

*Investigation into the structure and function of
Hsp47*

Umbreen Ahmed

A thesis submitted to the University of Birmingham
for the degree of Doctor of Philosophy

School of Biosciences
University of Birmingham
Birmingham
B15 2TT

May 2009

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Acknowledgements

Firstly, a huge thanks to my supervisor, Dr Timothy Dafforn for his continuous support, enthusiasm and patience throughout my PhD. It has been a pleasure working for him, even when the lab work has not been particularly rewarding! Our many conversations on cars and formula 1 have been especially enjoyable and will be missed.

I would also like to thank our collaborators from Imperial College (Andrew Miller, Dee Olerenshaw, Takayuki Homma, Micheal Wright and Firdaus Abdul) who provided me with the *Mus musculus* Hsp47 clones. Their mutagenesis work has made this thesis possible and our fruitful discussions have been both interesting and helpful. Thanks also to the James Whisstock group at Monash University, Australia, for providing me with the opportunity to work with such enthusiastic and dedicated scientists.

I am indebted to others for their valuable time, expertise and use of instrumentation with regard to specific techniques; Dr Corrine Smith for electron microscopy and Dr Klaus Fütterer for protein crystallography. There are also those who have assisted me in the laboratory – Rosemary Parslow for help with a bit of everything; Anshuman Shukla and Philip Robinson for work on *Xenopus laevis* and *Danio rerio* Hsp47; and two project students, Nguyen Le and Emma Pritchard for assistance in the purification of *Xenopus laevis* and *Mus musculus* Hsp47, respectively. Many thanks also go to past and present members of 7th floor laboratories; Matt, Denise, Karthik, David, Milly, Michelle, Yukie, Yu Pin, Rich and Raul for providing a friendly working atmosphere. A special thank you to Rachel Kendrick for listening, understanding and helping. Our fabulous trips away have made the last three years unforgettable and hopefully there are many more to come.

I also wish to thank Jyoti Patel, a great friend, whose support, humour and advice has made this experience particularly pleasant. Our countless conversations, scientific or otherwise, have provided me with the motivation to fulfil my true potential. Her friendship is truly appreciated.

I am forever indebted to my parents for their unconditional love, understanding and patience throughout. Without their continuous support this thesis would never have happened. Also I am extremely grateful to my brothers, Imran and Farhan, for their constant encouragement, belief and humour.

The financial support from Medical Research Council is greatly appreciated.

Abstract

Collagen is the most abundant protein in mammals and is the main protein of connective tissue in animals [1]. It is a fibrous structural protein that is the major component of the extracellular matrix, which provides great tensile strength to ligaments, tendons, bone and skin, for example. Hsp47 is a procollagen and collagen specific chaperone that is thought to be involved in the correct assembly and transport of collagen chains. It is a non-inhibitory member of the serine protease inhibitor superfamily and uses a pH-dependent mechanism for binding and release to/from its substrate.

In order to study the function of Hsp47 sufficient quantities of the protein need to be expressed and purified. Previous work on Hsp47 studied mammalian forms of the protein which expressed poorly in bacterial expression systems. This research studied Hsp47 in the poikilothermic vertebrate; *Xenopus laevis*, the ectothermic vertebrate; *Danio rerio* and the endothermic vertebrate; *Mus musculus* in a range of expression systems in order to overcome the low protein expression levels. *Mus musculus* Hsp47 proved the most successful as it yielded soluble protein that could be used for structural and functional characterisation.

A reduction in the pH of the environment triggers a conformational change that is vital for the function of Hsp47. Understanding this structural change is important in understanding the function of Hsp47. Histidine residues are thought to play a vital role in the conformational transition of Hsp47. Histidine variants were made to identify possible binding sites and/or areas involved in the instigation of the structural change in Hsp47. The mutations allowed the first identification of a more stable, latent, serpin state of the protein. Interestingly, this latent form comprised of enhanced biological activity as it reduced collagen fibril formation. One of the mutants lacked any collagen binding ability, which may be an indication of a possible collagen binding site on Hsp47.

The results presented in this thesis provide an insight into the function and structure of the non-inhibitory serpin Hsp47 and may enhance the understanding of the molecular mechanism by which the collagen-specific chaperone operates.

Abbreviations

Abbreviations used in this thesis were as recommended by the Journal of Biological Chemistry.

α_1 -AT	α_1 -anti trypsin
ACT	Antichymotrypsin
AMP	Ampicillin
ANGT	Angiotensinogen
ANS	Anilino-8-naphthalene sulphonate
AUC	Analytical ultracentrifugation
BSA	Bovine serum albumin
CBD	Chitin binding domain
CBG	Cortisol binding globulin
CD	Circular dichroism
DTT	Dithiothreitol
EK	Enterokinase
ER	Endoplasmic reticulum
FRET	Fluorescent resonance energy transfer
hPCI	human protein C inhibitor
HSE	Heat shock element
HSF1	Heat shock factor 1
Hsp	Heat shock protein
I	Intermediate state
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kDa	Kilodalton
LB	Liquid broth

N	Native state
Ni-NTA	Nickel-nitrilotriacetic acid
P4-H	Prolyl-4-hydroxylase
PAI-1	Plasminogen activator inhibitor-1
PDI	Protein disulphide isomerase
PEG	Polyethylene glycol
PI	Proteinase inhibitor
RCL	Reactive centre loop
RPM	Rotations per minute
SERPIN	Serine protease inhibitor
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TBG	Thyroxine binding-globulin
TNF	Tumour necrosis factor
U	Unfolded state
UTMP	Uterine milk protein
WT	Wild type
6xHis	Hexa-Histidine

Table of Contents

CHAPTER 1: Introduction	1
1.1 Protein Folding	1
1.2 Protein misfolding and aggregation.....	5
1.3 Macromolecular crowding.....	7
1.4 Molecular Chaperones in Protein Folding.....	8
1.5 Hsp47.....	15
1.6 Collagen biosynthesis.....	17
1.7 Hsp47 and collagen	21
1.8 The pH dependant conformational switch mechanism.....	29
1.9 Hsp47 as a member of the Serine Protease Inhibitor Superfamily.....	31
1.9.1 Introduction	31
1.9.2 Identification of the Serpin Superfamily	31
1.9.3 Serpin Folding	34
1.9.4 Serpin Structure	37
1.9.4.1 The Breach Region	39
1.9.4.2 The Shutter Region.....	40
1.9.4.3 The Hinge Domains.....	41
1.9.5 Serpin Function	44
1.9.6 Serpin conformations.....	47
1.9.6.1 The Native Conformation.....	49
1.9.6.2 The Cleaved Conformation	49
1.9.6.3 The Latent Conformation	50
1.9.6.4 The Polymers.....	51
1.10 The Location of Key Histidines in Hsp47	57
1.11 Objectives	60
1.12 Aims	61
CHAPTER 2: Materials and Methods.....	63
2.1 Materials	63
2.2 Methods	65
2.2.1 PCR.....	74
2.2.2 Insertion of Genes into Vectors	76
2.2.3 Ligation.....	80
2.2.4 Transformation of One Shot [®] Top10 Chemically Component <i>E. coli</i> Cells.....	84
2.2.5 Transformation of BL21 star (DE3) One Shot Cells.....	86
2.2.6 Determination of Protein Expression Levels.....	87
2.2.7 Purification of the Protein expressed in TOPO [®] and Xa/LIC vectors.....	87
2.2.8 Purifying Hsp47 from Insoluble Intracellular Aggregates	90
2.2.9 Cleavage of the Thioredoxin Tag.....	92
2.2.10 Purification of protein expressed in pTYB12.....	94
2.2.11 Purification of mHsp47 in pCR [®] T7/NT-TOPO [®]	96
2.2.12 Detection of Hsp47	97
2.2.13 Characterisation of Purified Hsp47	100
2.2.13.1 Circular Dichroism	100
2.2.13.2 Gel Filtration Chromatography	101
2.2.13.3 Analytical Ultracentrifugation.....	102
2.2.13.4 pH Titrations.....	104
2.2.13.5 Collagen Fibrillisation Assay	104
2.2.13.6 Collagen Binding Assay	105
2.2.13.7 Electron Microscopy	106

CHAPTER 3: Purification and Characterisation of <i>Xenopus laevis</i> and <i>Danio rerio</i> Hsp47	108
3.1 Introduction	108
3.2 Results	110
3.2.1 Cloning into the vectors.....	110
3.2.2 Expression of the protein from the Vectors.....	114
3.2.3 Protein Purification.....	123
3.2.4 Fusion tag cleavage	131
3.2.5 Collagen Binding Assays	138
3.2.6 Characterisation of the Protein	141
3.3 Discussion.....	149
CHAPTER 4: Purification and Characterisation of WILD TYPE <i>Mus musculus</i> Hsp47	156
4.1 Introduction	156
4.2 Results	160
4.3 Discussion.....	173
CHAPTER 5: Probing the Importance of the Breach Region in the Structural Transition of Hsp47 by mutating H191, H197 and H198	180
5.1 Introduction	180
5.2 Results	184
5.3 Discussion.....	202
CHAPTER 6: Probing the Importance of the Breach Region in the Conformational Transition of Hsp47 by mutating H220	209
6.1 Introduction	209
6.2 Results	210
6.3 Discussion.....	220
CHAPTER 7: Probing the Importance of the Gate Region in the Structural Transition of Hsp47 by mutating H255 and H256	226
7.1 Introduction	226
7.2 Results	229
7.3 Discussion.....	237
CHAPTER 8: Crystallography.....	242
8.1 Introduction	242
8.2 Results	244
CHAPTER 9: General Discussion, Conclusions and Future work.	248
9.1 General Discussion and Conclusions	248
9.2 Future Work.....	263
CHAPTER 10: References	267
CHAPTER 11: Appendices	280

CHAPTER 1: Introduction

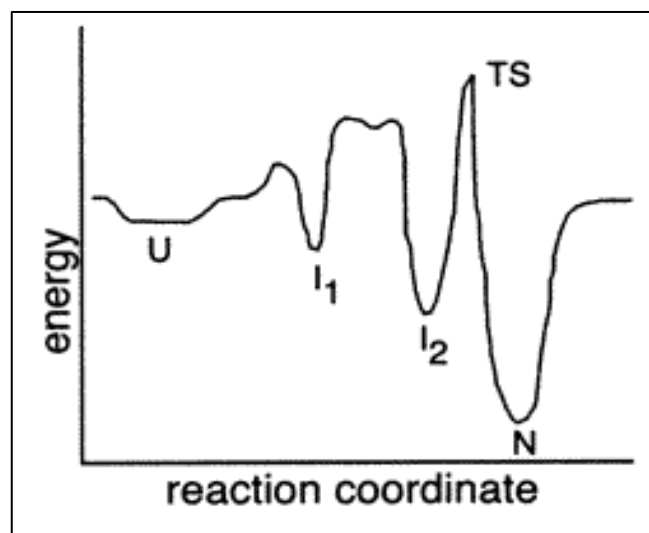
1.1 Protein Folding

Proteins are organic compounds that consist of amino acids which are joined by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues [2]. Their amino acid sequence determines their folding and ability to function and turnover [3], [4]. Proteins are multifunctional macromolecules that control chemical processes in the cell. To become biologically active they must fold into a unique stable structure [5], which involves polypeptide self-organisation [3]. To reach a balance between stability and flexibility, the free energy of stabilisation of globular proteins in solution usually involves only a few weak intermolecular interactions.

Protein folding is the process by which the primary amino acid sequence of a protein determines its specific three-dimensional conformation. The ability of a protein to fold depends upon the rotation of bonds in a protein chain. A fold is only marginally stable at room temperature. The more stable, native conformation, is stabilised by weak noncovalent interactions involving main chain and side chain atoms [6]. The findings that protein conformation can be disrupted by denaturants and refolding of the chain occurs following removal of denaturing agent, led to the proposal of the principle of self-assembly by Anfinsen [7]. This states that the amino acid sequence encodes all the information required for a protein chain to reach its correct conformation [6-8]. Internalisation of hydrophobic side chains inside the folded protein and exposure of the hydrophilic side chains to the aqueous environment is the major driving force for the phenomenon of spontaneous refolding [6]. Studies on smaller proteins suggest that a nucleation event is required, whereby a critical core is formed by several amino acids coming together, to drive the formation of the rest of the structure [9-11]. Vendruscolo and coworkers [11] found that definition of the overall architecture of the protein

fold by a small number of fundamental interactions was adequate for the instigation of the protein folding process.

The Levinthal paradox, ‘how a sequence finds its specific native structure in a finite time’ has been key to discussion on protein folding for decades [5, 12]. An initial suggestion of a solution to this problem proposed that folding proteins follow specific predetermined pathways which involve a series of mandatory steps through well-defined partly folded states [5, 13, 14]. This classical pathway was thought to include folding intermediates for which the unfolding transitions can be cooperative or non-cooperative (Figure 1.1) [15, 16].



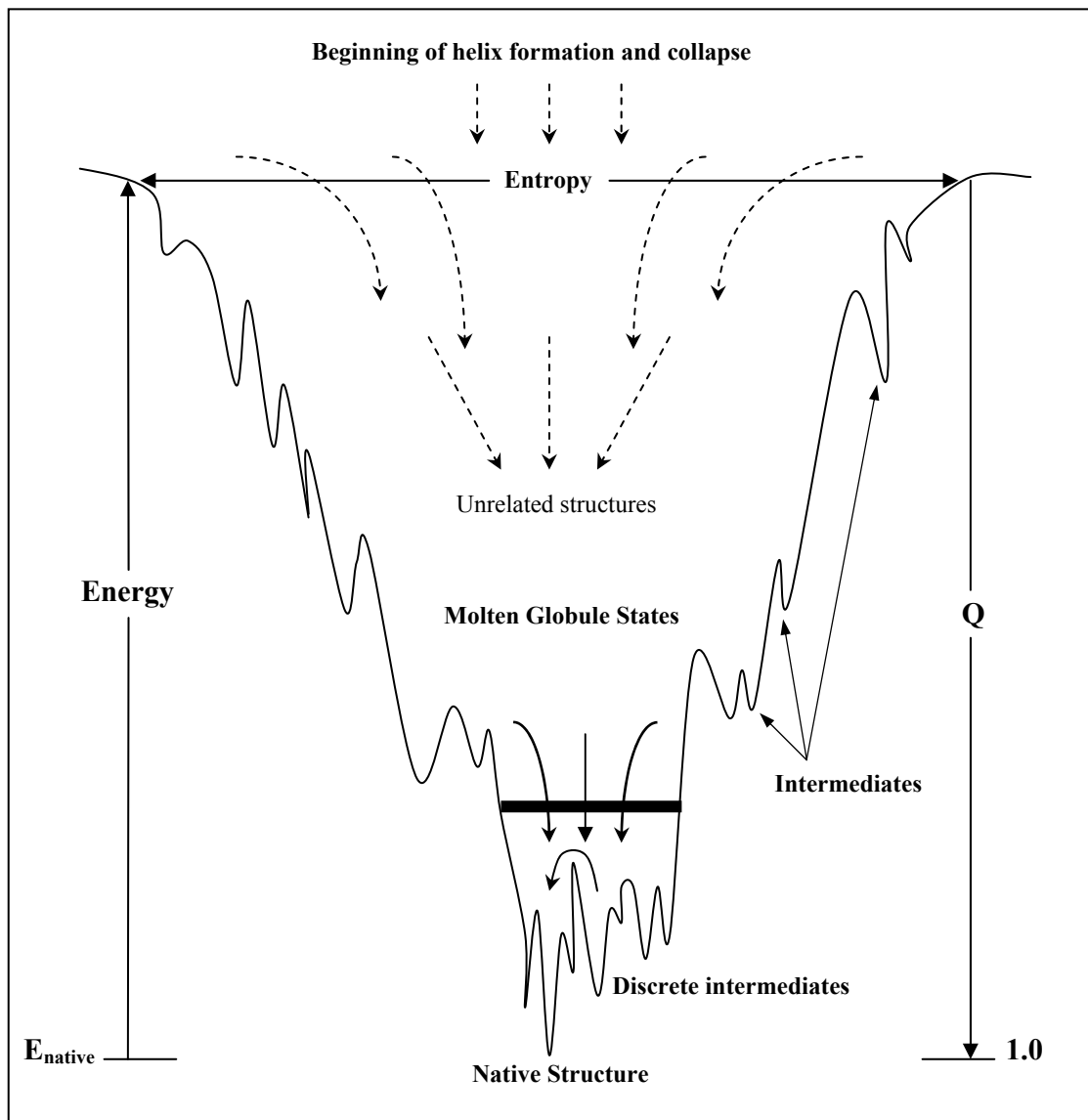
[17]

Figure 1.1. A classical kinetic folding pathway. U represents the unfolded state, I_1 and I_2 are classical intermediate states whilst N is the native state [17].

The Levinthal paradox could be avoided if these pathways were specific enough so that only a small region of conformational space would be sampled [5].

A more widely accepted theory states that the protein folding process involves a stochastic search of the conformations accessible to a polypeptide chain [17-19]. It describes protein folding in terms of energy surfaces and energy landscapes and proposes that folding involves

the organisation of statistical ensembles [5, 19]. This statistical landscape theory is represented as a funnel that has pockets which represent transiently trapped proteins as intermediate species (Figure 1.2). A rugged funnel-like model best represents the free energy landscape with a preferred direction of flow towards the native state. The folding landscape is rugged as biomolecular chains can take up many inappropriate conformations during the folding process and there are incorrect interactions between residues [19]. An ensemble of unfolded protein molecules occupies the top of the funnel. As proteins proceed down the funnel towards the native state, the energy pockets trap proteins in different intermediate conformations. Progression of the protein down the rest of the funnel requires overcoming of the energy barrier for each pocket. Therefore, in order for a trapped intermediate to be experimentally observed, the majority of molecules must occupy that particular state. Some of the trapped intermediates can act as precursors for aggregate formation, although it is thought that aggregate accumulation is prevented by an evolutionary fast-folding mechanism [20]. The folding pathway for each molecule is not necessarily identical to that of other protein molecules as it proceeds down the funnel towards the most stable, native state [19].



