted on a loop are also shown. B) A sample of a diffraction image at 0.1^{o} rotation with resolution rings labelled.

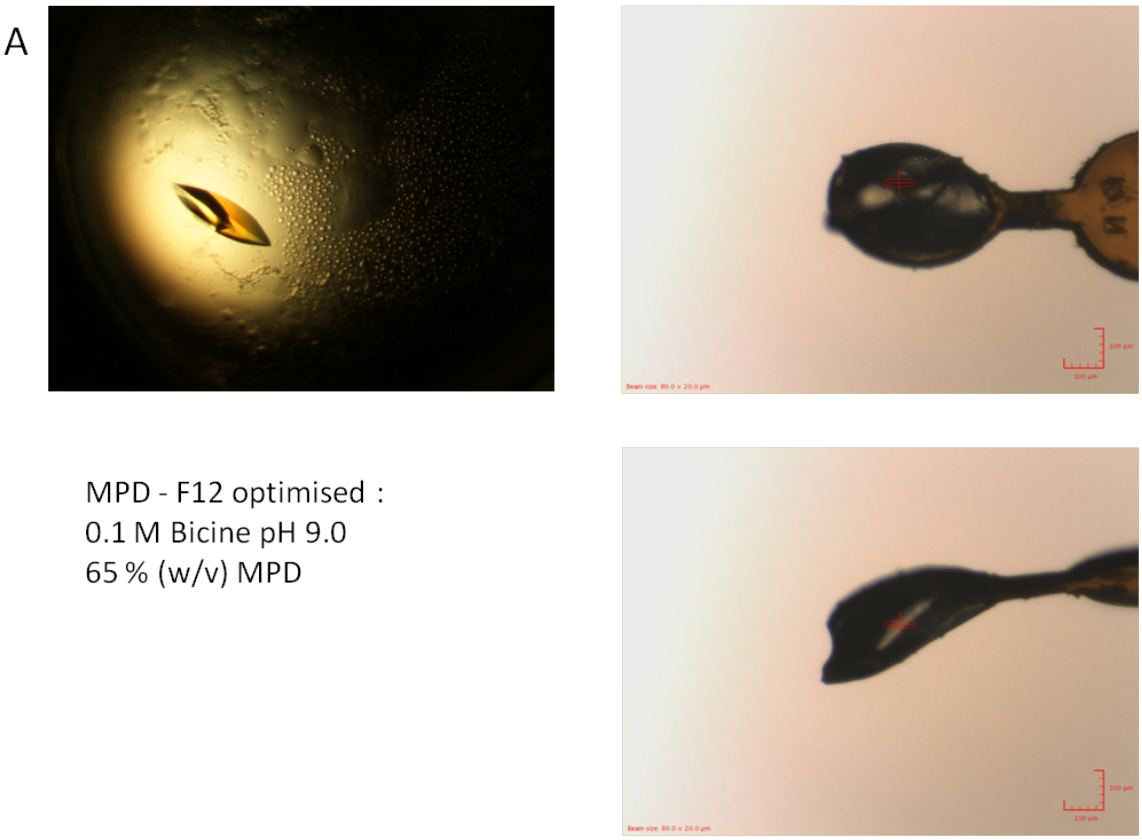
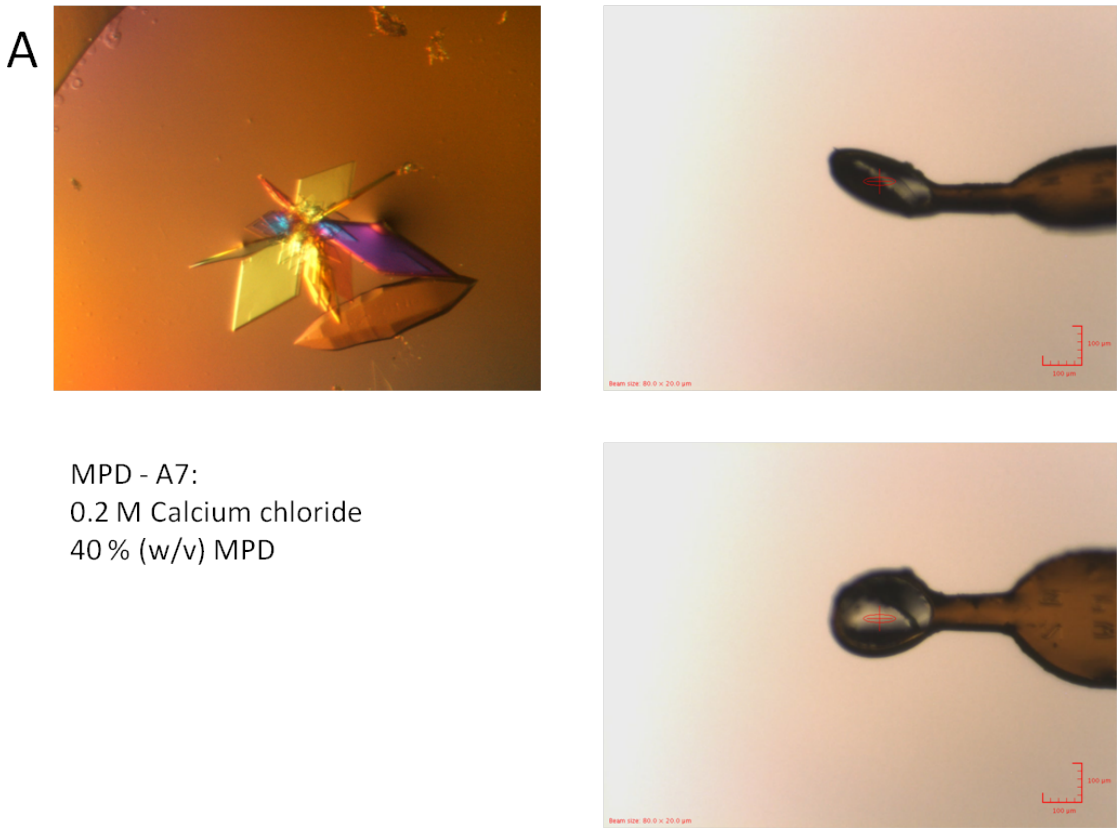


Fig. B.25 **Crystals of the monomeric form of SSR1 S192A C-His.** A) Crystals were obtained from an optimised crystallisation condition using hanging drop vapour diffusion. Snapshots of the crystals mounted on a loop are also shown. B) A sample of a diffraction image at 0.1^{o} rotation with resolution rings labelled.



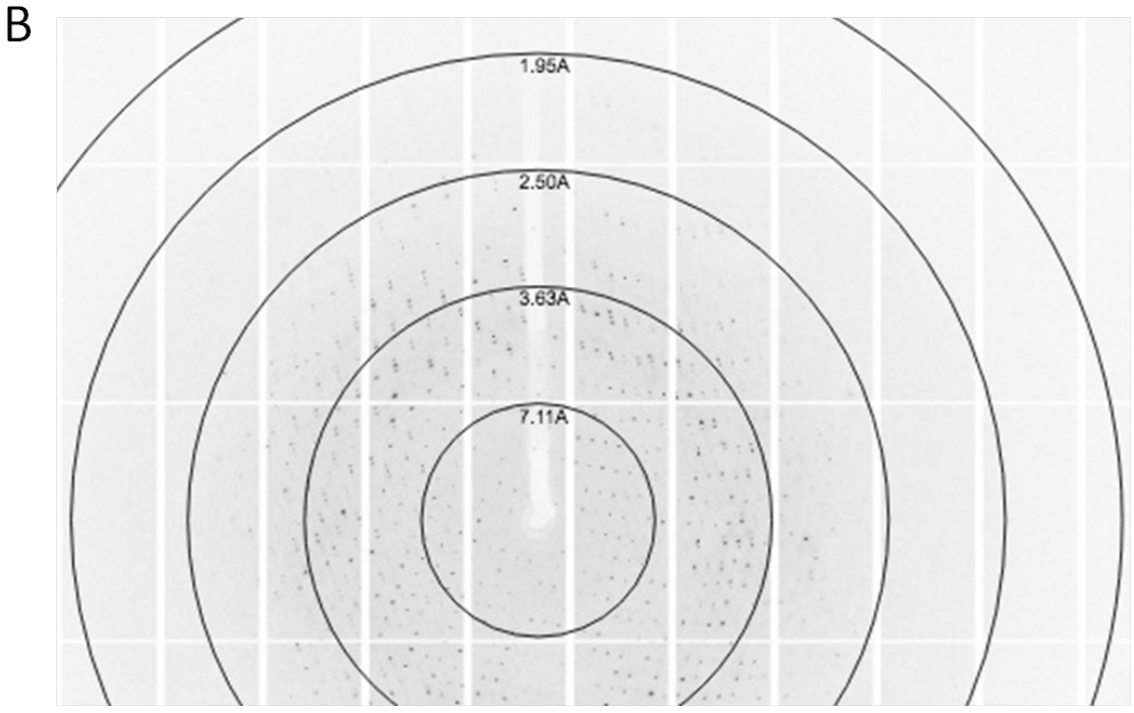
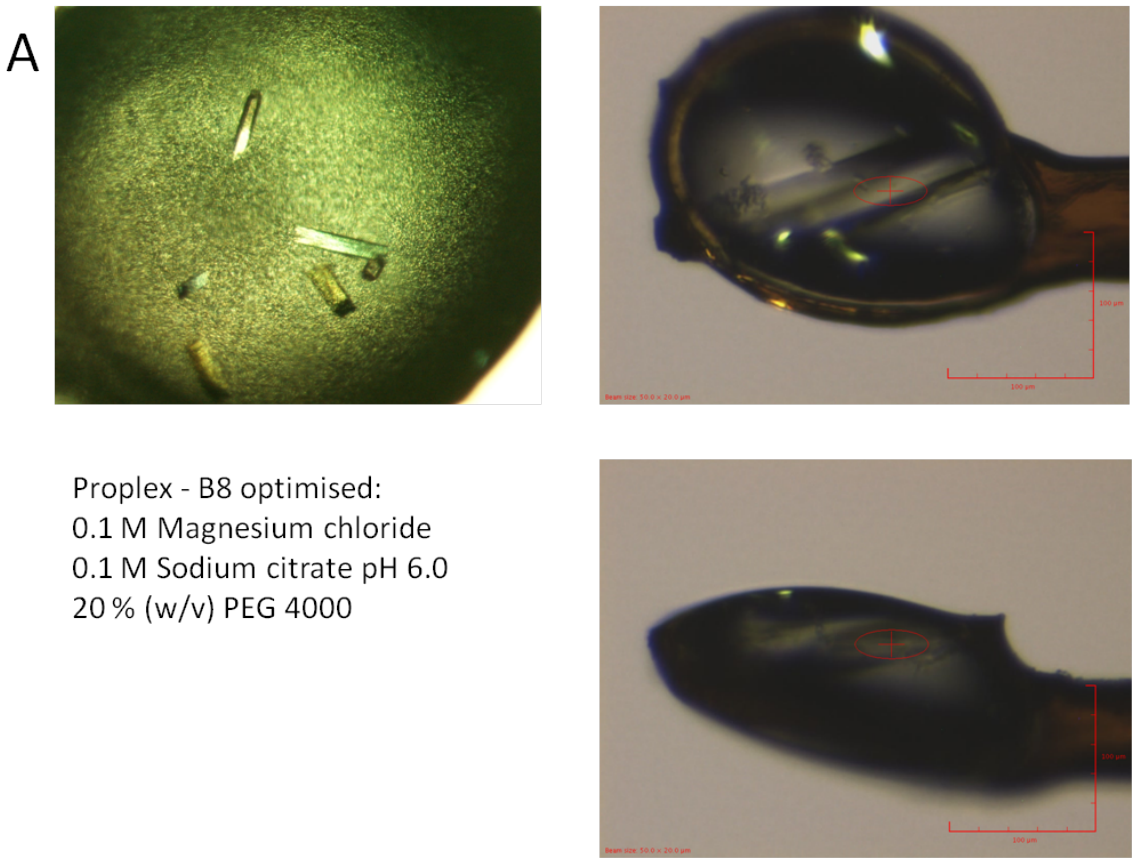


Fig. B.26 **Crystals of the monomeric form of SSR1 S191A-S192A C-His.** A) Crystals were obtained from a sitting drop plate with the crystallisation condition as shown. Snapshots of the crystals mounted on a loop are also shown. B) A sample of a diffraction image at 0.1^{o} rotation with resolution rings labelled.