Adapted from [19]

Figure 1.2. A protein folding free energy landscape. There is an ensemble of unfolded protein molecules at the top of the funnel. As the proteins move towards the native state they are trapped in energy pockets as stable intermediates. The energy barrier for each pocket needs to be overcome before the protein can proceed to the native state. The solvent-averaged energy (E) and the fraction of native-like contacts (Q) describe the position of states within the funnel [19].

The native protein conformation is generally the energetically most stable, having the lowest free energy [7]. Exceptions to this include haemagglutinin, the extracellular bacterial alpha-lytic protease and more notably, the serpin protein superfamily (described in detail in section 1.9). In these cases the native fold exhibits metastability, allowing for the structural flexibility

that is required for serpins to carry out their inhibitory function [21, 22]. This fragile equilibrium between stability and flexibility however is easily disrupted. For example, small changes in temperature, pH or amino acid sequence can instigate large conformation changes and subsequent destabilisation of the native protein state to more stable but inactive protein conformations, such as amyloids.

1.2 Protein misfolding and aggregation

Under normal physiological conditions, cells have developed the ability to maintain extremely high protein concentrations whilst avoiding protein aggregation [23]. This is particularly difficult given that proteins exist in different protein forms depending on their specific biological functions in the cellular environment. For example, whilst fully folded proteins are necessary for enzyme activity, unfolded or partially folded protein conformations may be required for functions such as translocation across membranes [3].

Therefore proteins (unfolded and folded) have to be carefully managed in the cell. Indeed, any extent of misfolding can cause malfunction and disease. Many familial diseases, associated with protein misfolding, can be due to a single amino acid substitution that results in the loss of cellular control and thereby the death of the organism which the cell is a part of [24, 25]. Alzheimer's disease, Creutzfeldt-Jakob's disease, late onset diabetes, spongiform encephalopathies and cystic fibrosis are all examples of protein misfolding diseases [14, 26]. These diseases are thought to be a consequence of aberrant protein behaviour resulting from the conditions of biological systems changing at a much faster rate than that set by evolution [27].

Evolutionary selection has enabled cells in physiological conditions to concentrate proteins to very high levels whilst avoiding aggregation [23]. Loss of this maintenance of solubility results in molecular aggregation, which plays a fundamental role in a number of ageing and degeneration diseases [27-29]. Refolding studies have demonstrated irreversible aggregation of chains that fail to refold correctly, forming non-functional aggregates [30]. It is perhaps obvious that aggregation initially results in dimers [31] which may be soluble but continued aggregation produces insoluble moieties. Aggregation is thought to be caused by the association of partially folded intermediates rather than native proteins [32]. The specificity of protein aggregation, such that only identical or very similar chains can aggregate, was demonstrated by Goldberg and coworkers. They studied the renaturation of *E. coli* tryptophanase and reported that it refolded to a native-like conformation following denaturation with 8 M urea. Additionally, they showed that the addition of crude total extract in 8 M urea from *E. coli* cells free of tryptophanase to denatured tryptophanase in 8 M urea does not change the percentage of tryptophanase that aggregates upon dilution of the urea dialysis [33]. Furthermore, *in vitro* aggregation studies of the coat protein and tailspike protein of Salmonella bacteriophage P22, revealed that partially folded chains of these two proteins aggregate with themselves but not each other [34].

Misfolding is thought to be distinct from aggregation but there are two different views on its concept. The first view interprets misfolding as ‘a partially folded polypeptide conformation that is stable enough not to be able to proceed to its functional conformation on a biologically relevant timescale’ [30]. An alternative view is that misfolded chains are on the correct folding pathway but require some unfolding to reach their native state [35]. Current protein folding models view misfolded chains as partially folded chains that vary in their tendency to aggregate with one another [36]. Therefore, protein misfolding is considered to be the molecular basis of protein aggregation whereby a polypeptide loses its ability to attain its

closely packed, native structure and consequently populating non-folded, partially folded or incorrectly folded states. The loose packing of these non-native structures results in the exposure of the hydrophobic core to the solvent; enhancing the propensity to nucleate oligomeric assemblies [14, 29, 37, 38]. These ‘aggregation nuclei’ act as templates for the recruitment of other misfolded or partially folded molecules, increasing the size of the assemblies and consequently yielding fibrillar aggregates [39].

1.3 Macromolecular crowding

Macromolecular crowding occurs in the inside of cells. Theories suggest that it effects interactions between macromolecules, resulting in it affecting the rates and equilibria of macromolecular reactions. Crowding, also termed ‘the excluded volume effect,’ describes the total concentration of macromolecules in the intracellular environment to be so high that a vast proportion of the interior of the cell is occupied and therefore unavailable to other molecules [23]. This volume exclusion is due to non-specific steric repulsion, whereby any two macromolecules cannot occupy the same place at the same time [40, 41]. The fraction of this excluded intracellular volume depends on the quantity, size and shape of the macromolecules in the cellular compartments [41]. Steric repulsion arises from the fact that each atom within a molecule occupies a certain amount of space. If atoms are brought too close together, there is an associated cost in energy due to overlapping of electron clouds, which may affect the preferred conformation and reactivity of the molecule. It is a destabilising interaction that causes an increase in the total free energy of a system.

Crowding is thought to enhance two properties of protein folding; the collapse of polypeptide chains and the tendency of partly folded chains to associate into aggregates [6, 23]. This favouring of aggregation is due to the effect of increased thermodynamic activity of partially

folded polypeptide chains by crowding. Polypeptide chains that fold slowly are particularly aggregation-prone as faster folding chains are able to internalise hydrophobic residues before they encounter those on other chains. Furthermore, the functional activity of molecular chaperones (described in section 1.4) is also thought to be enhanced in crowded milieu in order to stimulate their association with partly folded chains [41]. The folding patterns of lysozyme have been studied to test these hypotheses, under both reductive and oxidative conditions [42-45]. The reduced form of the protein is known for its high susceptibility to aggregation [46], allowing comparison of the effects of crowding on aggregation-prone and aggregation-resistant states of the same protein [45]. In a crowded environment, the refolding of oxidised hen lysozyme from denaturant is rapid and highly efficient whereas the refolding of reduced lysozyme is greatly reduced due to aggregation [23]. Protein disulphide isomerase (PDI) prevents this aggregation by acting as a catalyst, increasing the rate of disulphide bond formation and acting as a molecular chaperone, shielding hydrophobic surfaces [23, 45, 47]. Van den Berg and coworkers observed a 30% increase in the refolding yields of lysozyme on addition of PDI, in the absence of crowding agents. Moreover, they found that presence of crowding agents enhanced this refolding further. For example, in the presence of BSA and Ficoll, the refolding yield of lysozyme increased from less than 3% in the absence of PDI to 30-40% at the highest concentration of PDI used [45]. These findings suggest that the function of PDI is enhanced under crowded conditions.

1.4 Molecular Chaperones in Protein Folding

Frequently, the biological requirements of the cell are not met due to the low inherent efficiency of the folding process. Molecular chaperones function to overcome this problem without containing steric information that may facilitate correct protein folding [8]. Chaperones are prokaryotic and eukaryotic proteins that interact with unfolded or partially

folded protein subunits (for example, nascent chains emerging from the ribosome or extended chains being translocated across subcellular membranes) [6, 48, 49]. They recognise hydrophobic regions of proteins that are exposed to the solvent in the unfolded but not the native state [50].

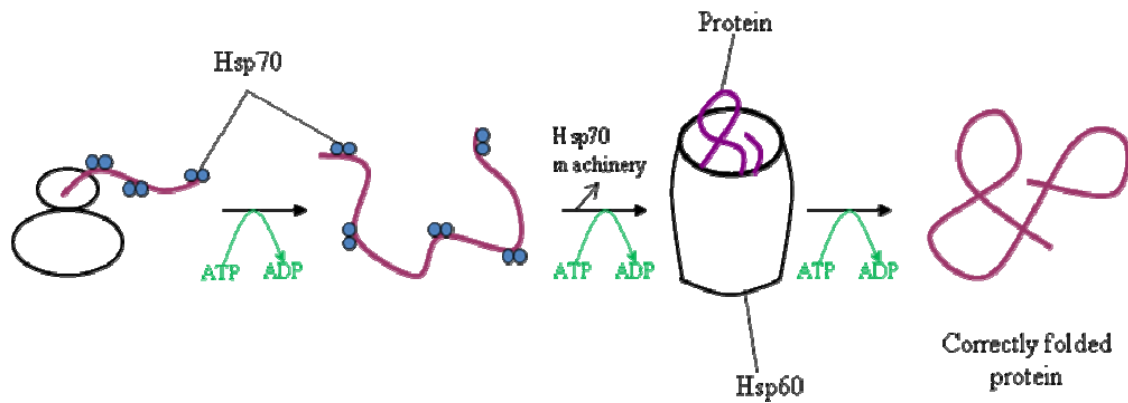
The term ‘molecular chaperone’ was first applied to the acidic nuclear protein, nucleoplasmin by Laskey and coworkers [51]. Nucleoplasmin is required in *Xenopus* eggs for the assembly of nucleosomes from DNA and histones [49, 52, 53]. In the absence of nucleoplasmin the mixing of separated DNA and histones at physiological ionic strength gives rise to aggregate formation. In the presence of nucleoplasmin however, this aggregation is prevented and the formation of the nucleosome cores arises [49]. Nucleoplasmin interacts with histones in such a way that shields their positive charge without binding to either DNA or nucleosomes. Thereby, it mediates nucleosome assembly by preventing incorrect interactions between histones and DNA [54]. It reduces electrostatic repulsion between histones thereby promoting histone-histone interactions and competes with the electrostatic attraction between DNA and histones, minimising the formation of non-specific aggregates [49]. Nucleoplasmin is not a part of the nucleosome, nor does it comprise steric information for nucleosome assembly; it is required solely for assembly [52, 54].

Molecular chaperones have since been characterised from animal, plant and viral sources. For instance, immunoglobulin heavy chain binding protein or BiP is found in the lumen of the ER of lymphoid cells and binds non-covalently to newly synthesised immunoglobulin heavy chains [49]. BiP binds to Ig heavy chains that do not have light chains attached, recognising hydrophobic regions (on the heavy chain) that are later covered by light chains. It is therefore thought that BiP may be involved in preventing and reversing heavy-chain aggregate formation, thereby assisting assembly of immunoglobulin molecules in antibody secreting

cells [55]. Grp 78, an ER protein the same as BiP, is produced by fibroblasts that are depleted of glucose. A possible role of grp 78 may be to bind to underglycosylated proteins in these fibroblasts and prevent the formation of insoluble aggregates [56]. Mutant forms of viral haemagglutinin that are unable to fold correctly have been shown to bind to BiP for longer periods whereas unfolded haemagglutinin molecules only bind the chaperone transiently as they enter the ER [57]. The presence of BiP in a variety of cells suggests that this heavy-chain binding protein may play a role in enhancing the solubility of proteins other than immunoglobulins [49].

The primary structure of BiP is related to that of a major heat shock protein in animal cells, Hsp70 [55]. The Hsp70 multigene family consists of genes that are related to the major heat shock protein and encode chaperones that are involved in the correct folding and transport of proteins. They are present in all cells except for archaebacteria. The Hsp70s are a group of highly conserved ATPases with molecular weights of approximately 70 kDa [50, 58]. Constitutively expressed members of this family are involved in the stabilisation of polypeptide chains as they are being synthesised on ribosomes. The Hsp70 and Hsp60 protein chaperone families are involved in protein folding under both normal and stress conditions. The binding and release of substrate proteins by both classes of chaperone is regulated by ATP-binding and hydrolysis. Close association of the polypeptide chain and molecular chaperone is maintained during the translation and folding processes [50]. The primary role of these interactions is thought to be to prevent ineffective reactions that are not part of the pathway. These two chaperone proteins act by two different mechanisms and there are two distinct chaperone actions that occur. Hsp70, which can act as either a monomer or dimer, is involved primarily in shielding the hydrophobic surfaces on nascent polypeptide chains until approximately 100-200 amino acids are accessible for effective folding. In contrast, the Hsp60s form double ring structures comprising of an internal cavity that acts as a barrier

between the yet unfolded protein and the cellular surroundings. In doing so, they preclude protein aggregation as well as enabling folding to the native state in an ATP-dependent reaction (Figure 1.3) [59].



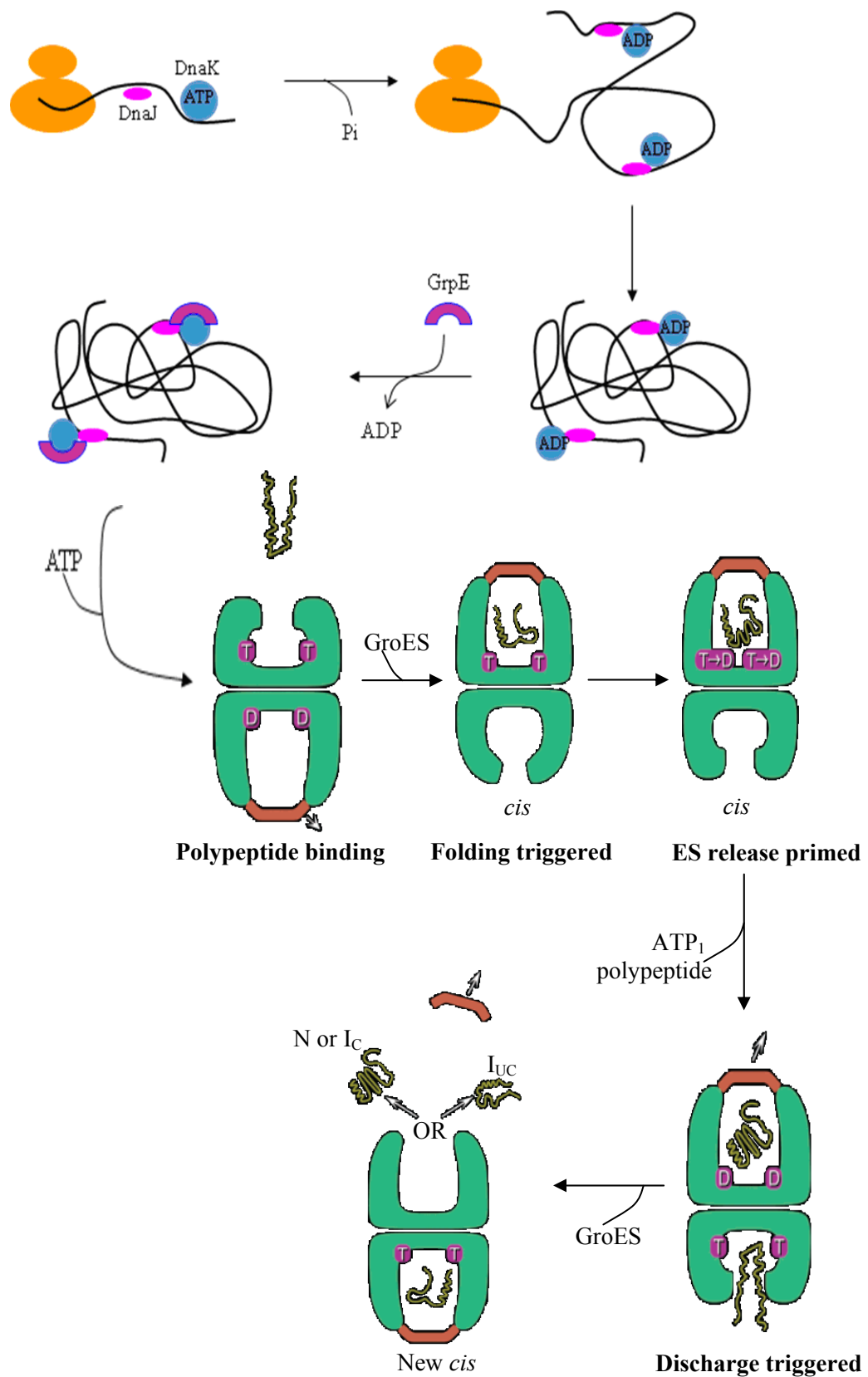
(Adapted from [59])

Figure 1.3. The sequential action of molecular chaperones Hsp60 and Hsp70. The Hsp70 acts early in the life of the protein by recognising and binding to amino acid chains. The Hsp60 forms a structure into which misfolded proteins enter and are transferred. Both Hsp70 and Hsp60 hydrolyse ATP, which in turn allows the binding and release of the subsequent protein that helps the correct folding of the protein.

Chaperones have been well studied in *E. coli* where it has been shown that functionality of Hsp70 and Hsp60 requires cooperation with other proteins. DnaK, the *E. coli* Hsp70, cooperates with DnaJ, another chaperone and GrpE, a nucleotide exchange factor. Both DnaK and its partner chaperone DnaJ have been shown to copurify with polyribosomes and have the ability to bind co-translationally to nascent polypeptide chains emerging from ribosomes. The binding of DnaJ to the unfolded protein requires the hydrolysis of DnaK-bound ATP and ADP. This allows the formation of a stable, ternary complex [60]. DnaJ stabilises the ADP-bound state of DnaK which has a high affinity for the unfolded peptide. Whereas the prevention of protein aggregation requires the binding of DnaK/DnaJ to exposed hydrophobic polypeptide regions, folding of the protein to the native state requires the release of bound structures. This is dependent on the regulation of DnaK by the regulator protein GrpE [60].

The regulation of low ATPase activity of DnaK is regulated by DnaJ so that ADP is tightly bound to DnaK, DnaJ and the unfolded protein complex. The interaction of GrpE with the ATPase domain of DnaK results in dissociation of ADP. Subsequently, there is binding of ATP to DnaK which results in dissociation of the ternary complex, allowing transfer of the unfolded protein to the bacterial Hsp60, GroEL.