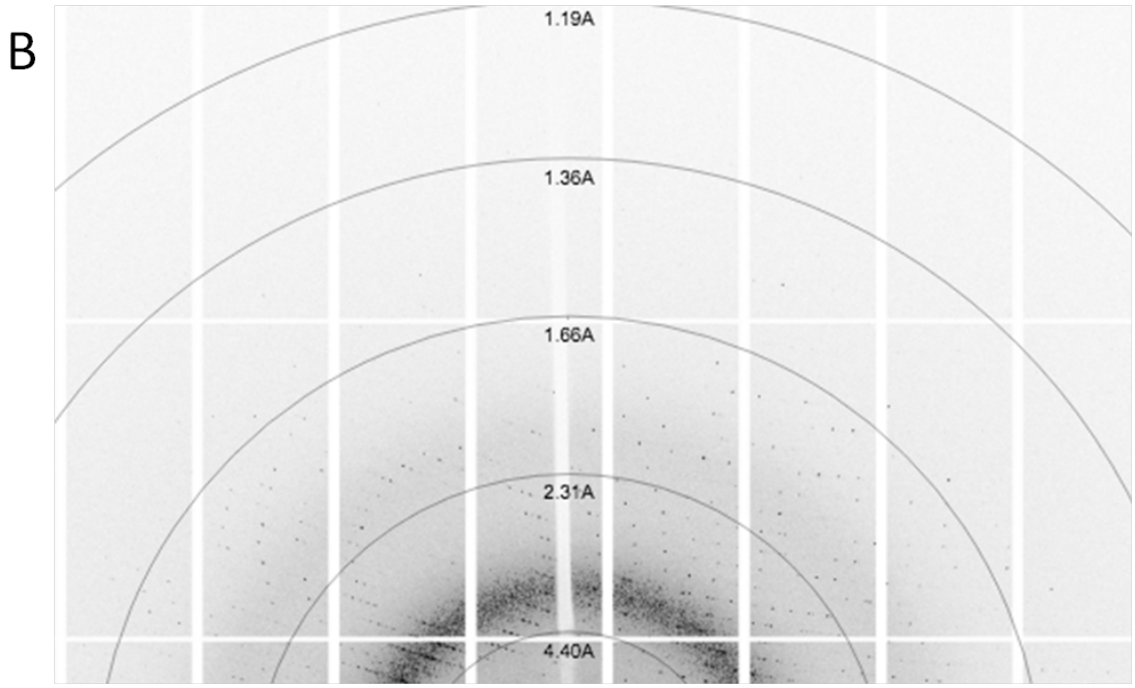


Fig. B.27 **Crystals of the monomeric form of SSR1 S192C C-His.** A) Crystals were obtained from an optimised crystallisation condition using hanging drop vapour diffusion. Snapshots of the crystals mounted on a loop are also shown. B) A sample of a diffraction image at 0.1° rotation with resolution rings labelled.

B.6 X-ray data collection and processing of SSR1 constructs

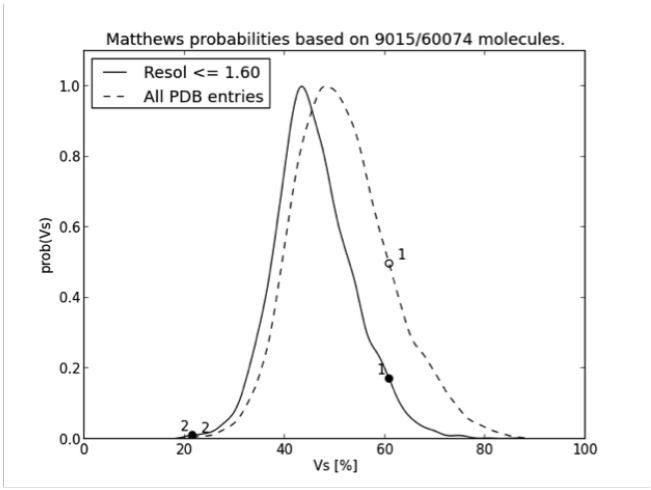
B.6.1 Data collection of the crystals of the untagged SSR1 WT monomer in two crystal forms and the C-terminal His_6-tagged SSR1 WT dimer

Details on data collection and processing statistics of the two crystal forms of the monomeric untagged SSR1 WT and the C-terminal His_6-tagged SSR1 WT dimer crystals are discussed in Paper 3. Crystals of form 1 of the SSR1 WT monomer, grown in the present of zinc ions, diffracted to atomic resolution with an anomalous mid-slope of 1.208 which allowed the anomalous scattering to be used for phase calculation. Structure determination of this crystal form will be discussed below and the final refined structure was then used as a search model in molecular replacement for the other crystal form of the monomeric SSR1 WT, and the His_6-tagged SSR1 WT dimer.

Calculations of Matthews coefficient and solvent content based on the unit cell parameters using an online tool MATTPROB (http://www.ruppweb.org/mattprob/default.html) (Weichenberger and Rupp, 2014) estimated that the crystals of form 2 most probably contain one molecule in the asymmetric unit with 60% solvent (Figure B.28). Whereas, the crystals from the dimeric SSR1 WT were estimated to have two molecules in the asymmetric unit with 41% solvent (Figure B.28).

A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.9391	0.9887	3.13	60.77	22500.00
2	0.0609	0.0113	1.57	21.53	45000.00



B

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0787	0.1977	4.20	70.72	22500.00
2	0.9213	0.8023	2.10	41.45	45000.00
3	0.0000	0.0000	1.40	12.17	67500.00

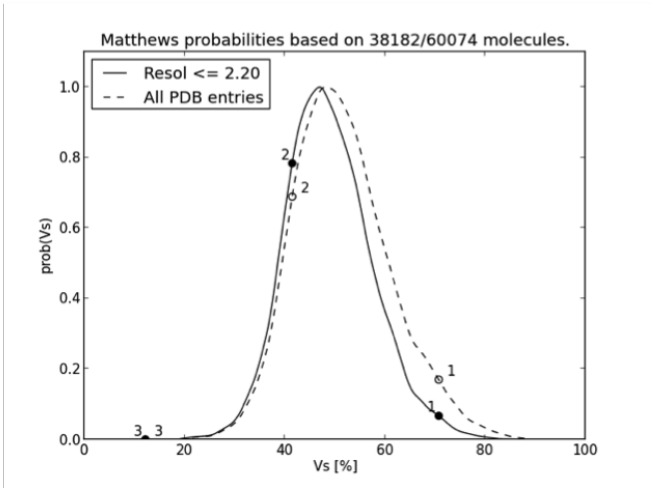


Fig. B.28 **MATTPROB calculations based on the unit cell parameters for the SSR1 WT crystals.** A) MATTPROB calculations on the SSR1 WT crystals of form 2 revealing the probability of the asymmetric unit content with the corresponding V_m and solvent percentage, shown in a table (top) and a graphical representation (bottom). B) The MATTPROB calculations on the crystals of SSR1 WT dimer.

B.6.2 Data collection of the crystals of the monomeric and the dimeric C-terminal His_6-tagged SSR1 C94S mutant

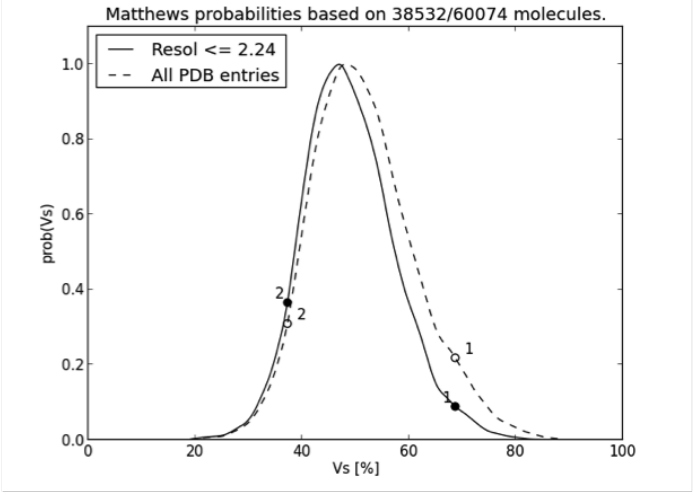
Details on data collection and reduction of these crystals are highlighted in Paper 3. MATTPROB estimated the crystals of the monomeric SSR1 C94S mutant to either contain one or two molecules in the asymmetric unit with 68 and 37 % solvent content, respectively (Figure B.29). Whereas for the crystals of the dimeric form, two or three molecules in the asymmetric unit would give a solvent content of 60 and 41 %, respectively (Figure B.29).

B.6.3 Data collection of the crystals of the various C-terminal His_6-tagged SSR1 clip site mutants

The MATTPROB analysis of the S191A-S192G, S191A-S192A, S192A and S192C mutants of SSR1 are shown in (Figure B.30). Crystals of the S191A-S192G mutant were estimated to have four molecules in the asymmetric unit resulting in a solvent content of 45 %. Crystals of the S191A-S192A mutant of SSR1 were evaluated to have either two or three molecules with solvent content of 58 or 36 %, respectively. Crystals of the S192A mutant of SSR1 most likely consist of one molecule with 60 % solvent. Finally, crystals of the S192C mutant most probably contain four molecules with 44 % solvent.

A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.1962	0.4140	3.92	68.63	22500.00
2	0.8038	0.5860	1.96	37.25	45000.00



B

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0020	0.0253	6.26	80.36	22500.00
2	0.2296	0.4195	3.13	60.73	45000.00
3	0.7605	0.5506	2.09	41.09	67500.00
4	0.0079	0.0047	1.57	21.45	90000.00

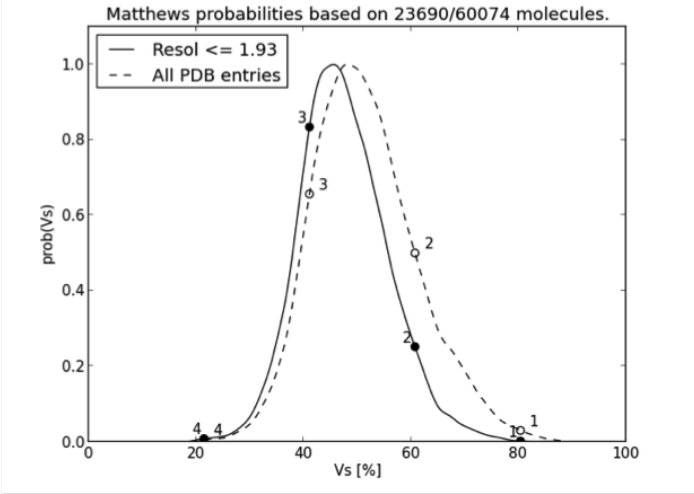
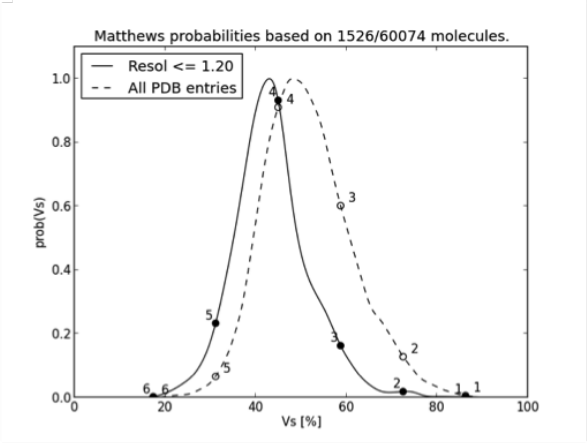


Fig. B.29 **MATTPROB calculations based on the unit cell parameters for the SSR1 C94S crystals.** A) MATTPROB calculations on the monomeric SSR1 C94S crystals revealing the probability of the asymmetric unit content with the corresponding V_m and solvent percentage, shown in a table (top) and a graphical representation (bottom). B) The MATTPROB calculations on the crystals of SSR1 C94S dimer.

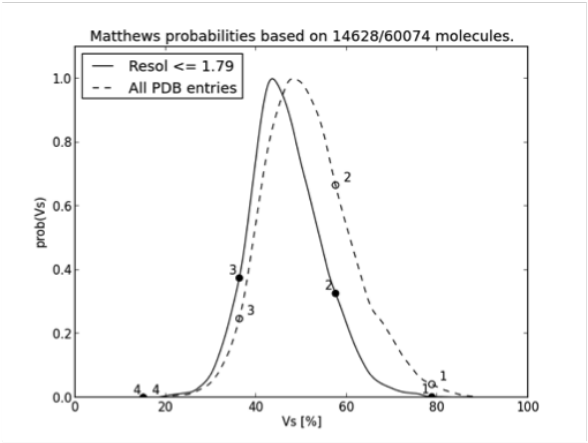
A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0018	0.0036	8.93	86.23	22500.00
2	0.0136	0.0751	4.47	72.45	45000.00
3	0.1207	0.3514	2.98	58.68	67500.00
4	0.6893	0.5312	2.23	44.91	90000.00
5	0.1726	0.0387	1.79	31.13	112500.00
6	0.0019	0.0000	1.49	17.36	135000.00



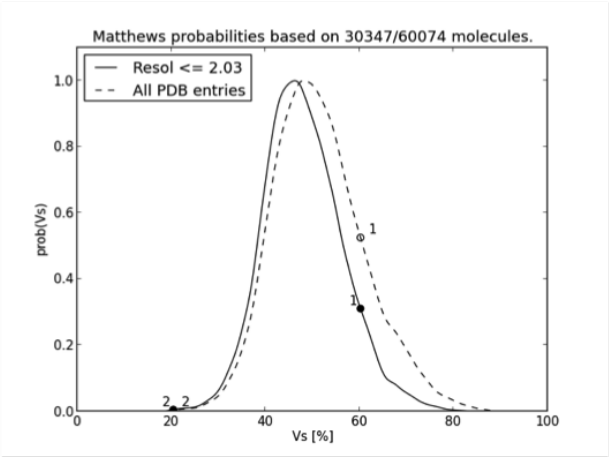
B

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0024	0.0434	5.79	78.76	22500.00
2	0.4644	0.6973	2.90	57.53	45000.00
3	0.5332	0.2594	1.93	36.29	67500.00
4	0.0000	0.0000	1.45	15.05	90000.00