GroEL is a large complex, which comprises two stacked rings of seven 60 kDa subunits, forming a quaternary structured double-toroid with a central cavity [60-62]. GroES, an essential cofactor of GroEL, forms a 70 kDa heptameric ring that binds to GroEL and regulates its ATPase activity. ADP is bound to the binary GroEL: GroES complex. The substrate protein itself is thought to induce the exchange of ADP for ATP. The binding affinity of GroEL for ADP is decreased upon binding of the substrate protein to the interior of the GroEL cylinder. This initiates a reaction cycle which involves the hydrolysis and exchange of nucleotides [50, 62]. Transient dissociation of GroES occurs as a consequence of substrate-protein binding, permitting ATP binding to the GroEL toroid. This results in a reduction in the binding affinity of GroEL for substrate protein, allowing its release into the cavity where folding is initiated (Figure 1.4). Multiple rounds of interaction with GroEL may be essential for complete folding of most proteins [50].



(Adapted from [50, 63])

Figure 1.4. A simplified pathway of chaperone-assisted protein folding in the bacterial cytosol.

Hsp70 (DnaK) and DnaJ interact with the polypeptide chain co-translationally. The ternary complex

between DnaK (in an ADP bound state), DnaJ and the unfolded polypeptide keeps the protein in the soluble form. Binding of GrpE to DnaK instigates a conformational change, leading to dissociation of ADP and binding of ATP. Subsequently, there is release of the bound polypeptide. Completion of folding of most polypeptides requires transportation to the GroEL-GroES reaction cycle. A GroEL-GroES-ADP asymmetric complex binds to non-native polypeptides in a collapsed, unstructured state via hydrophobic interactions with its apical domains. Binding of ATP followed by GroES to the same ring as the polypeptide results in large conformational changes in the GroEL apical domains. This causes movement of the hydrophobic regions away from the surface of the cavity, allowing GroES to cap the GroEL ring, forming a *cis* folding-active complex. The polypeptide is then released into the central cavity where it folds in this sequestered chamber whose walls are not hydrophilic. This environment favours internalisation of hydrophobic side chains and exposure of hydrophilic ones. ATP hydrolysis weakens the *cis* complex, discharging the ligand from the *cis* ring. The released polypeptide may be in the native state (N) or one committed to it (I_C), or may have failed to reach the native state (I_{UC}), in which case it can bind GroEL for another attempt at folding. The binding of GroEL, ATP and polypeptide to the same ring forms a new *cis* complex on the previous *trans* ring to begin the cycle again.

Bacteriophage head assembly had been known to be affected by mutations in genes encoding GroEL and its co-factor GroES [64]. A yeast mutant, which affected the function of mitochondrial Hsp60, was unable to oligomerise proteins imported from the cytosol into the mitochondria, providing evidence for the essential function of chaperones *in vivo*.

Chaperones are therefore a group of unrelated protein families that are involved in the stabilisation of unfolded polypeptide chains, unfolding of proteins for translocation across membranes or degradation and assistance to correct protein folding and assembly.

Hsp47 shares similarities and differences with the GroEL chaperone. Similarly, it is involved in preventing the aggregation of unfolded or partially folded collagen chains as well as

maintaining incomplete collagen chains in the ER until they are folded correctly before transporting them to the Golgi complex. However, while GroEL is not specific for the polypeptide chain it acts on, Hsp47 is substrate specific.

1.5 Hsp47

Molecular chaperones are also linked to changes in cellular environment such as heat-shock and environmental stress, as the rate at which they are synthesised is greatly increased when cells are exposed, for very short periods, to elevated temperatures [48] or other stressful stimuli such as increased ionic stress, changes in pH and exposure to toxic agents and detergents. The heat shock or stress response is a transitory reaction to stressful conditions that provides cellular protection and prevents organisms from being severely damaged [65]. Hsp47 has been characterised as a stress induced protein in animals [66].

1.5.1 The Discovery of Hsp47

In 1983, a glycoprotein of 47 kDa was isolated from mouse embryo parietal endoderm cells, which was shown to bind native type IV collagen and gelatin (denatured type I and type III collagen). As it possessed collagen binding abilities this protein was named colligin [67]. Subsequently a number of similar proteins were isolated from 3T3 fibroblasts and human epidermal keratinocytes, all of which were recognised as colligin [68], rat skeletal myoblasts (termed gp 46) [69], and chick embryo fibroblasts [70], which was termed Hsp47. Upon analysis of the gene sequences of the proteins, it was revealed that these proteins were highly similar and could be considered species variants of the same protein.

1.5.2 Expression of Hsp47

Hsp47 is a heat shock protein; therefore its expression is heat inducible. This is an unusual property for ER localised stress proteins that are usually induced by an increase in the concentration of the unfolded protein folded in the ER. Hsp47 is unique in that it is the only ER resident protein that is induced not by ER stresses but by cytosolic stresses [71]. The Hsp47 gene contains a conserved heat shock element (HSE) that is located -180 bp from the transcription initiation site (Figure 1.5). The expression of Hsp47 is induced upon the binding of heat shock factor 1 (HSF1) to the HSE [72]. The binding of HSF1 is in response to stress in the cytoplasm, such as heat shock and treatment with arsenite [73].

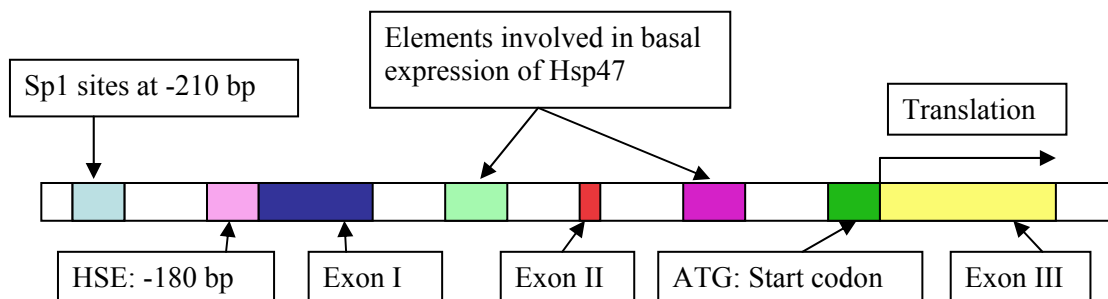


Figure 1.5. The regulatory elements of the Hsp47 gene.

Analysis of the promoter region has revealed that a Sp1 binding site at -210 bp from the transcription initiation site is essential for the constitutive expression of Hsp47. However, it is not responsible for regulating the level of the expression. This is done by two elements of about 30 bp that are present in the first and second introns of the Hsp47 gene (Figure 1.5). Regulation by these elements is at a level that correlates with that of the cellular expression of collagen [74]. Experiments studying the induction of Hsp47 by forms of stress such as heat shock showed an increase in the amount of Hsp47 but the collagen levels were either unaffected or decreased slightly. Analysis of the Hsp47 gene [75] revealed a heat shock element –nAAGn– at the promoter region. This demonstrated that the induction of Hsp47 was due to a heat-shock-element-heat-shock-factor system.

Low levels of Hsp47 are also constitutively expressed and its expression is known to correlate with collagen in several cells and tissues. Malignant fibroblasts show reduced expression of Hsp47 whereas in differentiating F9 cell lines, there is an increase in Hsp47 expression. On the other hand, PC12 cells or mouse M1 cells that do not produce collagen show no production of Hsp47 [76].

1.6 Collagen biosynthesis

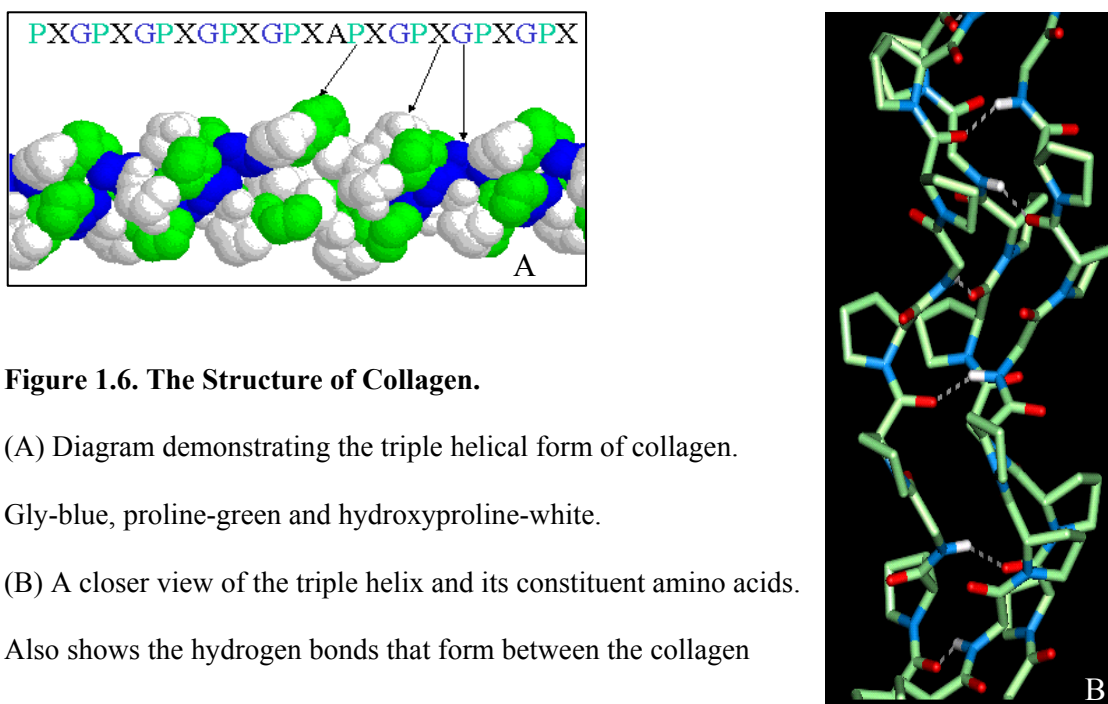
Collagen is the most abundant mammalian structural protein and is a major component of the extracellular matrix where it forms supermolecular structures that provide mechanical strength and elasticity in connective tissues [59, 74]. It has many significant roles, such as controlling cell shape and differentiation, allowing regeneration of broken bones, providing a scaffold for cells in our bodies and permitting wound healing.

Collagen, like other secretory proteins, is assembled, folded and processed in a specific manner. It is secreted from eukaryotic cells through the process of exocytosis and as with all proteins that undergo exocytosis, collagen enters the ER co-translationally during synthesis [77]. The ER provides an environment that promotes protein folding and resembles the extracellular space where the protein is ultimately localised. It also has a quality control mechanism, which ensures that the protein is correctly folded before it moves to the next stage of the secretory pathway [78].

Assembly of Collagen

At least 27 types of collagen have been identified in mammalian tissues, all of which are superpolymers with the same general structure of three collagen chains intertwined to form a triple helix (Figure 1.6). The triple helices can either consist of three identical α chains and be

homotrimeric or consist of three different α chains and be heterotrimeric. A right-handed triple helix is formed when each individual left-handed helix associates with two other α chains. The α chains are initially synthesised as polypeptides with propeptides located at the C- and N- termini. Assembly of the triple helix, which uses a zipper-like mechanism, initiates at the C-terminal propeptide and progresses in the C to N terminal direction. Once on the cell surface, N- and C- propeptidases are used to cleave the N- and C- propeptides leaving the remaining triple helical regions. Consequently, each α chain consists of 1050 amino acids and the excision of the propeptides allows polymerization of collagen molecules into fibrils in the extracellular space. The stabilisation of the triple helix is partly due to the presence of glycine moieties on the inside of the triple helix and the hydroxylated proline residues. Van der Waals forces, extensive solvent hydrogen bonds and interchain disulphide bonds hold the triple helix together [74, 78-81].



The triple helical conformation is stabilized by a Gly-X-Y repeating motif, where X is any amino acid and Y is proline or hydroxyproline. The triple helical structure of collagen is due to the unusual abundance of three amino acids; glycine, proline and hydroxyproline. It is these

amino acids that make up the characteristic Gly-Pro-X motif and each amino acid has a specific function [59]. Hydroxylation of the proline residue occurs while it is co-translationally transported to the ER. The glycine residue is highly conserved as its hydrogen sidechain is the only one small enough to accommodate the centre of the helix enabling the three strands to pack together tightly [79]. The three chains are held together by hydrogen bonds between the peptide bond NH of a glycine residue with a peptide carbonyl on an adjacent polypeptide chain. The fixed angle of the C-N peptidyl-proline or peptidyl-hydroxyproline bond allows each polypeptide chain to fold into a helix with such a geometry that allows three triple helices to twist together to form a three-stranded helix [59].

The repeated Gly-X-Y motifs have globular domains at the N- and C- propeptide terminals. These short segments at each end of the collagen chain are particularly important in collagen fibril formation [79]. They do not assume the triple helical structure and comprise of the unusual amino acid, hydroxylysine (described below). Covalent aldol cross-links between two lysine or hydroxylysine residues at the C-terminus of one collagen molecule and two similar residues at the N-terminus of an adjacent molecule stabilise the side-by-side packing of collagen molecules, thereby forming a strong fibril.

Collagen molecules are extensively post-translationally modified in the ER. Some of the proline residues in the Gly-X-Y motif are post-translationally modified via an ER resident enzyme, prolyl-4-hydroxylase, to form hydroxyproline. The addition of the hydroxyl group helps to stabilise the triple helical structure and facilitate the cross-linking of the secreted collagen molecules in the extracellular space. Lysine residues at the Y position can also be hydroxylated via a different enzyme, lysyl hydroxylase. This post-translational modification results in the production of hydroxylysine residues that again play a role in the cross-linking of secreted collagen molecules. Hydroxylysine residues are also sites for N-linked

glycosylations. These post-translational modifications can only occur on individual α chains prior to the formation of the triple helix. Once the collagen triple helical conformation is formed, the α chains can no longer act as substrates for the post translationally modifying enzymes [81, 82].

Vitamin C (ascorbate) is required for the numerous physiological functions in the human body, including synthesis of collagen. Vitamin C is known to induce lipid peroxidation and reactive aldehydes, a step thought to be required for collagen expression, collagen mRNA levels and collagen production in cultured human fibroblasts [83]. Vitamin C and iron in its ferrous state (+2) are required functional prolyl and lysyl hydroxylase. Iron is bound to the hydroxylase as a cofactor. Hydroxylation of proline and lysine results in the oxidation of iron from a ferrous state (+2) to a ferric (+3) state. Ascorbate acts as reducing agent, maintaining iron atoms in their reduced states. Thus, vitamin C enables the vital hydroxylation reactions to continue and subsequently the formation of collagen [84]. Vitamin C deficiency causes deactivation of prolyl hydroxylase which results in malformed collagen that denatures at 24 °C. Severe vitamin C deficiency results in scurvy [85].

Following the formation of the triple helices, the procollagen molecules are transported from the ER to the Golgi body, continuing their journey along the secretory pathway (Figure 1.7). Progression through the Golgi body results in the association of collagen trimers that are required for fibril formation following exocytosis. The cleavage of the C- and N- terminal propeptides instigates a dramatic diminution in the solubility of collagen, thereby driving the assembly of stable collagen fibrils.

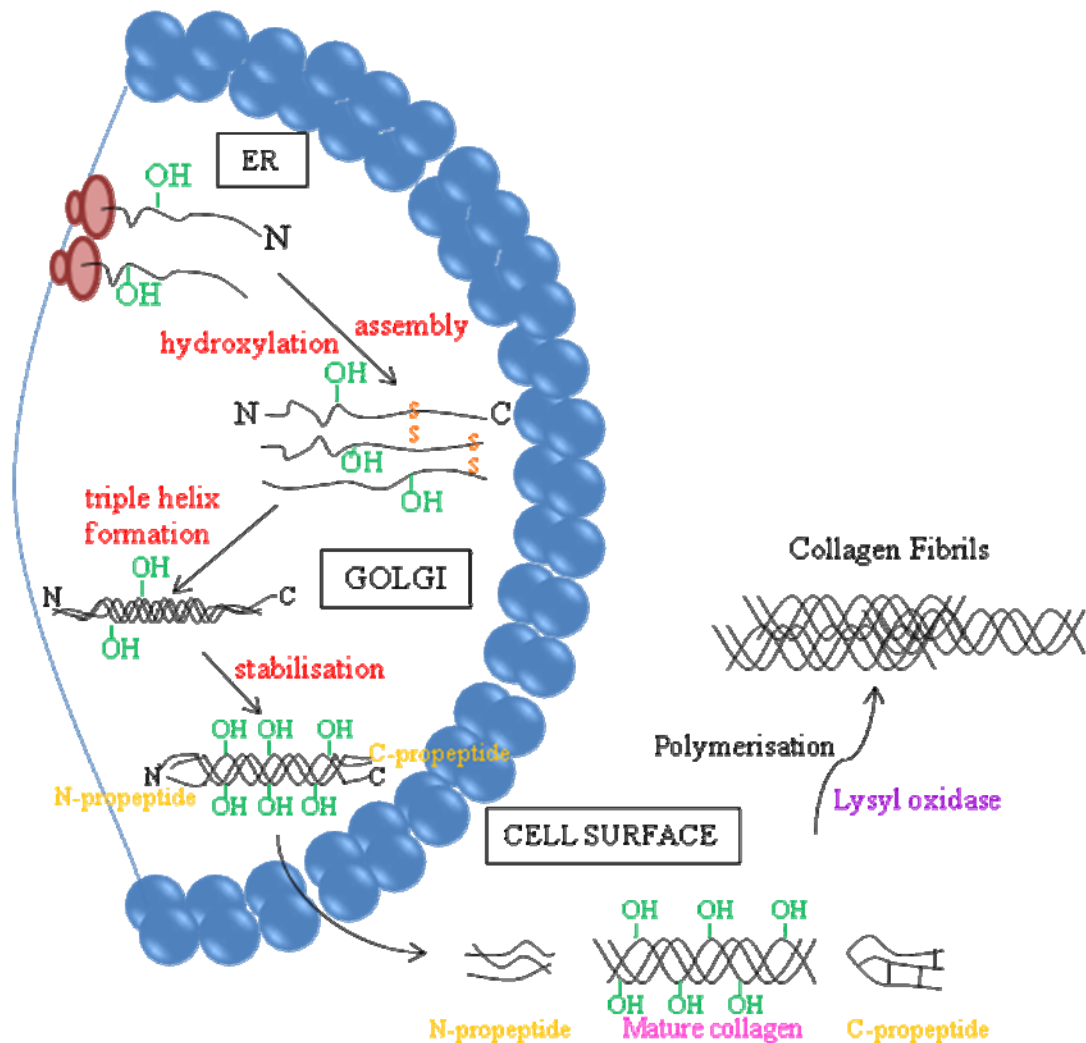


Figure 1.7. Collagen fibril formation. The formation of the procollagen helix requires three α helices. Disulphide bond formation and hydroxylation stabilise the triple helix, which moves from the ER to the Golgi apparatus. Subsequently, the procollagen is processed to transport vesicles that fuse with the plasma membrane, allowing the secretion of procollagen into the extracellular space. Once in the extracellular space, propeptidases cleave off the prodomain regions. Collagen fibrils are formed on polymerisation of the triple helices and lysyl oxidase enhances cross-linking and provides rigidity by the production of aldehydes.

1.7 Hsp47 and collagen

Hsp47 is a collagen-specific molecular chaperone. The glycoprotein is ubiquitous, abundant and highly conserved and like other molecular chaperones it assists in protein folding and refolding both *in vivo* and *in vitro* as well as protecting protein unfolding due to stress.

However, whilst it is likely that the majority of molecular chaperones interact with a large variety of globular proteins, Hsp47 specifically interacts only with polypeptides of procollagen and the fibrous protein collagen [86].

Immunofluorescence and immunogold-electron microscopy have been used to demonstrate that Hsp47 localises to the ER [87]. The amino terminal of Hsp47 consists of a signal sequence and the RDEL (Arg-Asp-Glu-Leu) sequence, an ER-retention sequence is found at the carboxyl terminal. Hsp47 also contains two glycosylation sites that are thought to be glycosylated by mannose-type sugar moieties. The presence of both the ER-retention sequence and the glycosylation sites are further evidence for Hsp47 localising to the ER.

Whilst Hsp47 is localised to ER, collagen is processed to the Golgi body where it is secreted via exocytosis. If Hsp47 bound to collagen does get carried to the Golgi body via vesicular transport, the RDEL sequence at its C-terminus is recognised by specific receptors, which retrieve it and return it to the ER [79]. Artificial removal of this sequence results in the secretion of the protein from the cell by exocytosis.

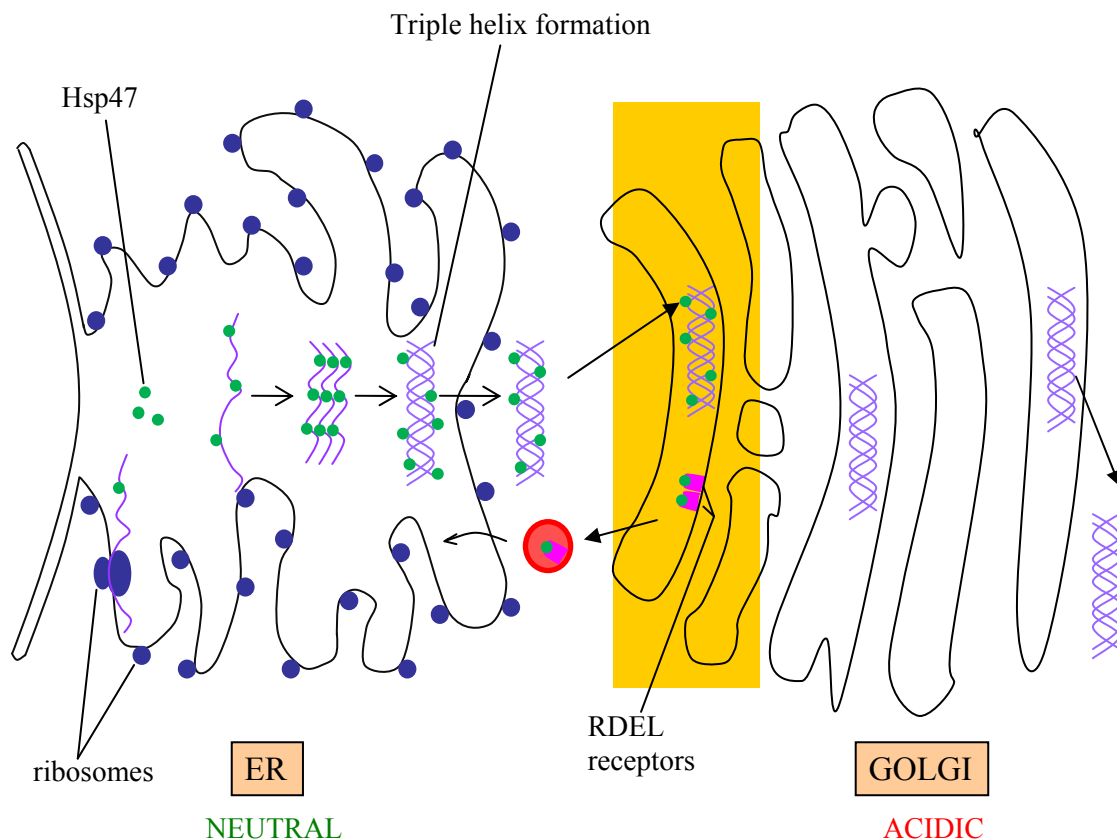


Figure 1.8. The role of Hsp47 in procollagen synthesis. Hsp47 transiently binds to procollagen prior to its secretion. Binding takes place immediately after the co-translation of the nascent polypeptide chain into the ER. Subsequently, triple helix formation occurs by three α chains interacting at their carboxyl termini via inter-chain disulphide bonds. Correct helical formation is ensured by progression in the C- to N- direction, allowing accurate alignment and orientation of the helix. The triple helix then continues to the *cis*-golgi compartment where it dissociates from the Hsp47 before proceeding to the *trans*-golgi for secretion to the cell surface. Receptors in the *cis*-golgi recognise the RDEL peptide and cycle Hsp47 molecules back to the ER. Misfolded procollagen aggregates in the ER and has been shown to be eliminated via an autophagy-lysosome pathway. The red sphere represents a transport vesicle.

1.7.1 The Ability of Hsp47 to Recognise the Triple Helix

As mentioned in the previous section, in cultured cells, the expression of Hsp47 is correlated with various types of collagen. Furthermore, Hsp47 is known to be a collagen-specific molecular chaperone. However, the specificity of Hsp47 binding was unknown so

experiments using peptide-libraries were performed in order to determine the specific binding sites for Hsp47 using the recombinant chaperone protein and synthetic collagen model peptides [88]. Binding assays allowed the identification of structural requirements for substrate recognition and regulation of the binding. The peptide (Pro-Pro-Gly)_n provided sufficient structural information necessary for recognition by Hsp47. A structure-activity relationship study using synthetic analogues of (Pro-Pro-Gly)_n identified chain length and primary structure to be important for the interaction. Koide and coworkers previously identified that prolyl-4-hydroxylation of the second Pro residue in the Pro-Pro-Gly repeat abolishes the chaperone-substrate interaction. Thus, they proposed a model whereby the dissociation of Hsp47 from collagen occurred during prolyl-4-hydroxylation in the ER. The peptides used included (Pro-Pro-Gly)₅, (Pro-Hyp-Gly)₅, (Pro-Pro-Gly)₁₀, (Pro-Hyp-Gly)₁₀, polyglycine and poly-L-proline. Solid-phase binding assays were performed in which peptides and porcine Type I collagen were immobilised onto cyanogen bromide-activated-Sepharose. Following incubation of Hsp47 lysate with affinity beads, proteins were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualised using coomassie brilliant blue (CBB) staining. A competition assay was also performed where a bed of type I collagen or (Pro-Pro-Gly)₁₀-immobilised sepharose was incubated with Hsp47. After washing the beads, a competitive peptide was added to the mixture. Proteins present in the supernatant and those retained on the beads were analysed by SDS-PAGE and CBB staining [88-90].

The results obtained showed that the residues that were found to be most frequently bound to the Xaa and Yaa positions were Pro and Hyp (4-hydroxyproline), respectively. Solid-phase pull down assays revealed that Hsp47 did not recognise a specific amino acid sequence but recognised structural similarities between these sequences. An initial assay contained poly-L-

proline, polyglycine, (Pro-Pro-Gly)₅ and (Pro-Pro-Gly)₁₀. Rat Hsp47 did not interact with polyglycine, poly-L-proline or (PPG)₅ [88].

The following points were noticed in relation to substrate recognition of Hsp47:

- (i) The minimum number of (PPG)_n repeats required for the interaction is between 5 and 10.
- (ii) It was thought that prolyl-4-hydroxylase (P4-H), an enzyme also present in the ER had similar substrate recognition to that of Hsp47.
- (iii) The Yaa position of Xaa-Yaa-Gly does not need prolyl 4-hydroxylation for Hsp47 interaction.

To further confirm these findings, analogues of (Pro-Pro-Gly)₈ with different Pro and Gly arrangements were used. Some of the analogues used were polyglycine, (Pro-Gly-Gly)₈, (Pro-Gly)₁₂, (Pro-Pro-Pro-Gly)₆ and Poly-L-proline. Only (Pro-Pro-Gly)₈ showed any interaction with Hsp47 demonstrating that the presence of Gly at every third residue is important for substrate recognition. This demonstrated that the simplest collagen model peptide (Pro-Pro-Gly)_n could be used as a binding substrate for Hsp47 [88].

Koide and coworkers also found that most of the Pro-Pro-Gly sequences were converted to Pro-Hyp-Gly in the procollagen biosynthetic pathway. This conversion in the Yaa position due to P4-H prevents the binding of Hsp47 before the formation of the triple helix. This inconsistency encouraged Koide and coworkers to identify Hsp47-binding sequences other than (Pro-Pro-Gly)_n. Various collagen model peptides were used which led to the discovery of Xaa-Arg-Gly triplets as an Hsp47-binding sequence. The Pro-Arg-Gly had a much greater affinity for Hsp47 than the previously identified (Pro-Pro-Gly)_n sequence. Fluorescent quenching-based binding assays revealed Hsp47 binding affinities for the triplets Pro-Arg-

Gly and Xaa-Arg-Gly were similar. These findings confirmed that Xaa-Arg-Gly triplets were dominant binding sites for Hsp47. Further studies showed that Hsp47 bound to (Pro-Pro-Gly)_n triplet at a temperature of 4 °C, however, significant binding was not observed at 25 °C. This demonstrated that Hsp47 binding was temperature dependant and is further evidence for the importance of triple – helical conformation of the substrates [91].

Collagen mimic peptides (PPG)₁₀, consisting of 10 repeating units of Pro-Pro-Gly, that can exist in both monomeric and trimeric states that resemble those of collagen have been used to study binding affinities of Hsp47 for its substrate. It has been demonstrated that Hsp47 binds preferentially to the triple helical form of collagen rather than the monomeric form [88]. It has also been shown that the binding of Hsp47 to the monomeric state is much weaker than the binding to triple helical collagen. Bulleid and coworkers have used newly synthesised procollagen α -chains to demonstrate that Hsp47 interacts with correctly folded, triple helical procollagen in comparison to the monomeric procollagen chains [78]. These results suggest that Hsp47 binds the triple helical form of collagen *in vivo* rather than the monomeric form. In contrast to this, P4-H interacts with the monomeric but not the triple helical form of collagen. The findings that Hsp47 binds cooperatively to triple helical type I collagen [92] and that the binding of Hsp47 to single stranded (PPG)₁₀ was non co-operative and had a low affinity, indicate that high affinity co-operative binding of Hsp47 requires the presence of triple helical procollagen.

Koide and coworkers also demonstrated that binding of Hsp47 to the Pro-Arg-Gly triplet was not disrupted by high ionic strength buffer implying that a hydrophobic interaction was mainly responsible for the Hsp47-collagen interaction. Thus it was suggested that the hydrophobic cleft of the Hsp47 directly interacted with a hydrophobic region formed on the Arg side chain [89-91].

The Importance of Hydroxylation

Hydroxylated collagen chains have been observed to have increased Hsp47 binding. This could be because binding of the chaperone to the triple helix requires hydroxylation or because triple helix formation requires hydroxylation and Hsp47-collagen binding only occurs once the triple helix has formed. Denaturation of the hydroxylated triple helix by an increase in temperature revealed that Hsp47 did not associate with denatured, single stranded collagen molecules. However, high amounts of Hsp47 were detected on reformation of the triple helix. This suggests that high affinity binding of Hsp47 depends upon the structure of the triple helix and not the extent of hydroxylation. Progressive hydroxylation of procollagen on the other hand, may cause dissociation of Hsp47. This is thought to be due to Hsp47 having a preference for proline rather than 4-hydroxyproline at the Y position [88-90]. The above results suggest that hydroxylation reduces the binding affinity of Hsp47 to procollagen and that the binding is dependant on the triple helical structure of procollagen. This is unusual as hydroxylation is required for the formation of the triple helix. This implies that Hsp47 may have a role in preventing the denaturation of the triple helix by binding to unhydroxylated regions of the triple helix and stabilising them. Consequently, dissociation of Hsp47 from the collagen triple helix follows the hydroxylation of the proline residues [88].

The Knockout Mouse

The essential role of Hsp47 in collagen biosynthesis was not proven until the establishment of Hsp47 knockout mice [93]. Hsp47 ^{-/-} mice were shown to secrete type I collagen molecules that were immature and partially processed, which resulted in embryonic death due to the formation of defective and discontinuous basement membranes. This led to the understanding that the molecular chaperone, Hsp47 was a requirement for the complete processing of type I collagen and that correctly processed and functional collagen cannot be produced without Hsp47. Subsequently, the precise role of Hsp47 in collagen biosynthesis was elucidated.

Recently, with the use of Hsp47 knockout mice, it was proved that Hsp47 is vital for the processing of type IV procollagen [94]. It was also discovered that knockout of Hsp47 instigates a delay in collagen secretion. This suggests that the ER possesses a quality control mechanism by which progression of procollagen through the secretory pathway is delayed in the absence of Hsp47. Thus, it is thought that Hsp47 facilitates the folding of procollagen so that it is in its correctly folded conformation in the ER.

Unlike most molecular chaperones that stabilise intermediates in order to favour the formation of a stable folded state, Hsp47 binds to the fully folded triple helical form of procollagen that has already reached the final stage of the folding pathway. A reason for this could be that Hsp47 is involved in preventing the thermal denaturation of the triple helix. Hsp47 may have a role in binding to and stabilising unstable regions, such as those resulting from the uneven distribution of hydroxyproline residues, during the movement of procollagen from the ER to the Golgi body [95]. This is consistent with the fact that heat shock results in the induction of Hsp47 expression and Hsp47 binds preferentially to the unhydroxylated form of triple helical procollagen [86, 95].

It has also been seen that Hsp47 inhibits the fibril formation *in vitro*, with complete inhibition requiring a 1:1 ratio of collagen:Hsp47 [96]. Hence, it may be that binding of Hsp47 to procollagen in the ER could be to prevent the formation of insoluble collagen fibres. However, this is unlikely to be the lone function of Hsp47, as it would not explain the observed secretion of immature collagen in the knockout mice.

In general, current findings and evidence support that Hsp47 is essential for the processing and folding of collagen. Although a number of roles have been proposed, a specific role is yet to be confirmed. It may be that Hsp47 has many roles, such as, aiding the formation of the

procollagen triple helix in the ER, preventing dissociation of the procollagen triple helix from the ER, precluding the unwinding of regions of the triple helix and preventing premature collagen aggregation in the ER. Further research is required to determine which, if any of the above is correct.

1.8 The pH dependant conformational switch mechanism

An unusual feature of Hsp47 is its pH-dependant interaction with procollagen/collagen. El-Thaher and coworkers have shown that the pH-dependant binding and release of procollagen/collagen coincides with the pH-driven trans-conformational changes in Hsp47 indicating that these conformational changes alter the binding affinity of the chaperone for its substrate. The use of CD and fluorescence spectra to examine changes in secondary and tertiary structure revealed three structural states. The Alkali state (above pH 6.4) consisted of 25% and 36% α -helix and β -sheet, respectively. The Intermediate state (pH 6.2) comprised of 15% α -helix and 51% β -sheet and the transition from the Alkali to the Intermediate state showed marked changes in secondary structure. The Acid state (below pH 5.7) contained 11% and 57% α -helix and β -sheet, respectively. Therefore a pronounced decrease in α -helical character and a substantial increase in β -sheet content are characteristic of the transition from the Alkali to Intermediate to Acid states [97] (summarised in Table 1.1 below).

State	α -helix character (%)	β -sheet character (%)
Alkali	25	36
Intermediate	15	51
Acid	11	57

Table 1.1. Summarizing the secondary structure composition of the Alkali, Intermediate and Acid states. The transition from the Alkali to the Acid state is associated with a decrease in α -helix and an increase in β -sheet character.

The protonation and deprotonation of histidines may be the initial trigger driving the conformational changes in Hsp47. Mature recombinant mouse Hsp47 (mrmHsp47) contains fourteen histidines and its overexpression in *E. coli* allowed El-Thaher and coworkers to use this for structural studies. It is thought that the transformation from the Alkali to the Intermediate state results from the formation of new electrostatic interactions following histidine protonation. Titration curves of CD intensities (at 220 nm and 289 nm) as a function of pH demonstrated points of inflection at approximately 6.2, the pK_a of the imidazole group of histidines [97]. They also found that the double mutant H197Q:H198N and a H335L mutant resulted in aberrant collagen binding activity. The H335L mutant remained bound to collagen even at pH 6.1 whereas the double mutant (H197Q:H198N) abolished collagen binding properties completely (Yokota and Nagata, unpublished).

The above findings, along with previous evidence of Hsp47-collagen binding above pH 6.3 but release at or below pH 6.3, is accounted for by the pronounced effect of pH on the secondary and tertiary structure of Hsp47. The three state theory explained above can be used to illustrate the correlation between the binding and release of Hsp47 from procollagen/collagen and the pH-induced structural changes observed in mrmHsp47, such that only in the Alkali state can Hsp47 bind its substrate and not in the Intermediate or Acid state. The work by El-Thaher and coworkers supports this hypothesis as protein purification was done using a collagen-sepharose column at pH 8.0. Subsequently, the protein was eluted at pH 5.8 after which it was stored at pH 8.0, restoring its collagen binding ability [97].