C

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.9834	0.9931	3.09	60.21	22500.00
2	0.0166	0.0069	1.55	20.42	45000.00



D

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0018	0.0040	8.72	85.90	22500.00
2	0.0130	0.0846	4.36	71.80	45000.00
3	0.1419	0.3878	2.91	57.70	67500.00
4	0.7280	0.4996	2.18	43.60	90000.00
5	0.1150	0.0240	1.74	29.50	112500.00
6	0.0002	0.0000	1.45	15.40	135000.00

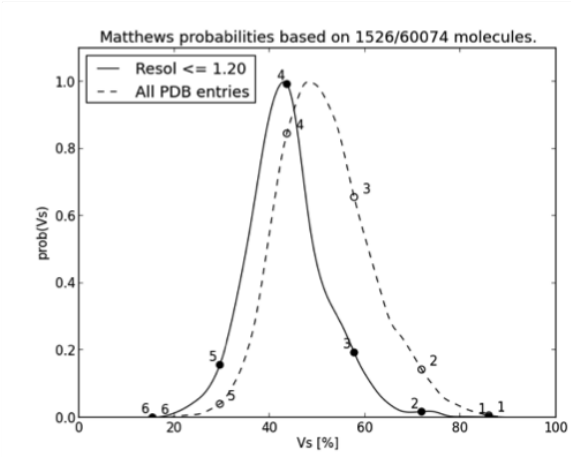


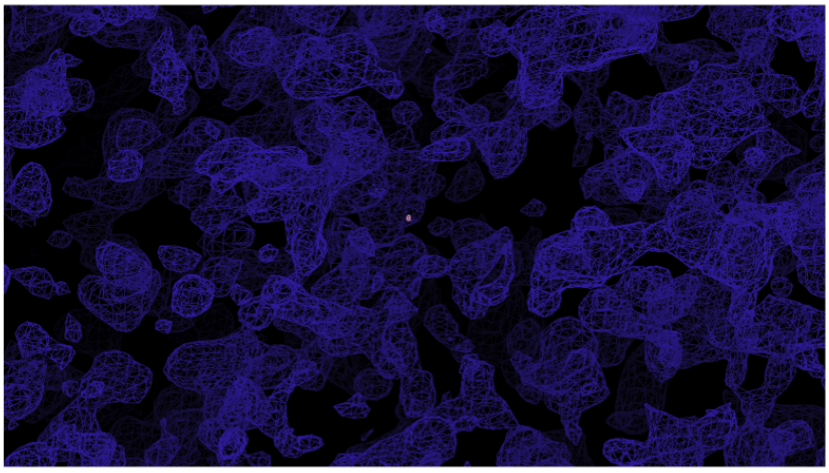
Fig. B.30 **MATTPROB calculations based on the unit cell parameters for the other SSR1 mutants crystals.** MATTPROB calculations on the crystals of A) SSR1 S191A-S192G, B) SSR1 S191A-S192A, C) SSR1 S192A and D) SSR1 S192C mutant crystals revealing the probability of the asymmetric unit content with the corresponding V_m and solvent percentage, shown in a table (top) and a graphical representation (bottom).

B.7 Structure determination of the untagged SSR1 WT monomer of crystal form 1

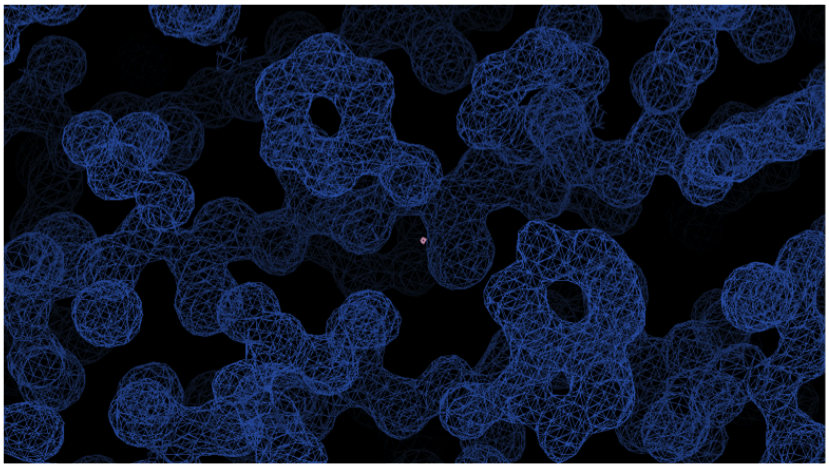
X-ray diffraction data from the crystals of form 1 of the SSR1 WT monomer were collected at X-ray wavelength 0.9763 Å which yielded diffraction to 1.09 Å resolution. Data processing statistics indicated an anomalous normal probability mid-slope value of 1.208, suggesting that the crystals contained anomalous scattering atoms from the precipitant (0.25 M $ZnSO_4$). The hkl2map GUI with built in SHELXC/D/E package (Sheldrick, 2008) was used to calculate initial phases. Initial SHELXC analysis indicated a strong anomalous signal with d"/sig > 0.80 across the whole dataset to 1.09 Å. Subsequent SHELXD identified 11 anomalous scatterers with occupancy ranging from 0.11 to 1.00 (Figure B.31). The heavy atoms sites were then fed into SHELXE to calculate the initial phases and carry out several rounds of density modification which was paired with poly-alanine model auto-tracing. Both enantiomorphs were tested, and due to the atomic resolution nature of the dataset, the inverted hand could easily be validated to be correct as opposed to the original hand, which clearly showed the the electron density of side chains that could be sequenced to be compared with that of SSR1. (Figure B.31).

The final poly-alanine model generated by SHELXE consisted of 8 polypeptide chains with each chain containing residues in the range of 7 to 45 (Figure B.32). Manual model building and refinement was carried out using *Coot* (Emsley et al., 2010) and REFMAC5 (Murshudov et al., 2011) and the resultant model along with the heavy atom sites were manually observed and validated. Analysis of the positions of the heavy atoms indicated that they are not located within the side chain atoms, therefore suggesting their origin being the zinc ions in the precipitant (Figure B.33). Further analysis of the geometry, the identity of the ligands and the corresponding metal-ligand distances provided clear evidence supporting the binding of zinc ions.

```
HETATM    1 ZN01 ZN   Z    1       9.401   19.600   87.883  1.00 15.79          ZN
HETATM    2 ZN02 ZN   Z    2      -9.929   27.405   71.662  0.90 15.79          ZN
HETATM    3 ZN03 ZN   Z    3     -16.989   29.426   87.217  0.41 15.79          ZN
HETATM    4 ZN04 ZN   Z    4     -20.718   39.176   69.047  0.67 15.79          ZN
HETATM    5 ZN05 ZN   Z    5      -9.562   16.562   87.217  0.33 15.79          ZN
HETATM    6 ZN06 ZN   Z    6     -14.324   39.375   77.180  0.47 15.79          ZN
HETATM    7 ZN07 ZN   Z    7     -13.196   43.637   73.481  0.28 15.79          ZN
HETATM    8 ZN08 ZN   Z    8     -14.095   18.710   61.838  0.21 15.79          ZN
HETATM    9 ZN09 ZN   Z    9       7.262   33.708   98.799  0.17 15.79          ZN
HETATM   10 ZN10 ZN   Z   10      -4.331   38.231   74.952  0.12 15.79          ZN
HETATM   11 ZN11 ZN   Z   11     -14.300   20.518   66.471  0.11 15.79          ZN
END
```

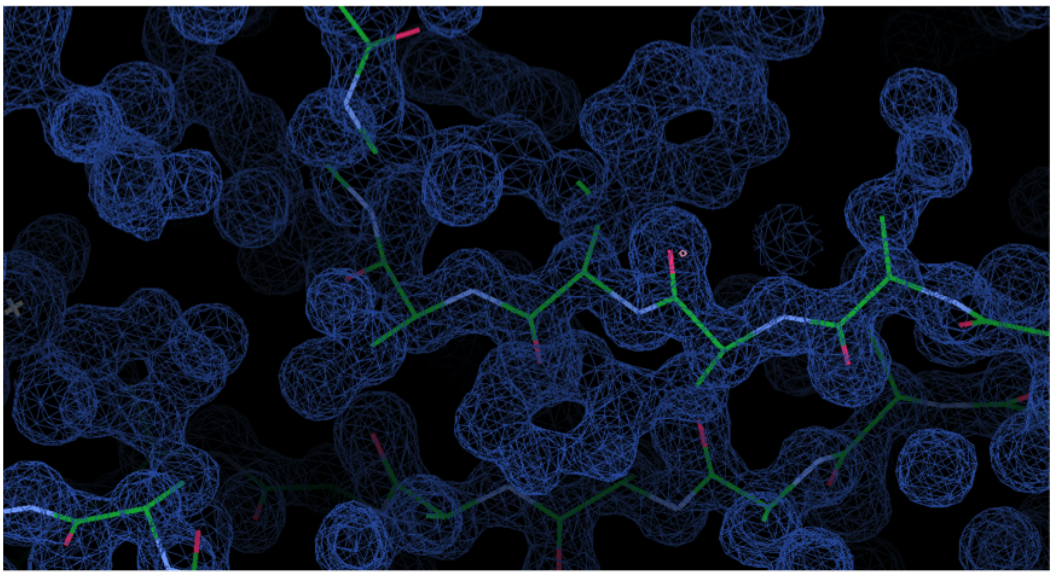


Original hand

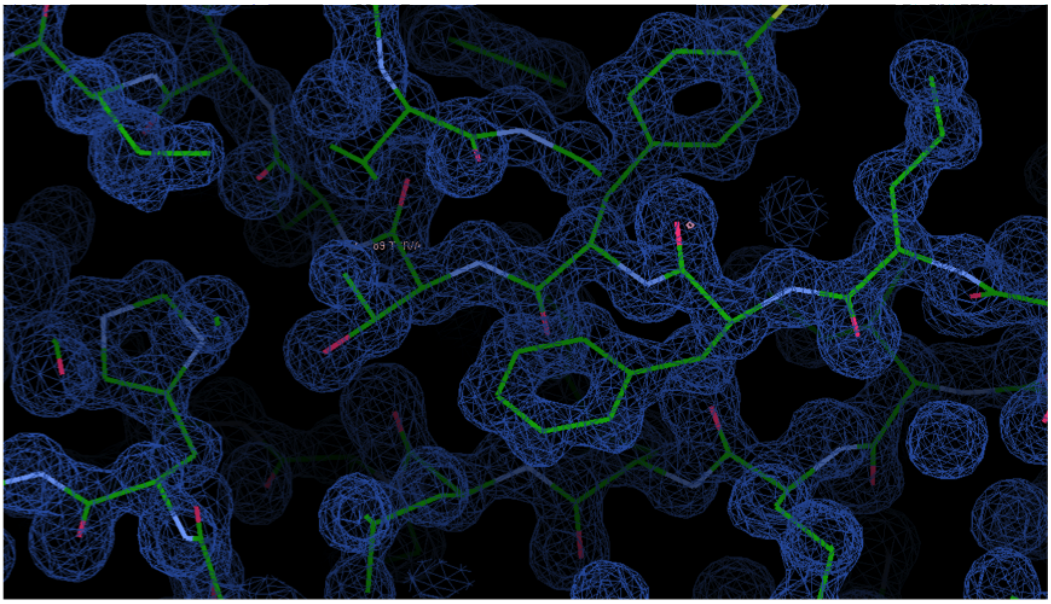


Inverted hand

Fig. B.31 **Structural determination of SSR1 WT crystals of form 1 via SHELXC/D/E.** 11 zinc sites were identified with the occupancy of each atom highlighted in red box (top). The initial electron density of the original hand (middle), and the initial electron density of the inverted hand showing clear view of the backbone and side chain electron density (bottom).



Initial model



Final model

Fig. B.32 **Model output from SHELXE of the SSR1 WT crystals of form 1 and subsequent model building.** Poly-alanine traces of the initial model built by SHELXE (top) which were subsequently completed manually on *Coot* (bottom). The electron density maps were contoured to 1.0 σ level.