Dafforn and coworkers also observed this pH-dependency when the binding of Hsp47 to a collagen mimic peptide (PPG)₁₀ was reduced following a drop in pH from 7.0 to 6.0. This reduction in pH is thought to induce a conformational change, leading to a 1000-fold decrease

in the binding affinity of Hsp47 for collagen, resulting in dissociation of Hsp47 and impairment of biological activity [87].

1.9 Hsp47 as a member of the Serine Protease Inhibitor Superfamily

1.9.1 Introduction

Serine protease inhibitors (serpins) belong to a family of proteolytic enzymes whose members are the enzymes of protein digestion. The serpins are known to bind and unfold serine proteases, many of which act as triggers for a number of molecular pathways, including immune activation and fibrinolysis [98]. It is the unique hypermorphic molecular mechanism of serpin function that provides specificity and irreversibility and has rendered serpins as the major serine protease inhibitors in the majority of higher organisms [99]. Inhibition of their cognate proteases involves a striking conformational transition after which the serpin itself becomes inactive. This highly flexible nature of the structure however, makes serpins prone to disadvantageous mutations and structural changes that can cause disease [100, 101].

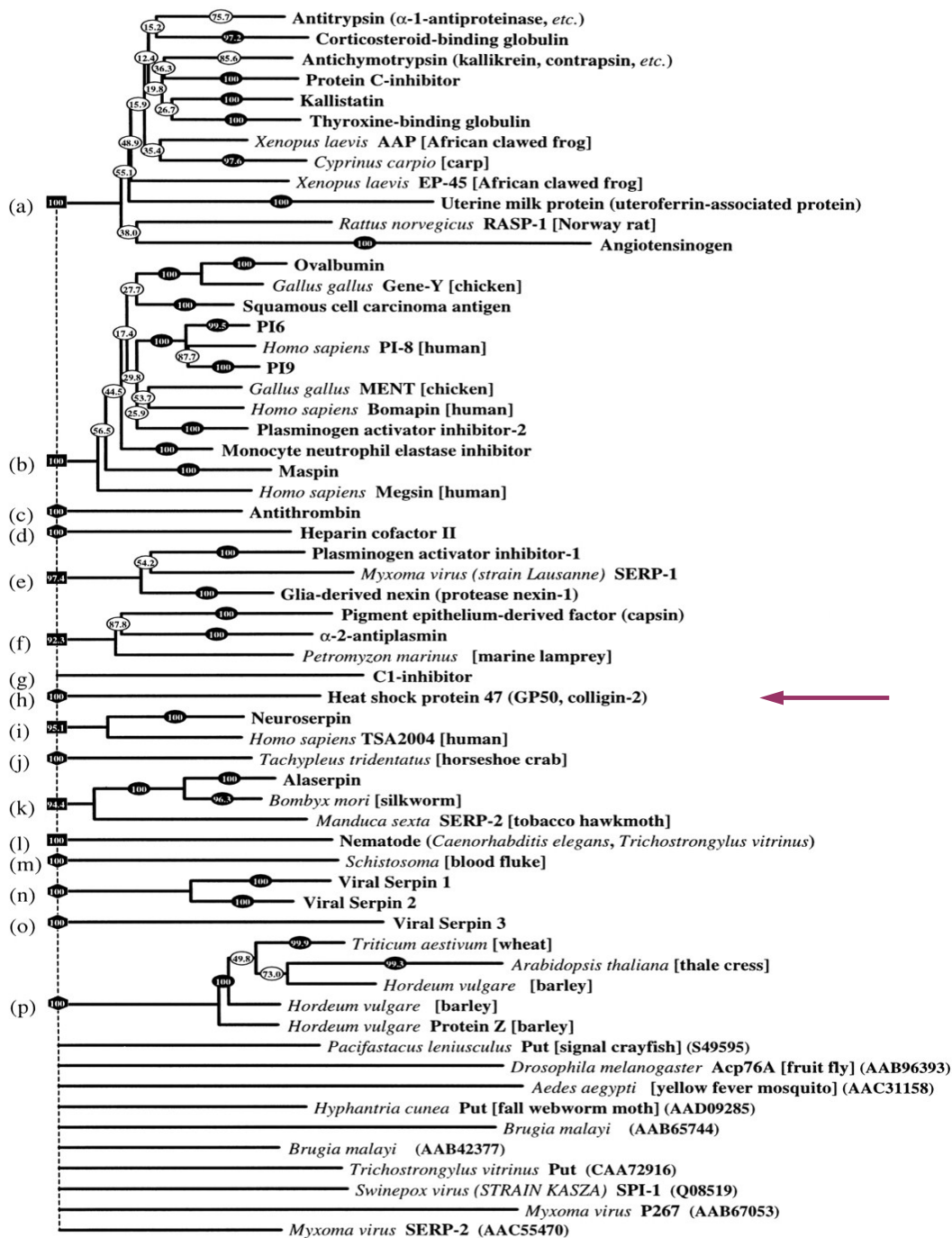
1.9.2 Identification of the Serpin Superfamily

The term ‘serpin’ superfamily first came about in 1985 as many of the members were serine protease inhibitors [102]. Previously, it had been observed that there was 30 – 50% sequence homology between α_1 -antitrypsin (α_1 -AT), chick ovalbumin and antithrombin III [103]. Currently, over eight hundred proteins in the sequence databases are assigned to the serpin superfamily with serpins being present in viruses, higher eukaryotes and prokaryotes. They are however yet to be found in yeasts [104].

Sequence comparisons have enabled the serpin superfamily to be classified into clades (Figure 1.9) and 10 orphan sequences [105], with similar sequenced serpins placed in the same clade. There are sixteen clades in total, eight of which are assigned to the extracellular serpins. An example is clade *a*, which is formed by α_1 -AT-like serpins such as ACT, thyroxine binding-globulin (TBG) and angiotensinogen (ANGT). This group of serpins has a broad range of functions, but are mainly associated with inhibition of serine proteases. However, some are non-inhibitory. These include cortisol binding globulin (CBG), TBG [106], ANGT and the uterine milk protein (UTMP). As antithrombin (clade *c*) and neuroplasmin (clade *i*) cluster close to the insect/ intracellular/ PAI-1 section of the phylogenetic tree, it implies that these groups diverged reasonably early and that intra- and extracellular serpins may be linked by antithrombin or neuroserpin [4].

Intracellular protease inhibitors such as PI-6 and PI-8 (proteinase inhibitor-6 and -8, respectively) are in clade *b* along with the other intracellular serpins or ov-serpins [104]. The intracellular serpin subfamily is the ancestor of the extracellular serpins and they function by modifying the behaviour of cells. For example, megsin, enhances maturation of megakaryocytes from bone marrow cells [107] and PAI-2 has the capability of inhibiting tumour necrosis factor- α (TNF) – induced apoptosis [108]. However, the functions of many of the serpins in this clade are yet to be identified. The one exception to this is the storage protein ovalbumin, where there is evidence of a conformational transition in ovalbumin during embryo development [109].

Hsp47 was identified as a serpin by virtue of sequence homology [99] and is another example of a non-inhibitory serpin, which has adapted a role as a collagen-specific molecular chaperone [110].



[104]

Figure 1.9. A phylogenetic tree demonstrating the relationship between the serpin superfamily. a-p represent the 16 clades; clade a-antitrypsin-like serpins, clade b-intracellular serpins, clade c-antithrombin, clade d-heparin cofactor II, clade e-PAI-1, clade f-capsin, clade g-C1 inhibitor, clade h-Hsp47, clade i-neuroserpin, clade j-horseshoe crab, clade k-insect, clade l-nematode, clade m-blood fluke, clade n-viral serpin 1/2, clade o-viral serpin 3, clade p-plants. The 10 orphans (at the bottom of

the tree) represent serpins that failed to group with a clade and include the *Aedes aegypti* factor Xa inhibitor and Acp76A (the accessory gland protein).

1.9.3 Serpin Folding

Studies on serpin protein folding have involved kinetic and equilibrium measurements. Bruch and coworkers investigated the unfolding pathways of α_1 -AT, antichymotrypsin (ACT), antithrombin and C1 inhibitor using circular dichroism (CD) [111]. A stable intermediate was observed for each of these serpins, implying that they shared similar unfolding pathways, comprising at least one stable intermediate.

The unfolding pathway of α_1 -AT was probed further by James and coworkers using intrinsic tryptophan fluorescence and cysteine labelling [112]. Of the two tryptophan residues (W194 and W238 found on the A and B β -sheets, respectively) in α_1 -AT, the majority of the intrinsic fluorescence signal is thought to be due to W194 [113]. James and coworkers observed ordered unfolding of the different regions in the protein. Early exposure of P3'Cys and C313 were indicative of early movement in the RCL and opening of β -sheet A and subsequent exposure of W194. Opening of the A β -sheet and movement of strand 1 in β -sheet C (s1C) are linked to the formation of a stable intermediate [112]. Helix F (hF) of this intermediate species has been shown to be unfolded, implying that this helix may be involved in the unfolding of α_1 -AT [114]. Subsequently, the A and C β -sheets of α_1 -AT are disrupted, prior to unfolding of β -sheet B and exposure of C232 [112]. This suggests that the B β -sheet is the most resistant region of α_1 -AT. Figure 1.13 illustrates the basic serpin fold. Similarly, the unfolding pathway of PAI-1 also exhibits a folding intermediate and β -sheets B and C are more resistant to denaturation than β -sheet A [115].

Powell and coworkers measured changes in tryptophan fluorescence lifetimes and quenching to investigate the kinetics of folding and unfolding of α_1 -AT [116]. Denaturation results were

more complex than general globular proteins, with curves demonstrating the presence of multiple equilibrium intermediates in the presence of denaturant. They found the folding process to be rapid (Figure 1.10), involving a series of intermediates (X_i) that lead to a more compact intermediate (I) [117]. Internalisation of tryptophan residues and formation of quenching interactions allowed progression from the intermediate to the native (N) state.



[116]

Figure 1.10. Proposed kinetic pathway of α_1 -AT. Folding of recombinant AT (r-AT) involves rapid collapse to a compact intermediate that folds more slowly to the native state. Rapid folding occurs through a series of intermediates (X_i) to a compact intermediate (I). Final packing of tryptophan residues forms the native (N) conformation.

This model suggests that the two-phase decrease in fluorescence is a result of consecutive reactions.

Further understanding of the folding/unfolding pathway of α_1 -AT was provided by Kim and coworkers. They used intrinsic tryptophan fluorescence and extrinsic 1-anilino-8-naphthalene sulphonate (ANS) fluorescence [118]. The α_1 -AT sequence comprises of two tryptophanyl residues; W194, which is conserved amongst the serpin protein superfamily and W238, substitution of which did not alter the kinetic parameters. Hence, the major changes in tryptophan fluorescence in the folding pathway of α_1 -AT are due to the environment surrounding W194. Refolding experiments revealed three kinetic phases. An initial fast phase that produced a small increase in fluorescence was followed by a slow decrease in fluorescence and a final third phase that was accompanied by a further slow decrease in fluorescence. A further, very slow phase (conversion from I_N to A), thought to be due to aggregation, was observed. Kim and coworkers proposed a folding pathway that was initiated

with the folding of the unfolded protein (U) to a molten globule-like state (I). Compactisation of I forms a native-like intermediate (I_N). Further internalisation of hydrophobic regions yields I_N' . I_N , the active intermediate, undergoes kinetic competition and can proceed to either the native protein (N) or aggregation (A) (Figure 1.11).

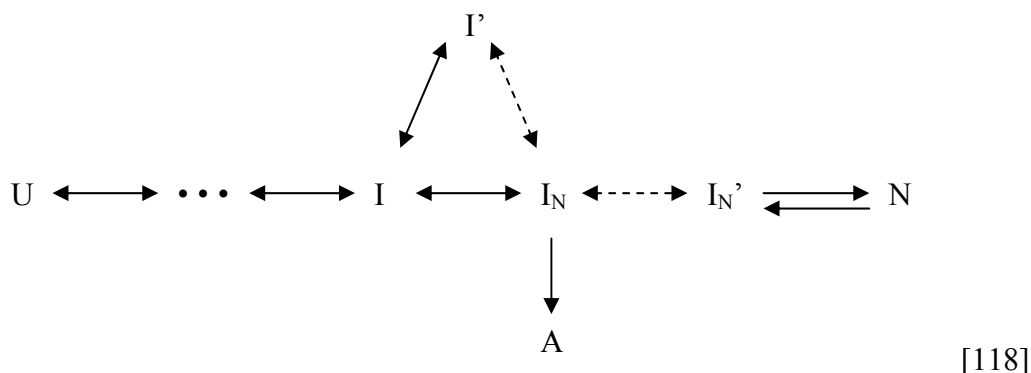
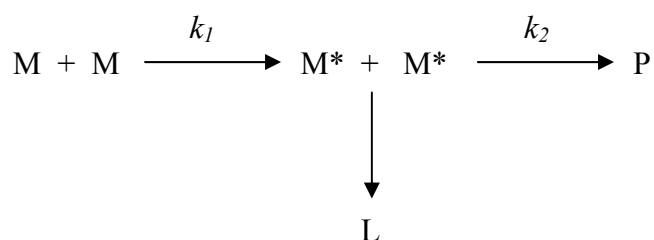


Figure 1.11. A proposed folding pathway of α_1 -AT. The first step involves conversion of the unfolded protein (U) to a molten globule-like state (I). This fast, folding phase is limited by *cis-trans* isomerisation. Folding of the molten globule-like state forms a native-like intermediate (I_N), either directly or via a more compact intermediate (I'). Further internalisation of hydrophobic residues forms I_N' . At high protein concentrations, I_N is aggregation prone and can result in aggregate formation (A). Trp 194 is located at the top of s3A, internalisation of which may be the final folding step in the closure of the A β -sheet, yielding native state (N) formation [118].

Dafforn and coworkers investigated in detail, the process by which polymer formation occurs in the α_1 -AT folding pathway, using intrinsic tryptophan fluorescence and extrinsic probe, 8-anilino-1-naphthalene sulphonic acid (ANS) [119]. In agreement with Kim and coworkers, polymerisation of α_1 -AT was found to require the formation of an intermediate species [119]. Dafforn and coworkers proposed that destabilisation of the native state, accompanied by a conformational change, results in the formation of a polymerogenic monomeric state (M^*) which can lead to polymer (P) formation. The polymerisation process was thought to be a two-step process (Figure 1.12) that was dependent upon the concentration of M^* and

therefore k_1 . At a high concentration of protein, the process was saturated thus non-specific aggregation was thought to compete with polymerisation.



[119]

Figure 1.12. A proposed model for the polymerisation of α_1 -AT. A conformational change in the α_1 -AT monomer (M) produces a polymerogenic intermediate (M*). Self-association of M* yields polymer (P). Progression of M* via a side pathway leads to the latent (L) conformation. Polymer formation is thought to be a two-step process; therefore the conversion from M to P (k_2) is dependent on the concentration of M* and hence the conversion of M to M* (k_1) [119].

1.9.4 Serpin Structure

The serpin superfamily consists of single domain proteins that are usually 350-400 amino acids in length. They have a highly conserved secondary structure which comprises of nine α -helices (A-I) surrounding a core of three β -sheets (A-C) and a 26-34 amino acid reactive centre loop (RCL) [120], which acts as bait for target proteases as it contains the protease recognition site (Figure 1.13). The central β -sheet is also known as the A-sheet and can exist in two forms. The first is the metastable, 5-stranded form that results from normal folding. The second is the 6-stranded hyperstable form, which is the result of proteolytic cleavage or part denaturation of the protein. Changes in conformation of the RCL, β -sheet A and β -sheet C permit serpins to undergo major structural rearrangements [121].

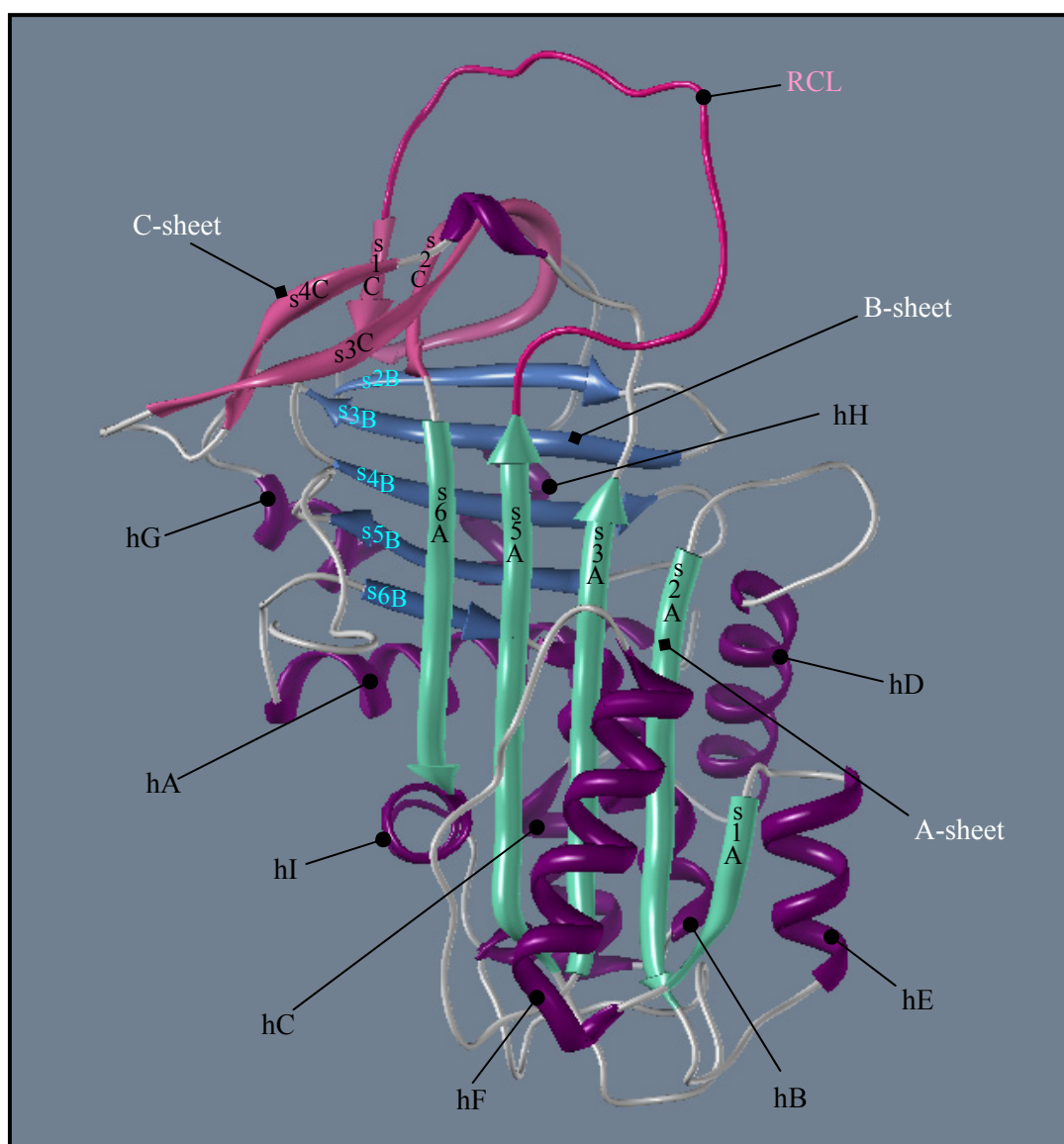


Figure 1.13. The basic serpin fold. Illustrating the positioning of the α -helices (in purple), β -sheets (β -sheet A comprises the green β strands, β -sheet B the blue β strands and β -sheet C the pink β strands) and the RCL (pink).