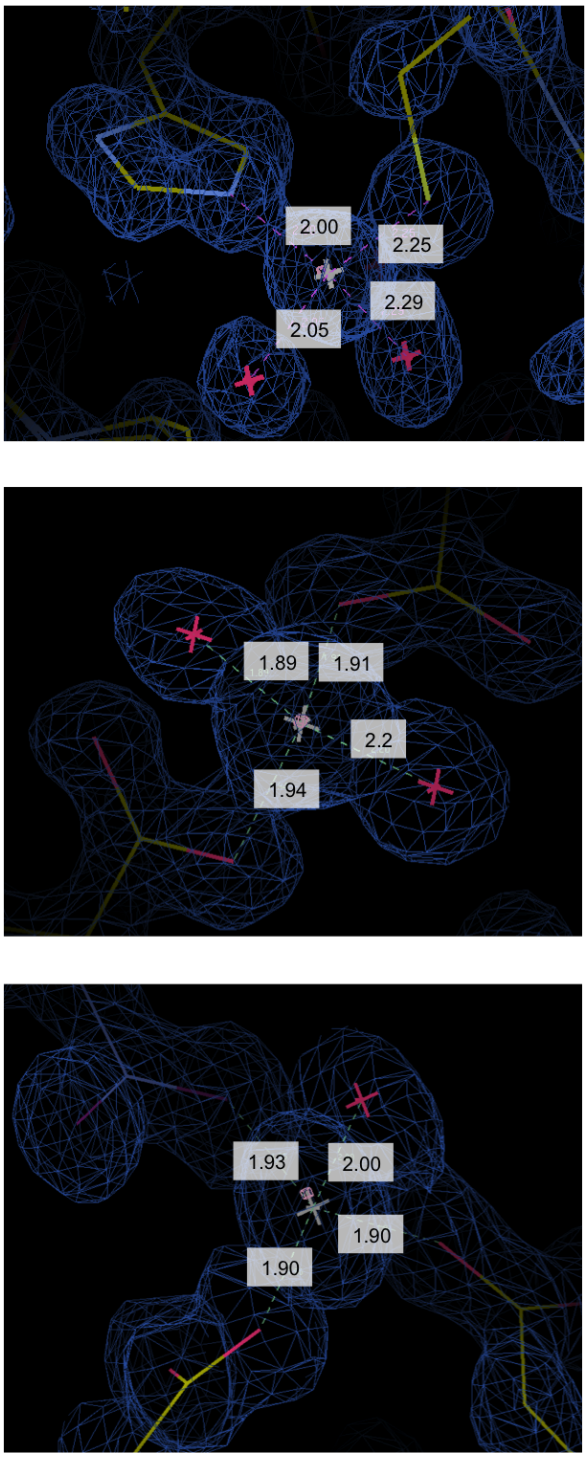


Fig. B.33 **Zinc ions binding sites and the corresponding ligand distances.** Zinc ions with ligands consisting of side chain atoms and water molecules forming tetrahedral geometry. The ligand distances (Å) are labelled in white boxes. An example of a zinc ion that forms crystal contact can also be seen (bottom). The electron density maps were contoured to 1.0 σ level.

B.7.1 Model building, validation and structural analysis of the untagged SSR1 WT monomer of crystal form 1

Cycles of model building and refinement on the structure of the crystal form 1 of SSR1 WT yielded the final R_{work} and R_{free} values of 15 and 16 %, respectively. The structure was then validated as described in Section 2.14.4. The final model of the crystal form 1 of SSR1 WT monomer consists of one molecule in the asymmetric unit with 11 zinc ions bound at the active site, on surface residues as well as bridging crystal contacts between symmetry-related molecules (Figure B.33). Details of the analysis of the structure and comparison with BLF1, especially regarding the clipping between residues 191 and 192 are discussed in Paper 3.

B.8 Model building, validation and structural analysis of a variety of SSR1 mutants

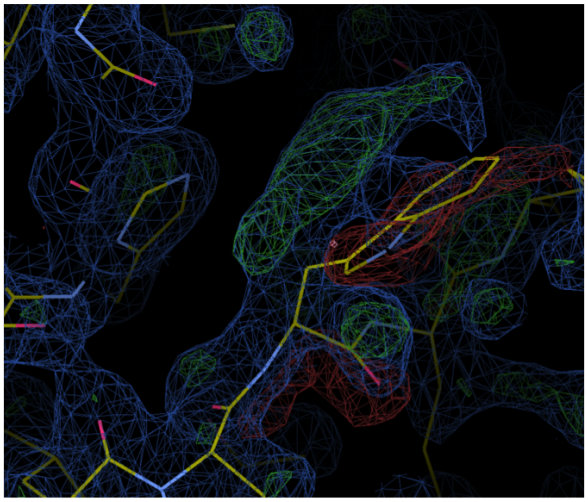
The final refined structure of the SSR1 WT crystals of form 1 was used as the search model for molecular replacement by PhaserMR (McCoy et al., 2007) for the structure determination of the rest of SSR1 variants. PhaserMR was run implementing the number of SSR1 chains to look for, based on the MATTPROB analysis described above. Where a single solution was found with reasonable scores (Top LLG and Top TFZ scores of >60 and >8, respectively, indicate a possible solution), the initial model and the electron density map were analysed for bias, symmetry clashes and crystal packing. When no obvious faults were found, the initial models were then subjected into iterative cycles of model building and refinement followed by validation as previously mentioned. Details of the final refined models and validation scores of all SSR1 structures are included in Paper 3, Table 1. Detailed analysis of these structures are discussed in Paper 3.

B.8.1 Model building, validation and structural analysis of SSR1 WT monomer of crystal form 2 and SSR1 WT dimer

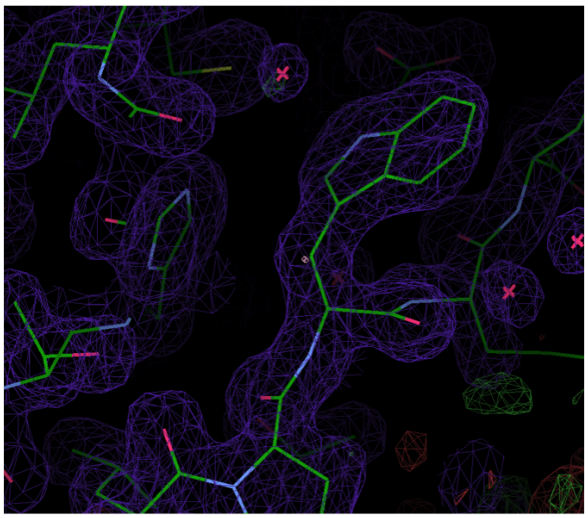
An examination of the electron density map of the structure of SSR1 WT of crystal form 2 following molecular replacement showed that the level of bias was low. For example, the position of Trp64 in the crystal form 1 of SSR1 WT structure is different than that in crystal form 2 but the resultant electron density map clearly identified the correct position Figure B.34. The crystal packing was also checked for bad clashes and these were found to be very low (Figure B.35). A similar approach was taken during model building and refinement of the structure of the SSR1 WT dimer, which revealed continuous electron density consistent of a domain-swapping event (Figure B.36).

B.8.2 Model building, validation and structural analysis of the dimeric and monomeric SSR1 C94S mutant

A comparison of the initial and the refined model of the SSR1 C94S monomer with their corresponding electron density maps is shown in Figure B.37. The refined electron density map showed clipping between Ser191 and Ser192 as seen in the SSR1 WT monomer structure. The structure of the dimeric SSR1 C94S mutant also revealed an extended conformation of the S191-S192 residues which are involved in domain-swapping with the 2-fold related subunit, albeit with a distinct pattern to that of the SSR1 WT dimer as discussed in Paper 3 (Figure B.38).



Initial model



Refined model

Fig. B.34 **A typical bias check, exemplified by the model building and refinement of the structure of crystal form 2 of SSR1 WT.** Top) A tryptophan (W64) side chain was in a different position in crystal form 2 compared to its position in the coordinates of crystal form 1 used for molecular replacement. The F_{obs}-F_{calc} map showed that the electron density is unbiased towards to model. Bottom) The refined model with its electron density map clearly reveals the correct side chain position. The electron density maps were contoured to 1.0 σ level.

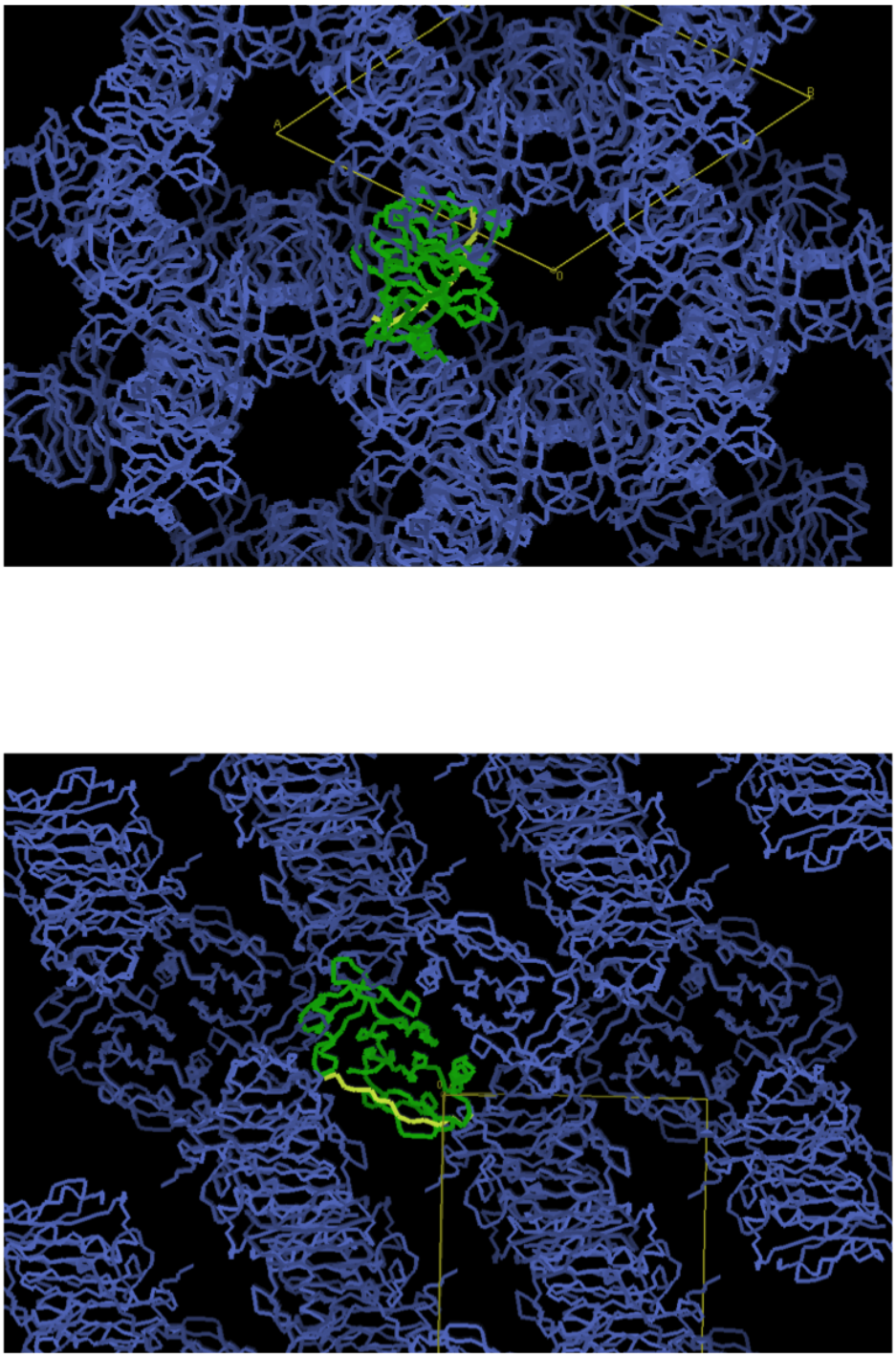


Fig. B.35 **A typical crystal packing and clashes analysis, exemplified by the model building and refinement of the structure of crystal form 2 of SSR1 WT.** The crystal packing within the crystal lattice down the **c** (top) and **a** axis (bottom), respectively. The packing of chains was analysed for obvious gaps in the lattice and checked for extra density unaccounted for. In this case, SSR1 is clipped producing a major chain (green) and a shorter chain (yellow).

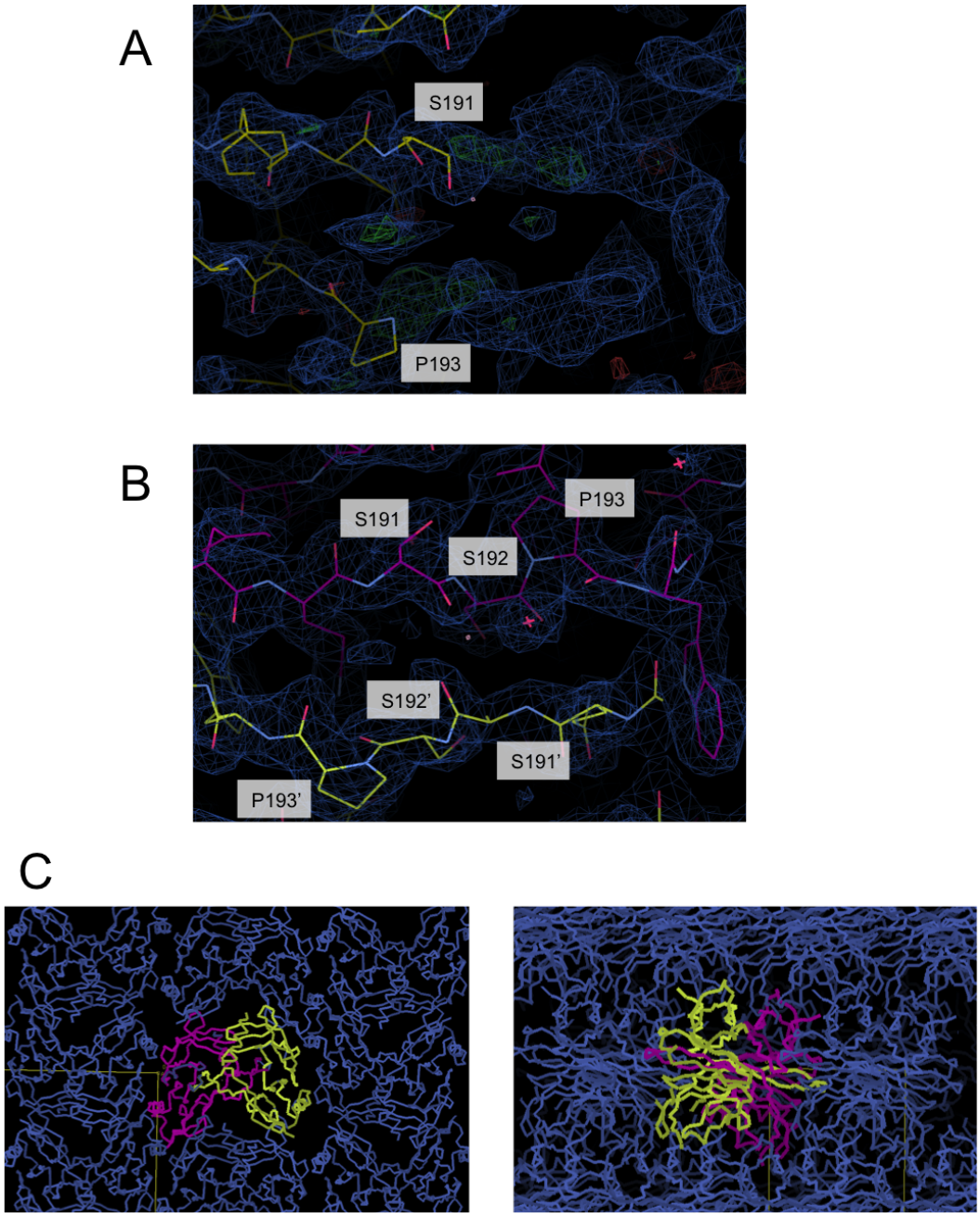


Fig. B.36 **Model building and refinement of the structure of SSR1 WT dimer.** A) The initial electron density map revealed that the electron density following S191 is not connected to P193 (S192 was initially omitted). B) The final refined model with the corresponding electron density map revealing clear position of an extended conformation of the S191-S192 motif which resulted in domain swapping (as discussed in paper 3). Residues marked with (') indicate those from a 2-fold symmetry related chain. The electron density maps were contoured to 1.0 σ level. C) The packing of SSR1 WT domain-swapped dimer along **a** (left) and **c** (right) axis, respectively.

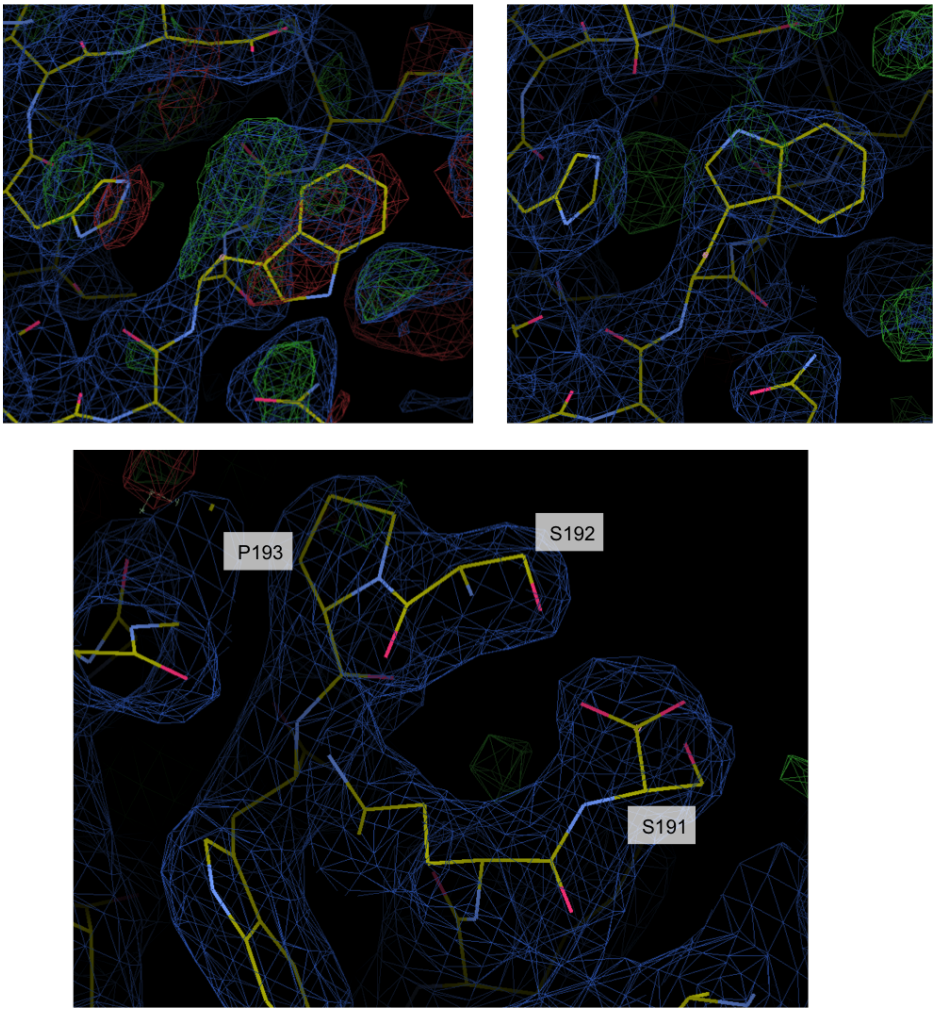


Fig. B.37 **Model building and refinement of the structure of SSR1 C94S monomer.** Top) Examination of the electron density map around W64 showing a low level of bias. Bottom) Consistent with the SDS-PAGE analysis of this mutant, electron density map confirms the clipping between residues S191 and S192. The electron density maps were contoured to 1.0 σ level.

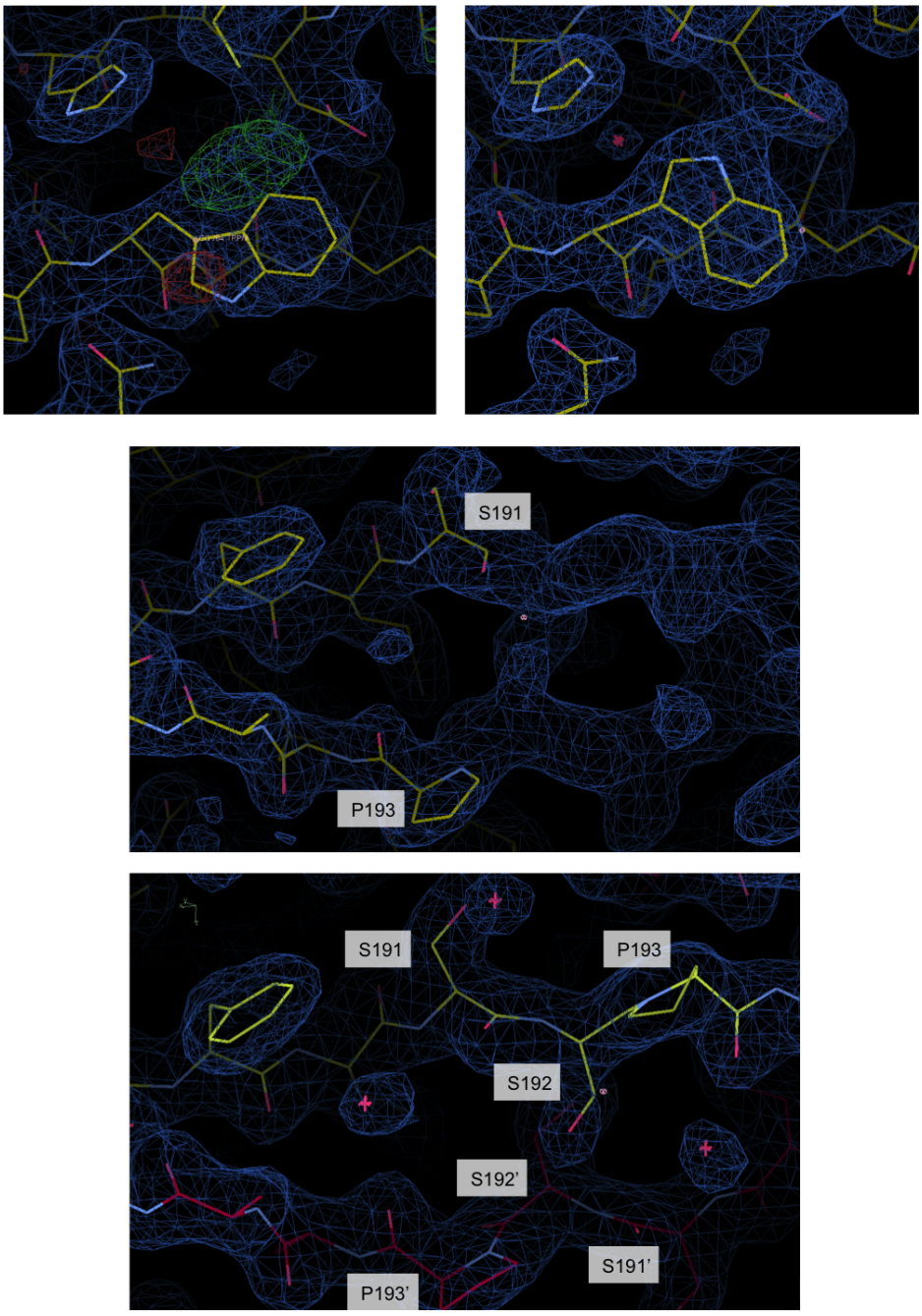


Fig. B.38 **Model building and refinement of the structure of SSR1 C94S dimer.** Examination of the electron density map around W64 showing a low level of bias (top). The initial electron density map revealing a continuous density following S191, not connected to P193 (S192 was initially omitted) (middle). The final refined map revealed the positions of the S191-S192 residues, lying in an extended conformation (bottom). Residues marked with (') indicate those from a 2-fold symmetry related chain. The electron density maps were contoured to 1.0 σ level.

B.8.3 Model building, validation and structural analysis of the variety of SSR1 clip site mutants

The initial and final refined models with their corresponding electron density maps of SSR1 S191A-S192G, S191A-S192A, S192A, and S192C mutants are shown in Figure B.39. Analysis of the structures identifying the absence of clipping and its implications in the folding of SSR1 are addressed in Paper 3. The structure of the SSR1 S192C mutant consists of 4 subunits in the asymmetric unit. The final model revealed that they comprised 2 disulphide-linked dimers with C192 of each subunit linked to that of a neighbouring subunit with clear indication of clipping between residues 191 and 192 (Figure B.40).

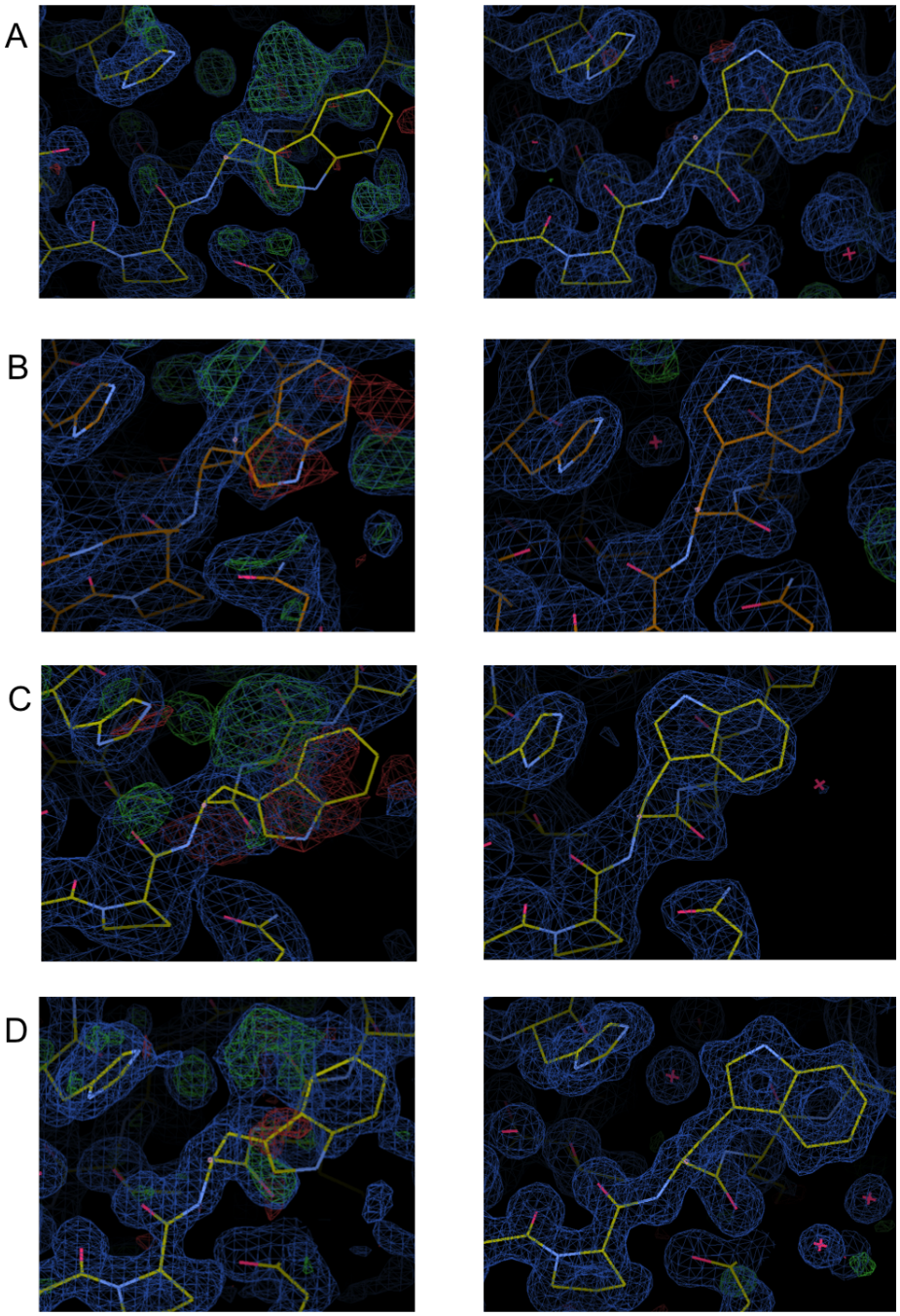


Fig. B.39 **Model building and refinement of the SSR1 clip site mutants.** Examination of the electron density map around W64 showing a low level of bias for A) S191A-S192G, B) S191A-S192A, C) S192A, and D) S192C mutants, respectively. Initial (left panels) and the refined models (right panels) with their corresponding electron density map are shown. The electron density maps were contoured to 1.0 σ level.