Table 1.2 below shows the serpins used to model the mHsp47 structure.

PDB Identifier	Sequence Identity to mHsp47	Sequence Homology to mHsp47	Resolution of structure	RMSD from Hsp47 model (C-alpha)
1by7	28 %	51 %	2.00 Å	1.36 Å
1ova	26 %	47 %	1.95 Å	1.45 Å
1qlp	28 %	49 %	2.00 Å	1.67 Å

Table 1.2. Summarising the statistical parameters of the model structure of mHsp47. The model is based on PAI-2 (1by7), ovalbumin (1ova) and α_1 -AT (1qlp). The sequence identities (over 20 %) for each of the named serpins with mHsp47 indicate a good correlation.

In addition to motion of the RCL, the conformational transition from the stressed to relaxed serpin state is associated with rearrangement of rigid segments within the serpin [101, 122, 123]. These include the breach, shutter and distal clusters (Figure 1.14).

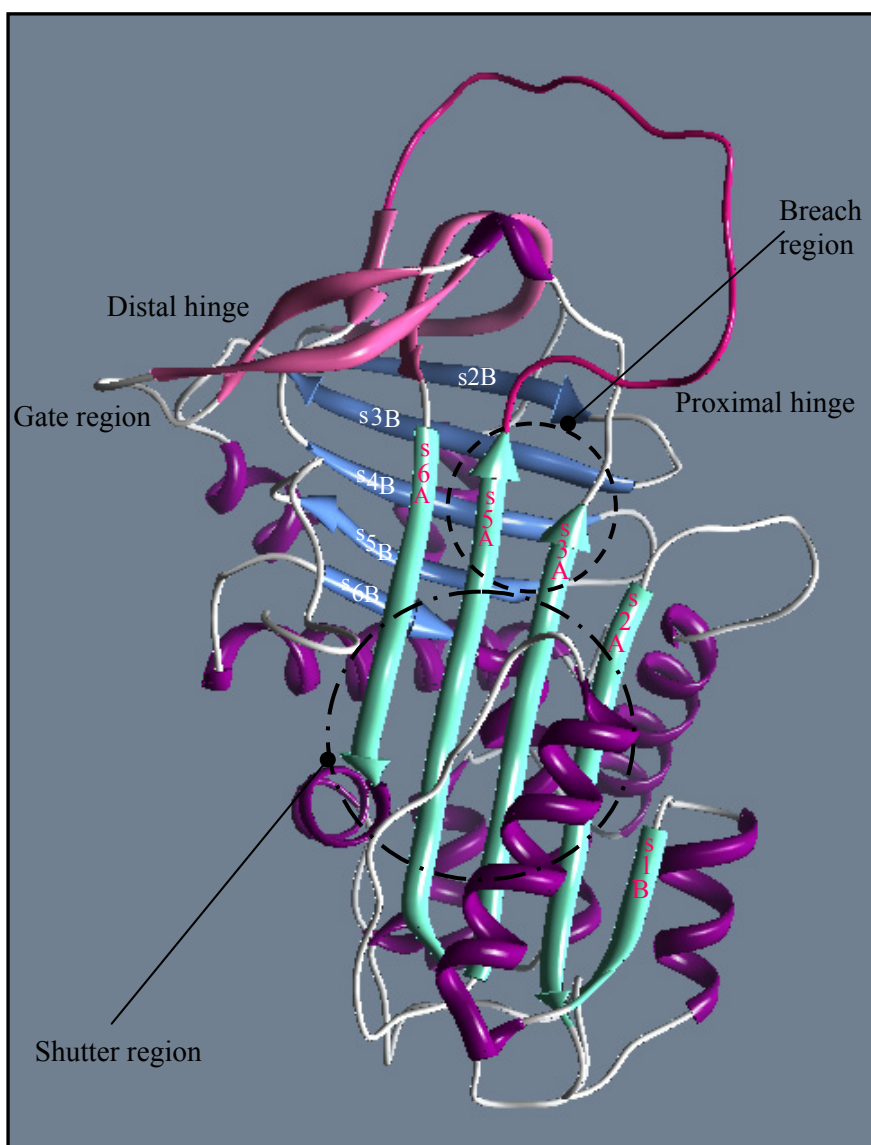


Figure 1.14. The domains of a serpin shown on a model structure of native Hsp47. The helices are shown in purple, the green strands make up β -sheet A, blue strands form the B β -sheet and pink strands represent β -sheet C.

1.9.4.1 The Breach Region

This domain consists of the top of strand 3 of β -sheet A (s3A), s5A, s2B, s3B and s4B and is located at the top of the A-sheet [101]. It is the region where the RCL initially inserts into the

serpin during the inhibitory process. Conservation of the residues within this region provides an indication of its importance in serpin structure [124] and mutations at several of these positions cause hyperstability [125]. Point mutations that perturb the conformation of the A-sheet render it prone to aberrant loop insertions and are the most common cause of serinopathies. For example, in the Z-mutant of α_1 -AT, replacement of the negatively charged Glu342 by a positively charged Lys residue disrupts the salt bridge with Lys290. This salt bridge is normally responsible for the narrow opening of the top of the cleft of the A-sheet. Replacement of the Glu by the Lys results in a charge repulsion between the two Lys residues which instigates widening of the breach region, leaving it susceptible to insertion of an RCL from a second molecule. The result is polymerisation and can cause diseases such as liver cirrhosis and emphysema [126]. Another example is a thermolabile variant of antithrombin in which the highly conserved Phe229, a residue deeply buried in the breach cluster, is replaced by a Leu residue [127]. As the side chain of Leu is more branched than that of Phe, a small internal cavity is formed that enforces a steric clash between residues neighbouring the hinge region of the RCL. Such deviations initiate conformational changes that destabilise the variant, which then has a tendency for spontaneous polymerisation at increased body temperature [127]. As the breach region is at the position of RCL insertion, it is thought to be responsible for the regulation of insertion of the polypeptide loop into the A-sheet. Thus, it may determine whether the serpin acts as an inhibitor or a substrate [122].

1.9.4.2 The Shutter Region

The shutter domain consists of part of s6B, the N-terminal part of helix B and the central parts of s5A and s3A. The fragments are located at the core of the serpin and interconnected by a cluster of hydrogen bonds. A shutter-like structure results and is responsible for controlling the opening of the A-sheet [101], so that aberrant loop insertion into the lower part of the cleft is prevented. Protease inhibition is accompanied by the formation of a novel set of hydrogen

bonds along the upper half of the cleft, breaking the hydrogen bridge between s3A and s5A. This instigates a conformational transition in the entire shutter region, including the translocation of helices B and F, which in turn forms an opening for the incoming loop (Figures 1.13 and 1.14). In native α_1 -AT, the shutter region consists of a triad of polar residues, Asn186, His334 and Ser56. In cleaved α_1 -AT, there is a shift in the side chain of Asn186 and it forms a new hydrogen bond to Ser56 [101]. The shutter domain, like the breach, is vital for controlling the conformational change in serpin structure; therefore 85% of residues in the shutter domain are conserved in known serpin structures. The naturally occurring α_1 -AT S_{iiyama} (Ser53Phe) and α_1 -AT M_{malton} (Phe52 deletion) mutants are in this region and lead to polymerisation and disease [128, 129].

1.9.4.3 The Hinge Domains

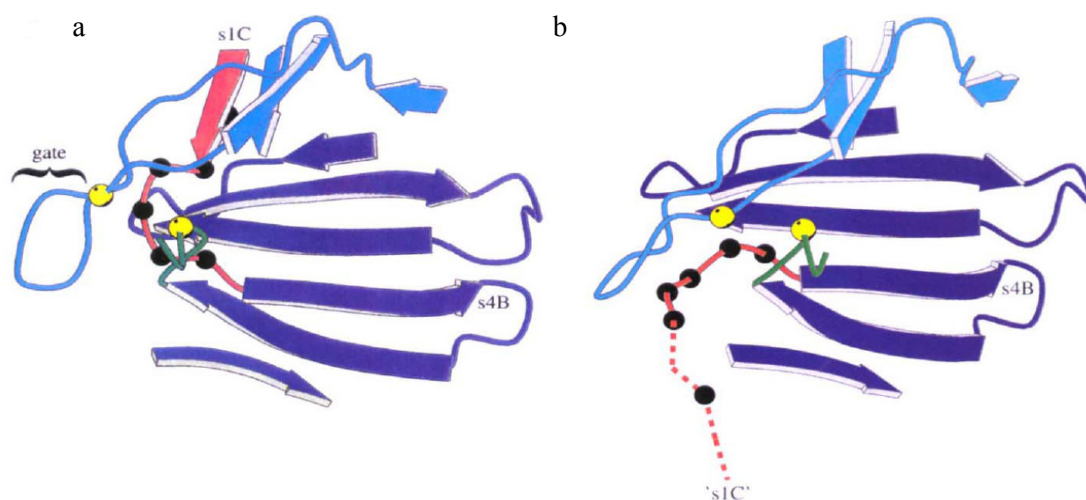
The Proximal Hinge

The proximal hinge domain comprises of residues P8-P15 on the N-terminal of the RCL. Studying variants with mutations in this region provided evidence for the necessity of the mobility of the RCL for inhibitory activity [123, 130]. Antithrombin variants, Cambridge-I (mutation of P10 Ala to Pro) [131], Cambridge-II (mutation of P14 Ala to Ser) [130] and Hamilton (mutation of P12 Ala to Thr) [132] have been characterised. Despite having a normal association rate with proteases, the variant inhibitor-protease complexes are less stable than those of their WT counterparts [133]. It is the locking of the serpin and protease, rather than their docking that appears to be affected by mutations in the proximal hinge region [134]. The majority of variants remain susceptible to cleavage at P1-P1' and become protease substrate, the outcome of which is loss of inhibitory activity [123]. Great variability has been seen in the characteristics of individual examples of variants with mutations in the proximal hinge. The consequence of a mutation at the P14 position (Val432Glu) in C1-inhibitor is conversion from an inhibitor to a substrate, which results in hereditary angioneurotic oedema

[135]. By comparison, a mutation at P10 from alanine to threonine results in a complete loss of inhibitory activity, causing the formation of polymers in the plasma [136]. Furthermore, mutation of the same residue in antithrombin causes only a minor loss of inhibitory activity [130]. Residues P10, P12 and P14 are proposed to be the most critical positions within the proximal hinge domain as mutation at other residues do not appear to affect the inhibitory characteristics of the serpin [137]. Moreover, the mutation of residues P10, P12 and P14 to threonine in leukocyte elastase inhibitor led to a loss of inhibitory activity [138].

The Distal Hinge Domain

The distal hinge region is formed by s1C and the turn connecting it to s4B [123]. The transition from the native to an alternative conformation is accompanied by a rearrangement of the distal hinge [139, 140]. This is supported by studies of antithrombin, which revealed extensive movement in this domain as a consequence of transition from the active to latent conformation [123] (Figure 1.15).



[123]

Figure 1.15. The native and latent forms of antithrombin. The distal hinge region (pink) adopts a different conformation, depending upon the conformation of the RCL. Such that, in active antithrombin (a), strand s1C is ordered, whereas in the latent form of the protein (b) this strand is

disordered. s1C is pink, the remainder of the C-sheet is light blue, sheet B is dark blue and the C-terminus is coloured green. The black spheres represent mutations that result in loss of inhibitory activity of the serpin. The yellow spheres are C_α atoms of cysteine residues that form a disulphide bond between the gate region and the C-terminus [123].

A loop, also described as a 'gate', formed by s3C and s4C and the turn connecting these strands has been shown to hinder the movement of s1C [139]. In PAI-1, this 'gate' is flexible, permitting spontaneous conversion to the latent form, however, in antithrombin; it is constrained by a disulphide bridge (Cys 247-Cys 430). Mutation of Cys 430 results in a loss of inhibitory activity of antithrombin. Thus, the hinge is considered as being the s3C and s4C gate and the C-terminal sequence that fixes its position, as well as the mobile hinge formed by s1C and the turn connecting it to s4B [123]. Stability of the distal hinge domain plays an essential role in protease inhibition as variants of antithrombin, C₁-inhibitor and α_1 -AT with mutations at s1C, the turn connecting it to s4B and the C-terminal sequence are dysfunctional [141].

Consequently, there appears to be a role for each of the domains. The breach is the initial point of insertion of the RCL and it consists of many conserved residues, mutations of which result in hyperstability of the molecule [125]. Therefore, the breach domain is thought of as being the first line of defence against unwanted conformational changes. The transition from the stressed to the relaxed state is fundamental for protease inhibition, and the breach, together with the shutter domain, is responsible for regulating the associated conformational changes [101]. The shutter region is known to control the rest of the serpin molecule as it contains critical hinge regions between the major fragments. Hence, it is essential that this domain is accurately packed [101]. The proximal and distal hinge domains are associated with conformational transitions and are once again vital for inhibitory function as mutations in this region lead to a defect in inhibitory activity [123, 141].

1.9.5 Serpin Function

The highly flexible RCL is the central component for the activity of an inhibitory serpin [142] (Hsp47 however, is not thought to be inhibitory). It acts as a pseudosubstrate that is specifically recognised and cleaved by the serine proteases. The interaction of the flexible loop region with active site of the proteases allows serpins to form covalent complexes with their corresponding serine proteases. The release of the serpin from the complex requires cleavage of its polypeptide chain by the protease in the middle of the serpin loop region, degrading the serpin molecule. Following cleavage, the RCL inserts itself as a β strand into the A-sheet. The result is conversion of the A-sheet from a parallel, metastable conformation to an anti-parallel, hyperstable form [99]. This zipping of the flexible loop into a groove on the side of the serpin results in the translocation of the protease from one side of the molecule to the other, over a distance of 70 Å. This pulls the active site of the protease out of shape, disrupting approximately 40% of the protease structure, ultimately resulting in removal of the conserved serine residue from the catalytic triad. The protease remains covalently trapped on the serpin as an acyl enzyme. Complete irreversible inhibition is the consequence of the loss of catalytic activity of the serine protease. This mechanism has led to serpins being referred to as ‘molecular mousetraps’ [99, 143].

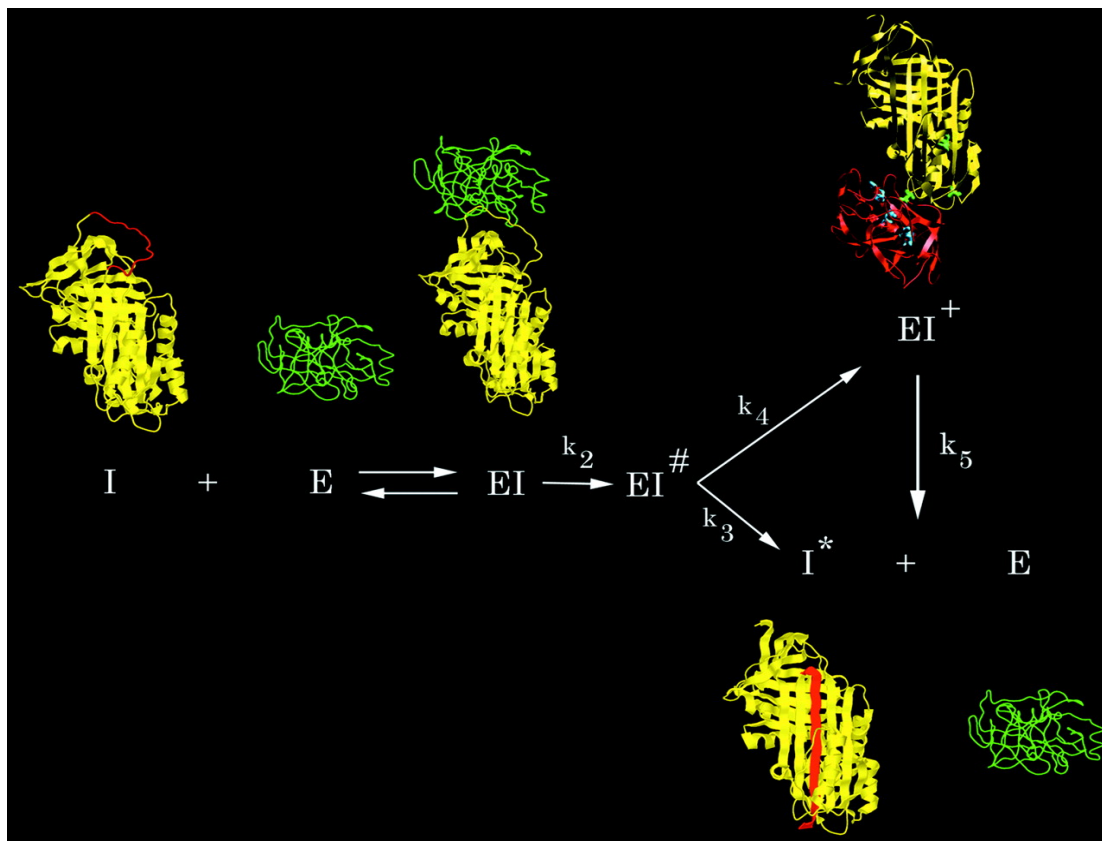
The Serpin Mechanism of Action

Serpins use an irreversible suicide mechanism to inhibit serine proteases (Figure 1.16) [143]. The kinetic trapping involves acylation of the serine hydroxyl at the active site of the protease, followed by deacylation and ultimately complete proteolysis of the serpin, leading to dramatic conformational transitions [144]. The branched pathway of inhibition is initiated by the formation of a noncovalent Michaelis-like docking complex (EI) by the protease interacting with residues flanking the reactive site P1-P1' (scissile) bond [143]. Both the highly flexible nature and conformation of the RCL are retained upon binding of the loop to

the protease [144, 145]. The covalent bond is formed between the serine-195 at the active site of the protease and the P1 residue of the serpin [144, 146]. The serpin is then cleaved between the P1 and P1' residues at the scissile bond and a covalent acyl-enzyme complex ($EI^\#$) is formed [146]. Following removal of the restraint, the RCL starts to insert into the A-sheet, transporting the bound protease with it. Translocation of the protease by 70 Å following complete loop insertion distorts its active site resulting in removal of the P1 side chain from the S1 pocket. The precise length of the RCL insertion results in compression of the protease against the base of the serpin which instigates protease distortion and hence inactivation [121]. Due to the dramatic conformational change caused by insertion of the loop at strand 4 in the A-sheet, the protease finds itself at the opposite end of the molecule [147]. This results in the formation of a loop-inserted covalent complex (EI^+) [121]. The dramatic conformational change was first observed by Loebermann and coworkers for α_1 -AT. They interpreted electron density maps and found residues adjacent to the cleaved peptide bond to be approximately 64 Å apart [148]. The formation of the (EI^+) complex begets a 37% structural loss in the protease [149]. Huntington and coworkers solved the crystal structure of the PI-2-trypsin complex and found the structure of trypsin to be distorted by approximately 40% [121]. Distortion of the protease active site prevents deacylation and the protease is released from the complex, leaving cleaved, inactive serpin (I^*) and free, active protease (E) [150]. Furthermore, as the trapping of the protease is kinetic, the serpin-protease complex eventually breaks down. The stability of the serpin-protease complex determines the rate of breakdown once the acyl-enzyme intermediate has formed. Thus, the half-lives of the covalent complexes vary from minutes to months [151].

The rates of RCL insertion and deacylation determine whether the serpin will act as an inhibitor or a substrate. If deacylation is faster than insertion the serpin will act as a substrate and so will dissociate from the protease once it has been cleaved. Conversely, if insertion of

the RCL is faster than the rate of deacylation, the serpin will act as an inhibitor and form a covalent complex with the protease [122]. Thus, the ratio of (EI^+) : $(I^*) + (E)$ depends upon the rate of hydrolysis (k_3) in relation to loop insertion (k_4) (Figure 1.14) [121].



[121]

Figure 1.16. The branched pathway, suicide substrate inhibition mechanism of a serpin. A peptide in the exposed RCL of the serpin (I) is recognised by the protease (E) and an initial non-covalent Michaelis complex (EI) is formed. Progression through the pathway involves conversion of the initial non-covalent complex into a covalent acyl enzyme intermediate ($EI^\#$). Cleavage of the scissile bond allows the insertion of the RCL into the A-sheet which involves concomitant protease translocation because of the ester linkage between the serpin and its cognate protease. This forms a covalently trapped serpin-protease complex (EI^+). Dissociation of the serpin-protease pair yields cleaved, inactive serpin (I^*) and active protease (E). Rate constants, k_3 and k_4 represent the rate of hydrolysis and insertion, respectively and determine the stoichiometry of inhibition.

1.9.6 Serpin conformations

The ability to undergo dramatic conformational changes due to high flexibility enables serpins to adopt a variety of conformations (Figure 1.17) and perform their inhibitory functions.

It will be seen in the results section that understanding these transitions in Hsp47 are important in understanding the activity of mutant Hsp47. In the past this part of Hsp47 function has been ignored.

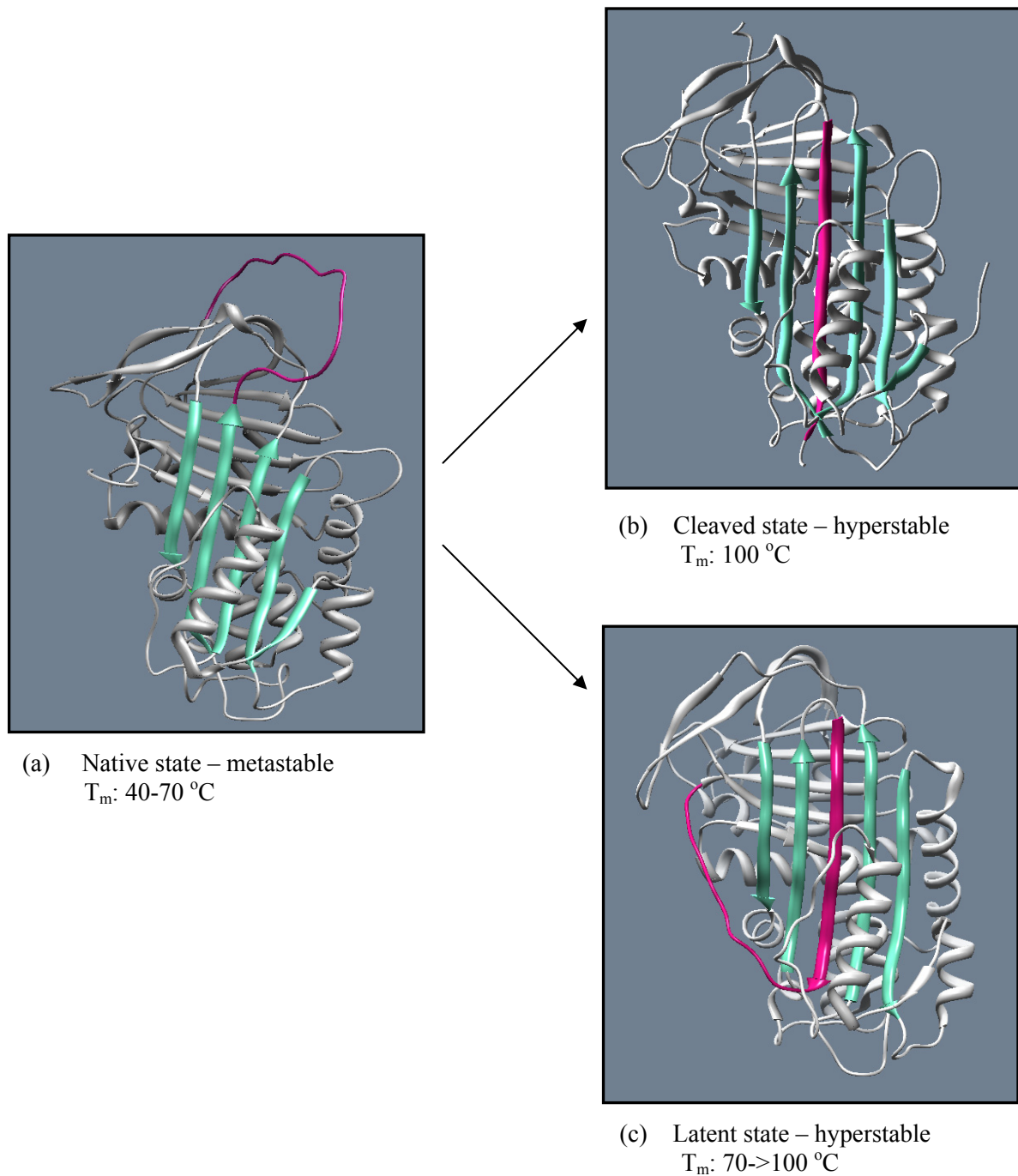


Figure 1.17. Possible conformations of a serpin. (a) Model structure of native Hsp47. The native state is the biologically active conformation and one in which the serpin is least stable (melting between 40 and 70 °C). (b) Cleaved α_1 -AT structure. The cleaved state results from insertion of the RCL into the A-sheet following cleavage by the protease. This increases the stability of the serpin which has a melting point of approximately 100 °C. (c) Latent PAI-1 structure. The latent state results from insertion of the RCL into the A-sheet without cleavage. Insertion of the RCL enhances the stability of the latent conformation, which has a melting temperature between 70 and 100 °C.

1.9.6.1 The Native Conformation

The native state has been defined as the biologically active conformation [120] or stressed serpin state. The transition from the native to the cleaved or latent (relaxed) forms is required for protease inhibition to occur. This change in transition is due to the metastability of the native serpin fold resulting in it possessing thermodynamic characteristics that allow the conversion from the native state, meaning that the active (native) form of the protein is not the most stable [21]. This contradicts a fundamental rule of folding, whereby the active form of the protein is the most energetically favourable [7]. As it consists of only 5 β -strands in the A-sheet, it lacks the full complement of strands, reducing the structural stability of this conformation. This state shows characteristics of a mesostable protein, which is denatured between temperatures of 40 and 70 °C [152].

1.9.6.2 The Cleaved Conformation

The cleaved conformation is an inactive state. Adoption of the cleaved conformation involves insertion of the RCL (residues N-terminal to the cleaved bond) as strand 4 into the A-sheet, resulting in a 6-stranded, hyperstable A-sheet. This requires movement of the first three strands of the A-sheet and an F helix [120, 153]. Bruch and coworkers used far UV-CD (circular dichroism) and near UV-CD to demonstrate that the rearrangement is associated with a change in secondary structure [111]. Cleavage of the RCL drives the transition from the stressed state to the relaxed state [101, 102]. The cleaved form can arise via two methods; the first is by breakdown of the serpin-protease complex following trapping of the cognate protease and the second is following non-specific cleavage by non-cognate proteases. The ability to develop into the cleaved state is a characteristic feature of inhibitory but not non-inhibitory serpins [154, 155]. The non-inhibitory serpin ovalbumin, for example, does not show a conformational transition or an increase in stability upon cleavage [111]. Cleavage of the RCL in inhibitory serpins is irreversible and results in a loss of inhibitory function but

greatly increases the conformational stability of the serpin in this state [156]. The cleaved conformation is hyperthermophilic as it does not denature at temperatures below 100 °C as well as being more resistant to chemical denaturation with protein unfolding requiring more than 6 M guanidinium HCl [152].

1.9.6.3 The Latent Conformation

As with the cleaved conformation, this conformation is inactive. Formation of the uncleaved, latent state involves conversion of the high energy, stressed state into an alternative lower energy, relaxed state, whereby the intact RCL is inserted into the A-sheet without cleavage [120]. In contrast to the cleaved conformation, in addition to the movement of the strands in the A-sheet and an α -helix, formation of the latent conformation requires the movement of s1C. The latent state was first discovered for plasminogen activator inhibitor-1 (PAI-1) which takes up this conformation spontaneously [139]. The intact RCL of PAI-1 inserts itself into the strand 4 position of the A-sheet but only up to residue P3, however. Complete RCL insertion requires movement of s1C away from the rest of β -sheet C, resulting in this β -sheet consisting of only 3 strands [157]. Antithrombin is another example of a serpin which spontaneously adopts the latent conformation [158]. The conformational stability of this state is intermediate between that of the native and cleaved conformations. It is more stable than the native but less stable than the cleaved conformer. Latent PAI-1, for example, denatures at temperatures of approximately 70 °C (although it can be as high as 90 °C) whereas the native form has a melting temperature of approximately 50 °C and the cleaved conformer denatures at temperatures in excess of 100 °C [152]. The spontaneity of adoption of the latent conformation by PAI-1 is thought to be part of a control mechanism by the serpin so that it circulates in the inactive form, only becoming active upon interaction with vitronectin [159].

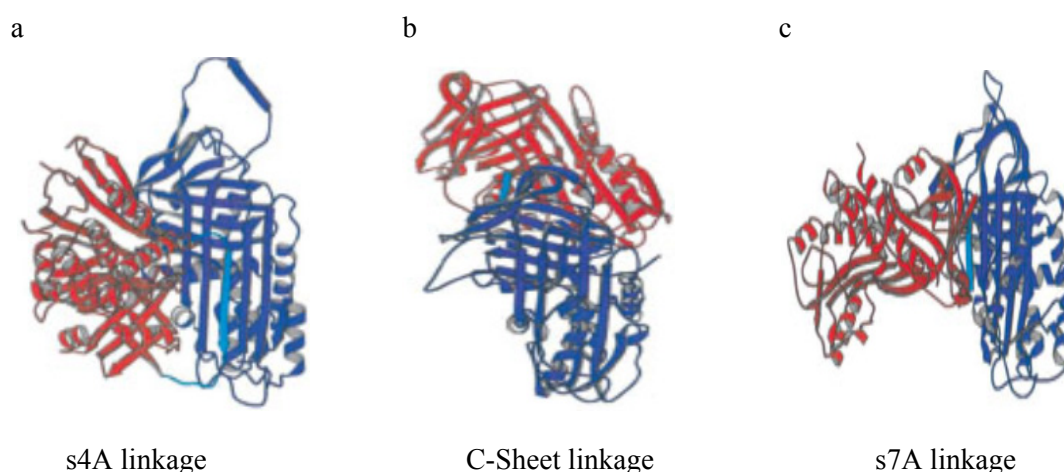
Studies on the latent form of antithrombin (L-antithrombin) have implied a two-step process; a rapid initial transition whereby the RCL is partially inserted into the A-sheet and s1C is displaced from β -sheet C, and a slower second phase involving complete loop insertion [160]. This proposal was due to the effect of citrate on latent state formation. At 60 °C, in the absence of citrate, antithrombin (like other serpins) rapidly polymerises. The presence of citrate however, inhibits polymer formation, resulting in the slow formation of L-antithrombin. As the formation of L-antithrombin was slow and increased with time, it implied that preventing polymer formation allowed the time required for the slow conversion to the latent state.

The prevention of the spontaneous conversion from the native to latent conformation is thought possible by two means. Firstly, the interactions of the A-sheet may prevent loop insertion and secondly, the interactions in the C sheet may prevent the insertion of s1C into the A-sheet [161]. Extension of the RCL by a flexible linker, (Gly-Gly-Ser)_n, resulted in loop insertion into the central β -sheet, thereby demonstrating that interactions of the A-sheet alone do not provide enough strength to prevent insertion of the reactive centre loop [162]. It was therefore proposed that interactions of the B/C barrel are critical for the prevention of loop insertion into the A-sheet. A number of mutations were made to destabilise native interactions of α_1 -AT. Substitutions such as V364A and F366A of s1C promoted conversion into the stable, latent conformation. It is therefore likely that conversion of the native into the latent state is obstructed by holding s1C in the barrel [161].

1.9.6.4 The Polymers

Serpin polymerisation occurs when the RCL of one serpin interacts with a β -sheet of another serpin. The initial insertion of the loop into the β -sheet of a second serpin molecule results in dimerisation. Repetition of this results in oligomerisation into polymer conformations [87].

Most of the polymers have greater conformational stability compared to that of the native form [162]. Polymer types identified thus far include the s4A linked polymer (Figure 1.18a), the C-sheet linked polymer (Figure 1.18b), the s7A linked polymer (Figure 1.18c) and the cleaved RCL linked polymer as well as polymers resulting from inter- and intramolecular disulphide bonds [163-168]. The resulting polymers possess pathogenicity and are associated with a range of conformational diseases. In such situations, mutations in the serpin variant cause a defect in movement of the RCL and the opening of the A-sheet. The result is polymer formation. Mutants of α_1 -AT, neuroserpin and α_1 -antichymotrypsin polymerise and have been linked to diseases such as liver cirrhosis, emphysema and dementia [87, 123, 169].



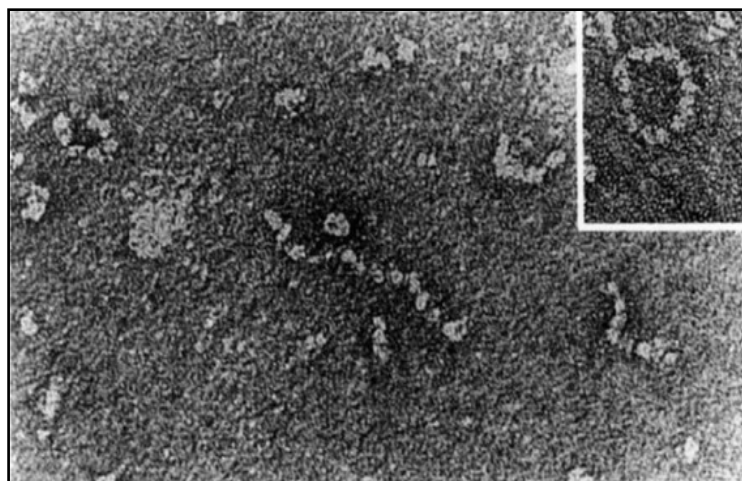
[170]

Figure 1.18. Possible polymer conformations of the serpin superfamily. (a) A s4A linkage resulting from the insertion of the RCL of one serpin into the strand 4 position of the A-sheet of a second serpin. (b) The C sheet linkage is formed when the RCL of one serpin molecule forms a β -strand in the C-sheet of a second. (c) The s7A linkage, which results from insertion of the loop into the strand 7 position of the A-sheet of an adjacent serpin.