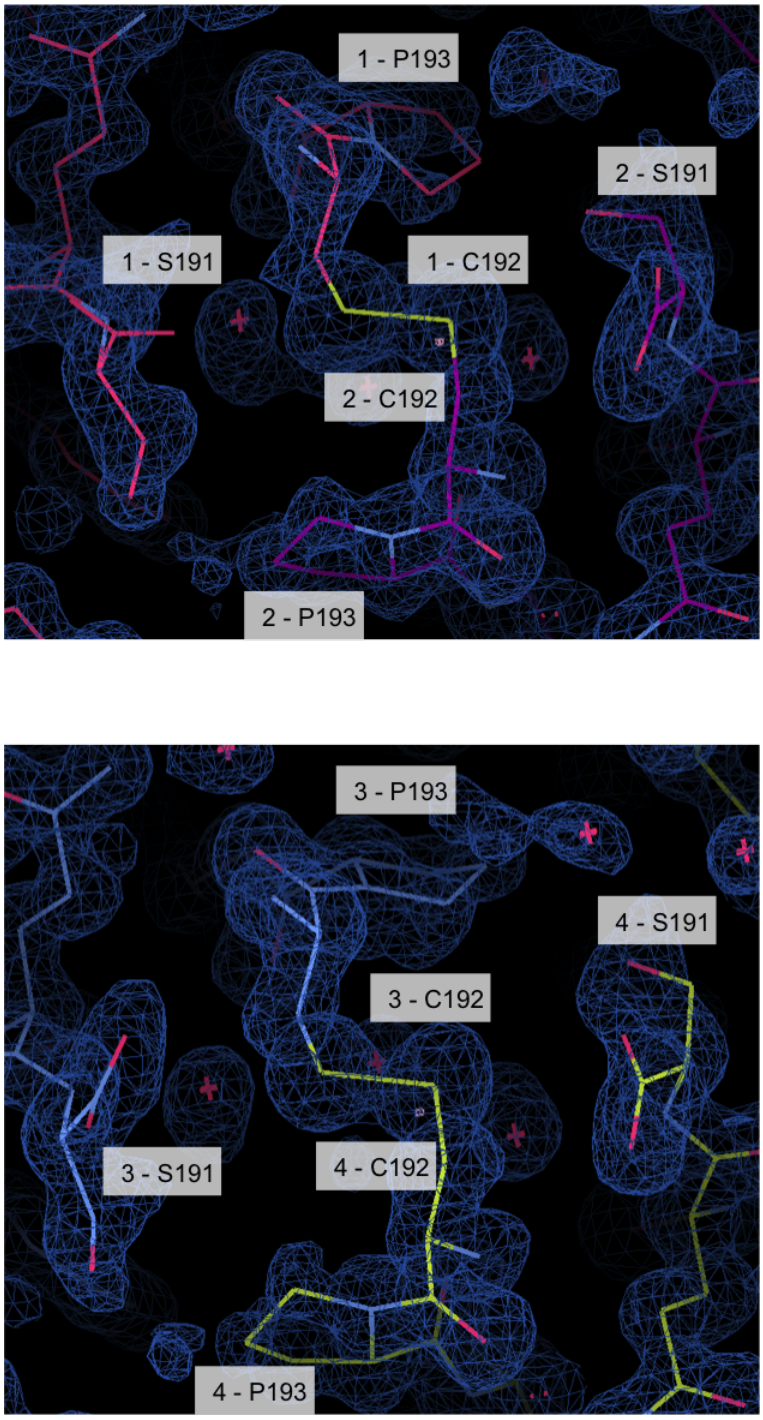


Fig. B.40 **Disulphide-linkage between the subunits of SSR1 S192C mutant.** SSR1 S192C crystallised with four subunits in the asymmetric unit consisting of two disulphide-linked dimers. There is clear evidence of complete clipping between residues 191 and 192, but C192 formed disulphide bridge to the equivalent residue in another subunit. Disulphide bond formed between subunits 1 and 2, and between subunits 3 and 4 are shown in the top and bottom panel, respectively. The electron density maps were contoured to 1.0 σ level.

Chapter 6

Conclusions and future work

6.1 Conclusions

This thesis aimed to expand the CNF1-like family of GDT by exploring and characterising other members of the family on their own or in complex with their molecular targets. In light of the discovery of BLF1, the second member of this family, and subsequent identification of its homologs as described in this thesis, it had become clear that more proteins belong to this family than initially thought. Unlike the other more established papain-like family of GDTs, with abundant apo and substrate-complex crystal structures allowing critical analysis on the aspects of toxin-substrate specificity and mechanism, such information in the CNF1-like family remains elusive. The conservation of a similar fold and the use of a Cys-His dyad in the papain-like family of GDTs suggest that they all share a similar mechanism to the cysteine protease, papain, which had emerged following divergent evolutionary events. The crystal structure of the C-CNF1 toxin had been solved revealing the involvement of a Cys-His pair in the active site in a similar relative orientation to that in papain. However, with the fold of C-CNF1 being completely different from papain, and the absence of crystal structures of C-CNF1 in complex with any protein target, many questions remain to be answered on their specificity, mechanism and function, which leaves an opportunity worth exploring.

6.1.1 Analysis of crystal structures of BLF1 C94S in complex with eIF4A

This study began with crystallisation trials which aimed to increase the resolution of crystals of a BLF1 C94S mutant in complex with eIF4A beyond the initial diffraction limit of 2.5 Å. The purification of the complex was straightforward due to a relatively strong binding between the inactivated C94S mutant of BLF1 to eIF4A making a 1:1 complex during gel filtration. In contrast, although BLF1 WT forms a complex with eIf4A during purification, crystallisation of the complex had been difficult presumably due to the complete deamidation leading to the release of the product and the dissociation of the complex. Whilst obtaining crystals of the BLF1 C94S:eIF4A complex that diffracted to higher resolution proved not to be successful, the structure determination of the complex in a distinct crystal form was possible. Unfortunately, whilst these crystals diffracted to 2.2 Å the structure could only be solved at 3.0 Å due to problems related to anisotropy.

Superposition of the crystal structures of forms A and B revealed that they have identical binding modes whereby BLF1 binds in the cleft between the two domains of eIF4A. Minor differences seen between the structures include the ordering of some residues in eIF4A and a slight rotation of one of its two domains with respect to the other probably due to distinct crystal lattice contacts. Both structures preserve a novel orientation of the two domains of eIF4A burying two relatively large patches, one from each domain, that interact, via extensive van der Waal's forces, hydrogen bonds and salt bridges, to the corresponding surfaces on BLF1. These intricate interactions form the molecular basis of the specificity of BLF1 towards eIF4A as discussed in Paper 1. Further analysis of the residues in eIF4A that interact with BLF1, revealed that about 75 % are conserved across many eukaryotes. In addition to the novel orientation of the two domains of eIF4A in the formation of this complex, the structure of eIF4A in the complex at 2.5 Å represents the highest resolution structure of the two domains of human eIF4A in the PDB.

Critical in the analysis of the structures of the two crystal forms is the position of Gln339 of eIF4A close to the Ser94 and His106, the two residues that would make up the Cys-His dyad in an active toxin. Due to the presence of an un-modelled solvent close to Gln339 of eIF4A in crystal form B, which results in a shift of Gln339 to a non-productive conformation, further analysis on the aspects of the mechanism of BLF1 toxin was carried out on the crystal form A structure.

Comparison of the coordinates of the BLF1 C94S in complex with eIF4A and those of BLF1 C94S and WT apo structures indicates that they are almost identical apart from the torsion angle of the Cys94 in the WT apo structure. This suggests that the interaction seen in the BLF1 C94S:eIF4A structure closely mimics that in the active complex. As papain has been generally accepted as utilising a thiolate-imidazolium ion pair for its function, the question arose as to whether a similar situation can be seen in BLF1 structures. This led to the structure determination of fresh crystals of BLF1 WT where the radiation damage was minimised in order to examine the interactions of the Cys94 side chain, which revealed that the Sγ of Cys94 forms a hydrogen bond to His106, a feature rarely seen in thiol proteases due to oxidation of the cysteine. Subsequent analysis on the geometry suggests that the hydrogen resides on His106, thus leading to the proposal that the Cys-His pair in BLF1 is in the thiolate-imidazolium ion form and is stabilised by hydrogen bonding. This is analogous to the situation in the serine proteases but where the shared proton lies at the Oγ of the serine presumably due to its higher pKa.

Further analysis by superposition of the Gln339 of eIF4A and the dyad, on those of the cysteine protease, TEV:substrate, the papain-like GDT, Cif:Nedd8 and the papain:inhibitor complex structures, identified that BLF1 contains an oxyanion hole to stabilise the tetrahedral intermediate during the course of glutamine deamidation. Investigation on the conformation of Ser94 of BLF1 C94S in the complex structure indicates that it adopts a Burgi-Dunitz angle, not suitable for the nucleophilic attack on the carbonyl group of the Gln 339 side chain of

eIF4A. However, as seen in the BLF1 WT apo structure, the change in the torsion angle of Cys94 might well position the Sγ of Cys94 in an ideal Burgi-Dunitz angle to carry out the nucleophilic attack. The conformation of the Ser94 in the complex is therefore thought to be in a Near Attack Conformation. Taken together, the oxyanion hole and the implementation of a thiolate-imidazolium ion pair for activity in BLF1 lead to a proposal that glutamine deamidation in this family indeed shares a similar mechanism to that of cysteine protease, papain. This observation is an interesting example of convergent evolution.

6.1.2 Structure determination and analysis on the crystal structures of SSR1

Following the discovery of a BLF1 homolog, SSR1, in the genome of *S. stiibiresistens*, a programme of structure determination and characterisation of this protein was initiated. The success of this programme would then be exploited to obtain more information on the BLF1 family of toxins by identifying SSR1's molecular target and co-crystallisation with the target protein with the aim of getting higher resolution crystal structures to further analyse the precise details of intermolecular interactions that form the basis of substrate specificity.

However, the study took an unexpected turn following the setbacks encountered during protein production and purification, as discussed in Paper 2. In short, over-expression of SSR1 in recombinant *E. coli* strains required growth at low temperature over a long incubation time, to produce soluble protein. However, it was then noticed that the purified protein was in a mixture of full-length SSR1 and a clipped variant, producing two fragments of 21 and 3 kDa, respectively. Subsequent mass spectrometry analysis suggests that the clip site lies in the peptide bond between Ser191 and Ser192. Purification of SSR1 WT proteins via gel filtration yielded a dimeric peak containing mostly the unclipped proteins and a monomeric peak consisting of the clipped variant. Subsequent structure determination of these two variants of SSR1 WT revealed a monomeric form with the peptide bond between Ser191 and Ser192

clipped but the two fragments remained associated, and an unclipped domain-swapped dimer. Although 3D domain-swapping is not uncommon in protein structures, especially those determined by X-ray crystallography, the mutually exclusive event of peptide clipping and domain swapping in SSR1 represented an observation rarely seen before.

Investigations of the factors influencing peptide clipping and the propensity of domain-swapping were carried out by the generation of several mutants, analysis on their clipping pattern and behaviour during purification and structure determination. The use of protease inhibitors and protease deficient *E. coli* strains ruled out the involvement of extrinsic proteases in the clipping. In light of the proposal that glutamine deamidation shares a similar mechanism to cysteine proteases, one possibility that could cause the observed clipping was the toxin activity itself. An SSR1 C94S mutant, whereby the nucleophile, Cys94, was mutated to serine, a mutation that abolishes deamidase activity in BLF1, still resulted in a similar pattern of clipping and dimerisation seen in the WT SSR1. Interestingly, the C94S dimer had a distinct pattern of domain swapping as discussed in Paper 3. The linear extended conformation of Ser191 and Ser192 in both distinct domain-swapped dimers combined with the clipping in the monomeric forms suggested the roles played by residues surrounding the clip site in cleavage.

Structure determination of a variety of clip site mutants has provided insights into the origins of the clipping and the domain-swapping events, as discussed in great detail in Paper 3. In summary, prior to clipping, partially folded or the so-called open monomers stall due to a strain during folding that was produced as a result of a clash between the C_β atom of residue at position 192 and the C_δ atom of Pro193. Owing to the high concentration of protein in recombinant expression system, these open monomers would encounter similar subunits to domain swap into dimers while maintaining a stable monomer-like fold. However, a significant proportion of the open monomers then fold into closed monomers which are then clipped to release the strain but, to do so requires an active nucleophile (Ser/Cys/Thr) at position

192. The absence of any functional group to assist clipping as seen in the S192A/V mutants would then leave the loop in a strained state, evident in the unusual Ramachandran angles and deviations from standard bond angles of the peptide bond. Although the requirement of Ser/Cys/Thr in allowing peptide clipping mimics the N-O acyl shift in the mechanism of self-splicing inteins, much remain unknown on the molecular basis underpinning the clipping event. Nevertheless, these mutually exclusive events of strain-induced clipping and domain swapping in SSR1 represent a novel post-translational modification in order to maintain a stable "functional" fold.

6.2 Future work

6.2.1 Optimisation of BLF1 C94S:eIF4A complex crystal structures

Although the current 2.5 Å structure revealed the first information on the molecular basis of the specificity of BLF1 towards eIF4A, higher resolution structures would permit a more precise analysis of the intricate details of the interaction which would then be utilised in a drug development programme to hopefully counter the effects of BLF1 toxin during the progress of melioidosis. Following the identification and characterisation of BLF1 homologs in *S. stiibiresistens* and *H. chejuensis*, each possessing interesting characteristics, these toxins could also be used to provide this information. This would then be followed by mutation and binding studies to investigate the roles played by the residues in the interfaces between these toxins and eIF4A in virulence. Although the habitat of *S. stiibiresistens* and *H. chejuensis* would exclude humans as their target hosts, the high sequence conservation of eIF4A across many eukaryotes provides a broad opportunity for this analysis. Whilst the structure of HCH_03101 in complex with eIF4A is yet to be determined, superposition of the coordinates of HCH_03101 on BLF1C94S:eIF4A complex places the β-protrusion around the C-terminal

domain of eIF4A, which might implicate its role in substrate recognition and binding and thus supporting the role of the toxin as targeting eIF4A or related proteins (Figure 6.1).

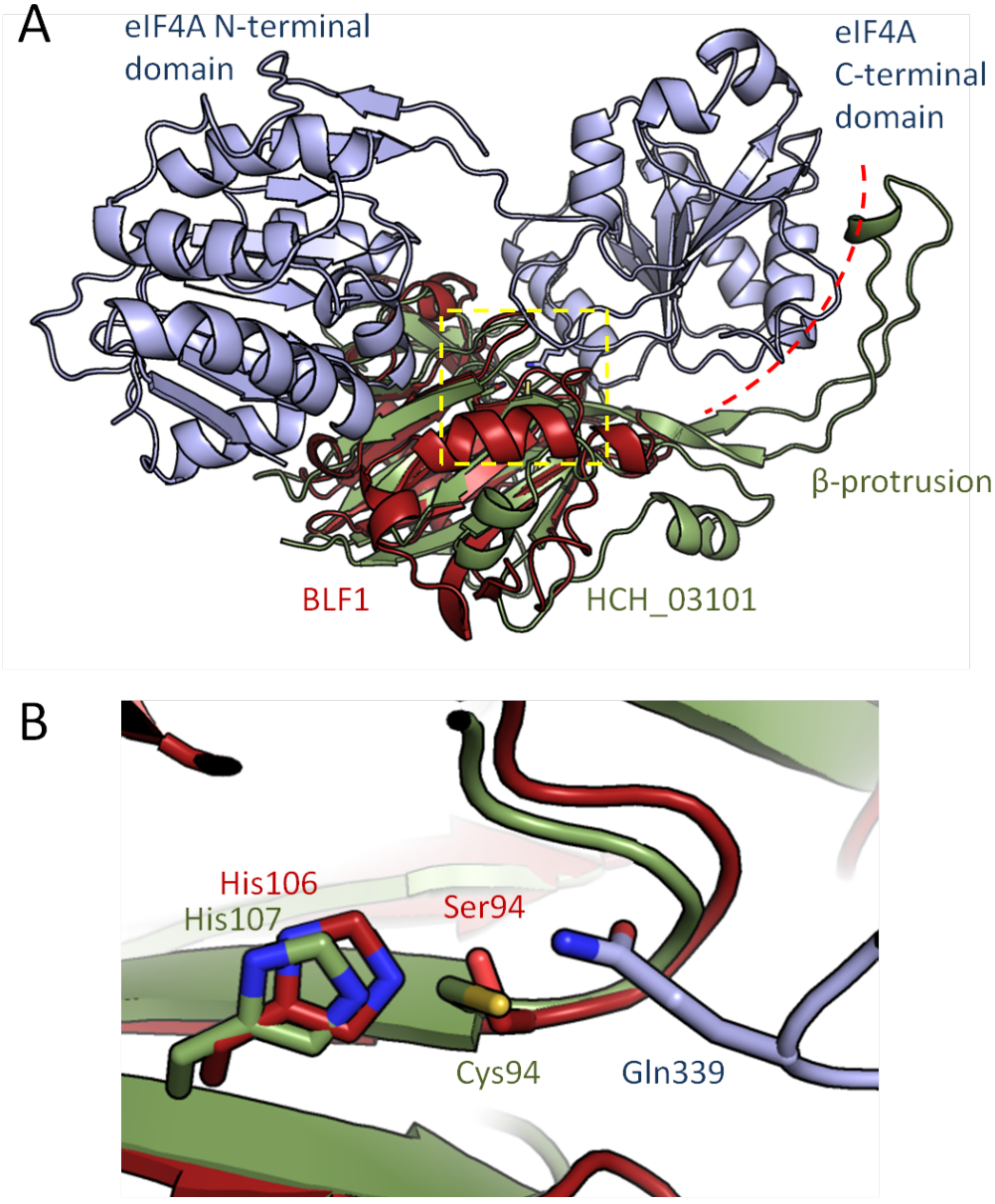


Fig. 6.1 **Superposition of HCH_03101 structure on BLF1C94S:eIF4A complex.** A) The β-protrusion of HCH_03101 seems to cradle around the C-terminal domain of eIF4A (red dashes) strongly suggesting the possibility of its involvement in recognition and binding. The target Gln339 of eIF4A is also placed closed to the active site residues in HCH_03101 (yellow dashed square) which is amplified in panel B.

6.2.2 Characterisation of *S. stiibiresistens* as an intracellular pathogen

As the pathogenicity of *S. stiibiresistens* is still poorly understood, the presence of BLF1 homolog in its genome, lacking the delivery warhead seen in CNF1, suggests that it is an intracellular pathogen. It is therefore crucial to identify the host organisms targeted by *S. stiibiresistens* to deepen our understanding of the involvement of this toxin in pathogenicity and disease.

6.2.3 The development of inhibitors for BLF1 or the CNF1-like family of GDTs

The structure of BLF1 C94S:eIF4A complex provided the first insights into the specificity of any toxin in the CNF1-like family which revealed a similarity to the mechanism of cysteine protease papain. Given the threats posed by *B. pseudomallei* and pathogenic *E. coli* strains, development of inhibitors becomes an urgent issue that needs to be addressed. Following the details of the specificity of BLF1 towards eIF4A discussed in Paper 1, and the addition of new candidates with the discovery of more BLF1 homologs, a more structured drug development programme can be developed, with the starting molecules being small peptides mimicking the substrate glutamine residue taking into consideration the successful design of papain inhibitors due to their similar mechanism.

6.2.4 BLF1-like toxins as anti-cancer or anti-carcinogenic agent

BLF1 deamidates Gln339 of eIF4A, an RNA helicase important in initiation-dependent protein translation in dividing cells. This provided an interesting scope of using these toxins as an agent to block or kill rapidly dividing i.e. cancer cells (Rust et al., 2018). The availability of several BLF1 homologs might provide opportunities to fine-tune the selectivity of these toxins towards cancerous cells without incurring any harm to normal cells.

6.2.5 Investigations of the relationship between the folding pathway of SSR1 and the evolutionary aspects of protein folding

The clipping and domain-swapping events in the folding pathway of SSR1 provide intriguing insights into the evolution of proteins and the protein folding events. However, many questions arise following these findings. How important is the role of strain in folding in the formation and therefore evolution of oligomeric forms of proteins? Is strain in folding a major force that governs misfolding and the formation of aggregates? Is strain-induced clipping a more widespread mechanism to ensure folding into a functional monomeric form as a safe passage to avoid misfolding, and hence loss of function? How much of the occurrence of domain-swapping observed following X-ray analysis is real and how much is an artefact of recombinant expression?

The observation of multiple domain-swapped/clipped structures in SSR1 clearly suggests that proteins are plastic and this could be beneficial in face of random mutations in order to maintain a stable functional fold. An activity assay as described above on the domain-swapped dimers would be an interesting avenue to support the view that strain and domain-swapping play significant roles in the evolution of protein folding. Unfolding of domain-swapped SSR1 proteins and refolding experiments in a much lower concentration would presumably prevent further domain-swapping, but due to the strain, it would be interesting to see if all the proteins would eventually be clipped in order to completely fold into a stable monomer.

Regardless, this strain-induced clipping event that has clear similarities with the mechanism of self-splicing inteins, provides an interesting relationship that raises many questions. The studies carried out on SSR1 in this thesis, therefore, provided a fundamental groundwork for future investigations. A structural sequence motif Ser/Cys/Thr-Pro at a turn on a loop would be an excellent starting point to search for more similar events in other proteins that might have been discarded at an early stage due to the clipping being treated as a problem.

Such studies would help strengthen our understanding of the clipping mechanism and protein folding in general.

References

Allwood, E. M., Devenish, R. J., Prescott, M., Adler, B., and Boyce, J. D. (2011). Strategies for Intracellular Survival of Burkholderia pseudomallei. *Frontiers in microbiology*, 2:170.

Aubert, D. F., Xu, H., Yang, J., Shi, X., Gao, W., Li, L., Bisaro, F., Chen, S., Valvano, M. A., and Shao, F. (2016). A Burkholderia Type VI Effector Deamidates Rho GTPases to Activate the Pyrin Inflammasome and Trigger Inflammation. *Cell Host & Microbe*, 19(5):664–674.

Bae, K. S., Lee, D. S., Lee, H. K., Lee, H. S., Chun, J., Moon, E. Y., and Ko, S. H. (2001). Hahella chejuensis gen. nov., sp. nov., an extracellular-polysaccharide-producing marine bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 51(2):661–666.

Blocker, A. J., Deane, J. E., Veenendaal, A. K. J., Roversi, P., Hodgkinson, J. L., Johnson, S., and Lea, S. M. (2008). What's the point of the type III secretion system needle? *Proceedings of the National Academy of Sciences of the United States of America*, 105(18):6507–13.

Boquet, P. (2001). The cytotoxic necrotizing factor 1 (CNF1) from Escherichia coli. *Toxicon*, 39(11):1673–1680.

Buetow, L., Flatau, G., Chiu, K., Boquet, P., and Ghosh, P. (2001). Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1. *Nature structural biology*, 8(7):584–588.

Büttner, D. (2012). Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiology and molecular biology reviews : MMBR*, 76(2):262–310.

Caprioli, A., Falbo, V., Roda, L. G., Ruggeri, F. M., and Zona, C. (1983). Partial purification and characterization of an escherichia coli toxic factor that induces morphological cell alterations. *Infect. Immun.*, 39(3):1300–1306.

Chao, X., Muff, T. J., Park, S. Y., Zhang, S., Pollard, A. M., Ordal, G. W., Bilwes, A. M., and Crane, B. R. (2006). A receptor-modifying deamidase in complex with a signaling phosphatase reveals reciprocal regulation. *Cell*, 124(3):561–571.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1):12–21.

Clayton, A. J., Lisella, R. S., and Martin, D. G. (1973). Melioidosis: a serological survey in military personnel. *Military medicine*, 138(1):24–6.

Cornelis, G. R. (2006). The type III secretion injectisome. *Nature Reviews Microbiology*, 4(11):811–825.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic acids research*, 16(22):10881–90.

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13(6):343–359.

Crow, A., Hughes, R. K., Taieb, F., Oswald, E., and Banfield, M. J. (2012). The molecular basis of ubiquitin-like protein NEDD8 deamidation by the bacterial effector protein Cif. *Proceedings of the National Academy of Sciences of the United States of America*, 109(27):E1830–8.

Cruz, J. W., Rothenbacher, F. P., Maehigashi, T., Lane, W. S., Dunham, C. M., and Woychik, N. A. (2014). Doc toxin is a kinase that inactivates elongation factor Tu. *The Journal of biological chemistry*, 289(11):7788–98.

Cruz-Migoni, a., Hautbergue, G. M., Artymiuk, P. J., Baker, P. J., Bokori-Brown, M., Chang, C.-T., Dickman, M. J., Essex-Lopresti, A., Harding, S. V., Mahadi, N. M., Marshall, L. E., Mobbs, G. W., Mohamed, R., Nathan, S., Ngugi, S. a., Ong, C., Ooi, W. F., Partridge, L. J., Phillips, H. L., Raih, M. F., Ruzheinikov, S., Sarkar-Tyson, M., Sedelnikova, S. E., Smither, S. J., Tan, P., Titball, R. W., Wilson, S. a., and Rice, D. W. (2011). A Burkholderia pseudomallei Toxin Inhibits Helicase Activity of Translation Factor eIF4A. *Science*, 334(6057):821–824.

Currie, B. (2003). Melioidosis: an important cause of pneumonia in residents of and travellers returned from endemic regions. *European Respiratory Journal*, 22(3):542–550.

Currie, B. J. (2015). Melioidosis : Evolving Concepts in Epidemiology , Pathogenesis , and Treatment. *Seminars in Respiratory and Critical Care Medicine*, 36(01):111–125.

Deane, J. E., Abrusci, P., Johnson, S., and Lea, S. M. (2010). Timing is everything: the regulation of type III secretion. *Cellular and molecular life sciences : CMLS*, 67(7):1065–75.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4):486–501.

Falke, J. J. and Hazelbauer, G. L. (2001). Transmembrane signaling in bacterial chemoreceptors. *Trends in Biochemical Sciences*, 26(4):257–265.

Falzano, L., Fiorentini, C., Donelli, G., Michel, E., Kocks, C., Cossart, P., Cabanié, L., Oswald, E., and Boquet, P. (1993). Induction of phagocytic behaviour in human epithelial cells by Escherichia coli cytotoxic necrotizing factor type1. *Molecular Microbiology*, 9(6):1247–1254.

Fisher, S. J., Levik, K. E., Williams, M. A., Ashton, A. W., and McAuley, K. E. (2015). SynchWeb: a modern interface for ISPyB. *Journal of applied crystallography*, 48(Pt 3):927–932.

Flatau, G., Landraud, L., Boquet, P., Bruzzone, M., and Munro, P. (2000). Deamidation of RhoA glutamine 63 by the Escherichia coli CNF1 toxin requires a short sequence of the GTPase switch 2 domain. *Biochemical and biophysical research communications*, 267(2):588–92.

Flatau, G., Lemichez, E., Gauthier, M., Chardin, P., Paris, S., Fiorentini, C., and Boquet, P. (1997). Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine. *Nature*, 387(6634):729–733.

Fraser, J. D. (2011). Clarifying the mechanism of superantigen toxicity. *PLoS biology*, 9(9):e1001145.

Fürstner, A. (2003). Chemistry and Biology of Roseophilin and the Prodigiosin Alkaloids: A Survey of the Last 2500 Years. *Angewandte Chemie International Edition*, 42(31):3582–3603.

Galán, J. E. and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, 444(7119):567–573.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31(13):3784–8.

Green, E. R. and Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology spectrum*, 4(1).

Hanahan, D. (1983). Studies on transformation of escherichia coli with plasmids. *Journal of Molecular Biology*, 166(4):557 – 580.

Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009). *EDNA* : a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of Synchrotron Radiation*, 16(6):872–879.

Istivan, T. S. and Coloe, P. J. (2006). Phospholipase A in Gram-negative bacteria and its role in pathogenesis. *Microbiology*, 152(5):1263–1274.

Jeong, H., Yim, J. H., Lee, C., Choi, S.-H., Park, Y. K., Yoon, S. H., Hur, C.-G., Kang, H.-Y., Kim, D., Lee, H. H., Park, K. H., Park, S.-H., Park, H.-S., Lee, H. K., Oh, T. K., and Kim, J. F. (2005). Genomic blueprint of Hahella chejuensis, a marine microbe producing an algicidal agent. *Nucleic acids research*, 33(22):7066–73.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923.

Korotkov, K. V., Sandkvist, M., and Hol, W. G. J. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, 10(5):336–351.

Krissinel, E. and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, 60(12):2256–2268.

Krissinel, E. and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, 372(3):774–797.

Kristich, C. J. and Ordal, G. W. (2002). Bacillus subtilis CheD is a chemoreceptor modification enzyme required for chemotaxis. *The Journal of biological chemistry*, 277(28):25356–62.

Kubori, T., Sukhan, A., Aizawa, S. I., and Galán, J. E. (2000). Molecular characterization and assembly of the needle complex of the Salmonella typhimurium type III protein secretion system. *Proceedings of the National Academy of Sciences of the United States of America*, 97(18):10225–30.

Kulp, A. and Kuehn, M. J. (2010). Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annual review of microbiology*, 64:163–84.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259):680–685.

Lara-Tejero, M., Kato, J., Wagner, S., Liu, X., and Galán, J. E. (2011). A sorting platform determines the order of protein secretion in bacterial type III systems. *Science (New York, N.Y.)*, 331(6021):1188–91.

Lebeda, F. J., Cer, R. Z., Mudunuri, U., Stephens, R., Singh, B. R., and Adler, M. (2010). The zinc-dependent protease activity of the botulinum neurotoxins. *Toxins*, 2(5):978–97.

Lerm, M., Selzer, J., Hoffmeyer, A., Rapp, U. R., Aktories, K., and Schmidt, G. (1999). Deamidation of Cdc42 and Rac by Escherichia coli Cytotoxic Necrotizing Factor 1: Activation of c-Jun N-Terminal Kinase in HeLa Cells. *Infect. Immun.*, 67(2):496–503.

Leslie, A. G. W. and Powell, H. R. (2007). Processing diffraction data with mosflm. *Evolving Methods for Macromolecular Crystallography*, pages 41–51.

Lockman, H. A., Gillespie, R. A., Baker, B. D., and Shakhnovich, E. (2002). Yersinia pseudotuberculosis produces a cytotoxic necrotizing factor. *Infection and immunity*, 70(5):2708–14.

Lowe, G. (1976). The Cysteine proteinases. *Tetrahedron*, 32(3):291–302.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography*, 40(Pt 4):658–674.

Miller, J. H. (1972). Experiments in molecular genetics. *Cold Spring Harbor Laboratory*.

Mobbs, G. (2014). Structural Studies on Glutamine de-amidase enzymes. *PhD Thesis, The University of Sheffield Library*.

Montecucco, C. and Molgó, J. (2005). Botulinal neurotoxins: revival of an old killer. *Current Opinion in Pharmacology*, 5(3):274–279.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011). *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*, 67(4):355–367.

Ngauy, V., Lemeshev, Y., Sadkowski, L., and Crawford, G. (2005). Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II. *Journal of clinical microbiology*, 43(2):970–2.

Nivaskumar, M., Bouvier, G., Campos, M., Nadeau, N., Yu, X., Egelman, E. H., Nilges, M., and Francetic, O. (2014). Distinct docking and stabilization steps of the Pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers. *Structure (London, England : 1993)*, 22(5):685–96.

Nivaskumar, M. and Francetic, O. (2014). Type II secretion system: A magic beanstalk or a protein escalator. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8):1568–1577.

Orden, J. A., Domínguez-Bernal, G., Martínez-Pulgarín, S., Blanco, M., Blanco, J. E., Mora, A., Blanco, J., Blanco, J., and de la Fuente, R. (2007). Necrotoxigenic Escherichia coli from sheep and goats produce a new type of cytotoxic necrotizing factor (CNF3) associated with the eae and ehxA genes. *International microbiology : the official journal of the Spanish Society for Microbiology*, 10(1):47–55.

Orth, J. H. C., Aktories, K., and Kubatzky, K. F. (2007). Modulation of host cell gene expression through activation of STAT transcription factors by Pasteurella multocida toxin. *The Journal of biological chemistry*, 282(5):3050–7.

Orth, J. H. C., Preuss, I., Fester, I., Schlosser, A., Wilson, B. A., and Aktories, K. (2009). Pasteurella multocida toxin activation of heterotrimeric G proteins by deamidation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):7179–84.

Oswald, E., Rycke, J., Guillot, J. F., and Boivin, R. (1989). Cytotoxic effect of multinucleation in HeLa cell cultures associated with the presence of Vir plasmid in Escherichia coli strains. *FEMS Microbiology Letters*, 58(1):95–99.

Overbye, L. J., Sandkvist, M., and Bagdasarian, M. (1993). Genes required for extracellular secretion of enterotoxin are clustered in Vibrio cholerae. *Gene*, 132(1):101–106.

Peraro, M. D. and van der Goot, F. G. (2016). Pore-forming toxins: ancient, but never really out of fashion. *Nature Reviews Microbiology*, 14(2):77–92.

Pérez-Tomás, R., Montaner, B., Llagostera, E., and Soto-Cerrato, V. (2003). The prodigiosins, proapoptotic drugs with anticancer properties. *Biochemical Pharmacology*, 66(8):1447–1452.

Pradel, N., Ye, C., Livrelli, V., Xu, J., Joly, B., and Wu, L.-F. (2003). Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157:H7. *Infection and immunity*, 71(9):4908–16.

Read, R. J. and Schierbeek, A. J. (1988). A phased translation function. *Journal of Applied Crystallography*, 21(5):490–495.

Rietschel, E. T., Kirikae, T., Schade, F. U., Mamat, U., Schmidt, G., Loppnow, H., Ulmer, A. J., Zähringer, U., Seydel, U., and Di Padova, F. (1994). Bacterial endotoxin: molecular relationships of structure to activity and function. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 8(2):217–25.

Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, 42(Web Server issue):W320–4.

Rust, A., Shah, S., Hautbergue, G. M., and Davletov, B. (2018). Burkholderia Lethal Factor 1, a Novel Anti-Cancer Toxin, Demonstrates Selective Cytotoxicity in MYCN-Amplified Neuroblastoma Cells. *Toxins*, 10(7).

Sanada, T., Kim, M., Mimuro, H., Suzuki, M., Ogawa, M., Oyama, A., Ashida, H., Kobayashi, T., Koyama, T., Nagai, S., Shibata, Y., Gohda, J., Inoue, J.-i., Mizushima, T., and Sasakawa, C. (2012). The Shigella flexneri effector OspI deamidates UBC13 to dampen the inflammatory response. *Nature*, 483(7391):623–6.

Sandkvist, M., Morales, V., and Bagdasarian, M. (1993). A protein required for secretion of cholera toxin through the outer membrane of Vibrio cholerae. *Gene*, 123(1):81–86.

Schmidt, G., Sehr, P., Wilm, M., Selzer, J., Mann, M., and Aktories, K. (1997). Gln63 of Rho is deamidated by Escherichia coli cytotoxic necrotizing factor-1. *Nature*, 387(6634):725–729.

Schraidt, O. and Marlovits, T. C. (2011). Three-dimensional model of Salmonella's needle complex at subnanometer resolution. *Science (New York, N.Y.)*, 331(6021):1192–5.

Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallographica Section A Foundations of Crystallography*, 64(1):112–122.

Singh, M. and Mahmood, M. (2017). Melioidosis: the great mimicker. *Journal of Community Hospital Internal Medicine Perspectives*, 7(4):245–247.

Washington, E. J., Banfield, M. J., and Dangl, J. L. (2013). What a difference a Dalton makes: bacterial virulence factors modulate eukaryotic host cell signaling systems via deamidation. *Microbiology and molecular biology reviews : MMBR*, 77(3):527–39.

Weichenberger, C. X. and Rupp, B. (2014). Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallographica Section D Biological Crystallography*, 70(6):1579–1588.

White, N. J. (2003). Melioidosis. *Lancet*, 361(9370):1715–1722.

Wiersinga, W. J., Currie, B. J., and Peacock, S. J. (2012). Melioidosis. *New England Journal of Medicine*, 367(11):1035–1044.

Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P., and Peacock, S. J. (2006). Melioidosis: insights into the pathogenicity of Burkholderia pseudomallei. *Nature reviews. Microbiology*, 4(4):272–282.

Williams, J. M. and Tsai, B. (2016). Intracellular trafficking of bacterial toxins. *Current opinion in cell biology*, 41:51–6.

Winter, G. (2010). *xia2* : an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43(1):186–190.

Winter, G. and McAuley, K. E. (2011). Automated data collection for macromolecular crystallography. *Methods*, 55(1):81–93.

Wongtrakoongate, P., Mongkoldhumrongkul, N., Chaijan, S., Kamchonwongpaisan, S., and Tungpradabkul, S. (2007). Comparative proteomic profiles and the potential markers between Burkholderia pseudomallei and Burkholderia thailandensis. *Molecular and cellular probes*, 21(2):81–91.

Yao, Q., Cui, J., Wang, J., Li, T., Wan, X., Luo, T., Gong, Y.-N., Xu, Y., Huang, N., and Shao, F. (2012). Structural mechanism of ubiquitin and NEDD8 deamidation catalyzed by bacterial effectors that induce macrophage-specific apoptosis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50):20395–400.

Zhao, J., Zeng, Y., Xu, S., Chen, J., Shen, G., Yu, C., Knipe, D., Yuan, W., Peng, J., Xu, W., Zhang, C., Xia, Z., and Feng, P. (2016). A Viral Deamidase Targets the Helicase Domain of RIG-I to Block RNA-Induced Activation. *Cell Host & Microbe*, 20(6):770–784.