The s4A Linked Polymer

As suggested by the name, these polymers arise following self-association of the RCL of a serpin into the strand 4 position of a second serpin molecule (Figure 1.18a) [126]. The

accumulation of the disease causing Z-variant of α_1 -AT (Glu342Lys) in hepatocytes was studied by Lomas and coworkers. They discovered that the heating of monomeric α_1 -AT resulted in the formation of non-covalently linked polymers. The resulting polymers, when visualised by electron microscopy had a 'beads-on-a-string' appearance [126, 171] (Figure 1.19). Consequently, it was believed that this insertion is the cause of accumulation of Z variant α_1 -AT in the liver [126]. Other rare mutations such as M_{malton} (F52 deleted) and S_{iiyama} (S53F), have been described to follow a similar pattern of events that ultimately lead to inclusions in the liver [172]. The affected residues were proposed to be vital for the opening of the A-sheet, following cleavage of the RCL [173]. Lomas and coworkers therefore hypothesised that the above variants would also undergo spontaneous loop-sheet polymerisation [126]. Electron microscopy confirmed the presence of polymers and moreover, these polymers had the distinct 'beads-on-a-string' morphology with occasional self-terminating circular necklaces [174]. Furthermore, Sivasothy and coworkers have used Fluorescent Resonance Energy Transfer (FRET) to demonstrate the distances between monomeric units in polymeric α_1 -AT and confirmed that RCL-s4A linkages were responsible for polymer formation [175]. The s4A linkage is the polymeric conformation that is most important in serpin related disorders. The stability of this oligomer is similar to that of the latent and cleaved conformers [152].



[171]

Figure 1.19. Electron micrograph of the antitrypsin S_{iiyama} variant. The appearance was identical to that of Z antitrypsin with the presence of circlelets or ‘beads-on-a-string’ morphology [171].

Recently, Huntington and coworkers investigated the molecular basis of serpin polymerisation by protein crystallography [174]. They used the previously reported self-terminating antithrombin dimer as a prototype [176] and observed an extensive domain swap that involved the insertion of strands 4 and 5 of the central β -sheet of one monomer into β -sheet A of another monomer (Figure 1.20). A similar domain-swapping observation for the prototypical serpin α_1 -AT implied that this intermolecular linkage was not only confined to antithrombin but could be applied to serpins in general [174].



Figure 1.20. A model of the linear serpin dimer. RCL is coloured yellow. Balls indicate AspN cleavage sites found in antithrombin (magenta) and α_1 -AT (cyan) polymers. The antithrombin dimer was linearised by breaking of intermolecular contacts in order to produce a dimer capable of further polymerisation. Cleavage by the metalloprotease AspN allowed the unravelling of helix I that was required to create a flexible between folded regions. The yellow ball represents the Leu303Cys substitution used to further investigate the effect of unravelling helix I on polymerisation of α_1 -AT [174].

The C-Sheet Linked Polymer

C-sheet polymerisation involves a ‘head-to head’ association of two serpins whereby the RCL of one forms a β -strand in the C-sheet of the second [152]. The RCL of one protein inserts as

strand 1 into β -sheet C of the second and requires over-insertion of the loop with s1C subsequently peeling away from the C-sheet [177]. C-sheet polymers have been observed for AT where the native and latent states of the protein resulted in dimer formation [158]. Furthermore, it is believed that in addition to forming A-sheet polymers, α_1 -AT can also form C-sheet polymers, depending upon the buffer conditions [164]. The weak interactions in this polymer type, namely hydrogen bonds between the 2 β -strands, dissociation of C-sheet dimers can take place using chemical denaturants [152]. This distinguishes between the s4A and C-sheet polymers, which are stable to high temperatures and concentrations of denaturants.

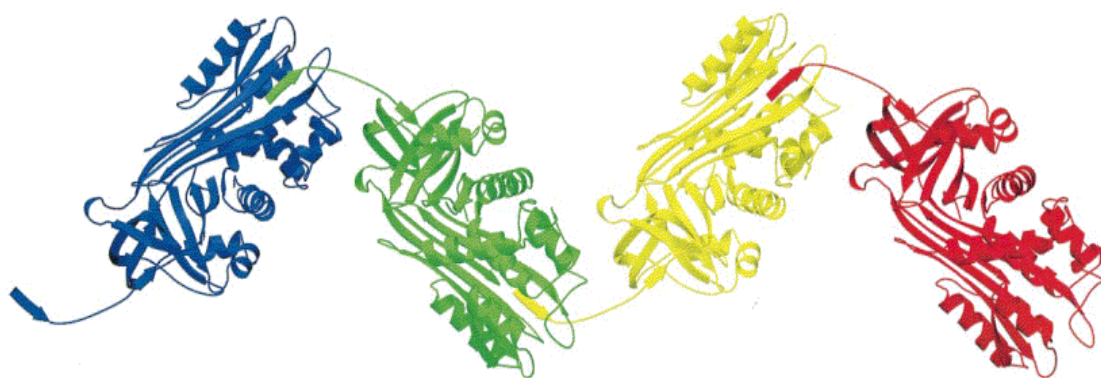
The s7A Linked Polymer

Formation of s7A polymers involves insertion of the RCL of one serpin at the end of the A-sheet, as strand 7 of another (Figure 1.18c). This conformation was first observed for native PAI-1 when its crystals revealed polymer contacts in the structure [178]. Acid-induced polymerisation studies of PAI-1 indicate that these polymers may be present in solution [166]. As of yet, PAI-1 is the only serpin known to adopt this conformation.

The Cleaved Polymer

Formation of the RCL-cleaved serpin polymer involves cleavage of the serpin at the P7-P6 position of the RCL, which produces a free peptide end that inserts into the A-sheet at the strand 4 position [165] (Figure 1.21) . Self-insertion of residues P14-P7 of the cleaved RCL precedes insertion of P6-P3 into another molecule that is partially occupied by P14-P7 of its own RCL. Cleaved polymers were first identified by Mast and coworkers using electron microscopy and were thought to have a ‘beads-on-a-string’ type appearance [126, 179]. As with the cleaved monomeric conformation, the cleaved polymer is hyperstable, thereby having similar thermal denaturation characteristics [179]. Differentiation of the cleaved

polymer from the other polymers can be done using SDS-PAGE where it is the only one that migrates to the same position as the cleaved, monomeric form of the protein [152]. Analysis of native PAGE and electron microscopy results for cleaved polymers revealed that they were structurally similar to serpin polymers associated with disease [165].



[165]

Figure 1.21. A model of the polymer conformation of cleaved α_1 -AT. The polymer is formed by the repeated insertion on residues P6-P3 of the RCL inserting into the A-sheet of a neighbouring α_1 -AT molecule.

1.10 The Location of Key Histidines in Hsp47

As described previously (section 1.8), the reversible pH dependent conformational changes in Hsp47 are driven and regulated by histidine residues in the serpin structure [87, 97, 180]. It is proposed that a change in the charge of at least one of these histidines is responsible for triggering the conformational change as the reduced binding affinity of Hsp47 for its substrate occurs between pH 6.0 and 7.0, close to the pK_a of histidine [97]. Dafforn and coworkers used the highly conserved serpin scaffold to construct a molecular model of mouse Hsp47 [87]. Upon examination, three histidine clusters were identified (Figure 1.22), each one being located in a serpin domain known to be involved in the transition from the metastable, stressed state to the relaxed state [120]. The breach cluster comprises of H197 and H198

(s3A-s4C) and H220 (s3C) and is found below the hinge region at the top of the A-sheet. As mentioned earlier, histidines 197 and 198 were probed by El-Thaher and coworkers [97], however they focused on the conformational transition of Hsp47 in relation to collagen binding and did not study the effect of the mutations on serpin conformation. Furthermore, the variants in this study differ from those made by El-Thaher and coworkers (explained in sections 1.9.8 and 1.9.9). Residues H255 and H256 (s3B-hG) form the gate cluster and lie adjacent to the gate region towards the end of the C-sheet. Finally, the shutter cluster consists of H302 and H335 and is found at the base of the A-sheet, neighbouring the shutter region [181]. The positioning of the histidine clusters in regions concerned with the conformational dynamics of serpins implies that they have an involvement in the pH-dependent conformational switch of Hsp47 [87]. With the exception of H255, which is conventionally substituted with tyrosine or asparagine, each of these histidines is conserved within the Hsp47 family. Only H335, however, is also conserved in the serpin superfamily. These histidines are thus likely to be involved in maintaining the serpin fold [181].

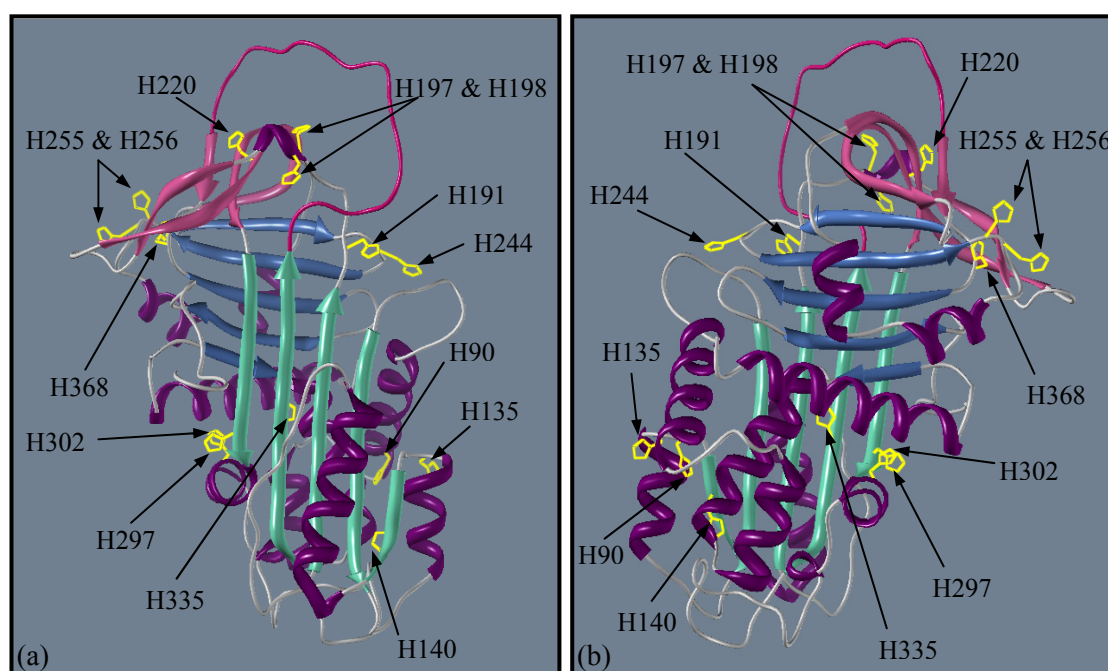
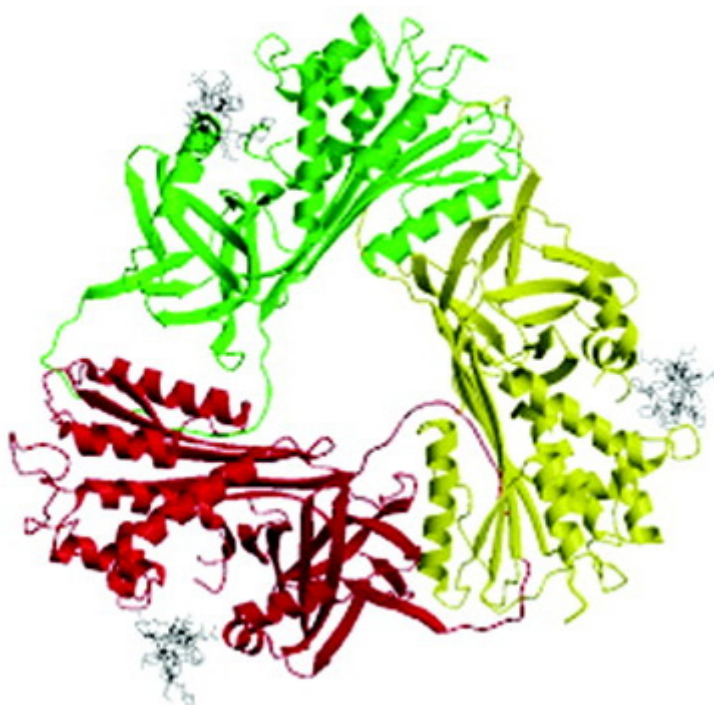


Figure 1.22. Models illustrating the positions of the histidine residues in Hsp47. Image (b) is a 180° rotation of image (a).

Observed Conformations of Hsp47

Monomeric Hsp47 is thought to form trimers by the insertion of the RCL of one molecule as strand 4 in the A-sheet. Dafforn and coworkers used extensive biophysical analyses to model the structure of the serpin polymer [87]. They proposed a ring structure (Figure 1.23), such as that seen in the S_{iiyama} mutant of α_1 -AT [171]. It is thought that such ring structures are sterically constrained, yielding only a few possible configurations of trimeric Hsp47. A proposed model suggests that the A-sheet of each Hsp47 monomer be pointing towards the centre of the ring and the helices that pack behind the A-sheet be exposed to solvent [87]. This includes the proposed collagen binding cleft between helices hA and hG/hH [182], leaving it plausible that trimeric Hsp47 has three binding sites for three collagen peptides [87].



[87]

Figure 1.23. Model of a closed trimer of Hsp47. Trimer formation is a result of individual monomers (coloured red, green and yellow) being linked by insertion of the RCL of one monomer into the A-sheet of another, resulting in a 6-stranded A-sheet. The proposed binding cleft (between helices hA and hG/hH) contains a $(PPG)_{10}$ peptide (shown in white).

1.11 Objectives

To date, studies of the substrate specific chaperone, Hsp47, have focused mainly on its specificity for collagen and the role of the glycoprotein in fibrotic disease. As a consequence, the function and mechanism of action of the protein are not well understood. This may be partly due to low expression levels of the protein and the tendency of the protein to aggregate following purification thereby making the purification procedure rather difficult. In order to overcome these limitations, the Hsp47 gene needs to be cloned into a stable plasmid that would yield a clone which produced consistent expression of the protein. Subsequently, a purification procedure comprising of a minimal number of steps would be required to obtain good yields of pure, soluble Hsp47. In addition to this, attempting to purify Hsp47 from different organisms could be a valuable strategy to increase the yields of pure Hsp47 for biophysical and conformational studies.

Although it has been shown that Hsp47, like the other members of its superfamily is conformationally mobile, regions of particular importance are yet to be discovered. Identification of histidine residues involved in the trans-conformational changes of Hsp47 is required to better understand the pH-dependent conformational switch mechanism associated with binding and release of collagen. A closer look at the locations of the histidines and comparison to mobile domains of other serpins should allow key histidines to be recognised. While in some ways Hsp47 possesses properties that are characteristic of a serpin, in other ways it is an unusual member of the inhibitory superfamily. For example, albeit that a number of serpins form oligomers, the existence of only two distinct forms (in this case a monomer and a trimer) of a serpin is not the norm. Antithrombin is the only other serpin that is reported to form a discrete set of oligomers [158, 183]. It would therefore be interesting to investigate the possibility of the existence of other serpin conformations for Hsp47.

1.12 Aims

*1.12.1 Cloning and Purification of *Mus musculus*, *Xenopus laevis* and *Danio rerio* Hsp47 (mHsp47, xHsp47 and zHsp47, respectively)*

The genes for wild type (WT) *Mus musculus* (mouse), *Xenopus laevis* (frog) and *Danio rerio* (zebra fish) Hsp47 were cloned into different vectors in order to discover a stable construct that produced large quantities of protein for structural characterisation studies. The difference in the protein sequence of each of the above organisms may be advantageous as it could affect protein properties, such as solubility and stability.

1.12.2 Cloning and Purification of mHsp47 Mutants

Mutants of mHsp47 were designed using site-directed mutagenesis in an attempt to identify the any key histidines. As with WT Hsp47, Hsp47 variants were cloned into a stable plasmid and purified from the *E. coli* expression system in order to obtain good yields of pure, soluble protein.

1.12.3 Biophysical Analyses of mHsp47 Mutants

Circular dichroism (CD) and analytical ultracentrifugation (AUC) were used to determine the secondary structure composition and oligomerisation state of each of the mutants. Any differences or similarities between the mutants were established and results previously obtained for WT mHsp47 were used for comparison.

1.12.4 Probing the pH-Dependent Conformational Change

pH titrations were carried out in order to distinguish the pH at which conformational transition of the serpin occurs. These involved minute additions of phosphoric acid to reduce the pH from 8.0 to 4.0, whilst constantly measuring the change in secondary structure. The

point of inflexion acquired for each mutant was compared and the pH at which Hsp47 dissociated from collagen was determined in each case.

1.12.5 Effect of Mutants on Collagen Fibrillisation

A collagen molecule undergoes many conformational changes prior to its materialization as the major component of connective tissues. The assembly of three procollagen chains with all peptide bonds being in the *trans* conformation allows the formation of a triple helical collagen molecule in the ER. Along with the many other suggestions for roles of Hsp47 in collagen biosynthesis, it has been proposed that Hsp47 may prevent the aggregation of collagen in the ER [96]. To investigate this further, the effect, if any on collagen fibre formation was measured by reducing the pH and monitoring the change in absorbance of the solution. The results for WT mHsp47 were compared to that of each of the mutants so as to investigate the effect of the mutation(s).

CHAPTER 2: Materials and Methods

2.1 Materials

Molecular Biology reagents/Plasmid expression vectors

Taq DNA Polymerase, all restriction endonucleases and T4 DNA ligase was purchased from New England Biolabs (NEB) (Hitchin, UK). dNTPs used for PCR were purchased from Bioline (London, UK). All primers were custom synthesised by Alta-Biosciences (University of Birmingham, UK). The sequence of each of the primers is shown in the Appendices (Chapter 9, Appendix A). The vectors used in this study were pET101/D-TOPO[®] (Figure 2.1), pET102/D-TOPO[®] (Figure 2.2), pET-32 Xa/LIC (Figure 2.3), pTYB12 (Figure 2.4) and pCR[®]T7/NT-TOPO[®] (Figure 2.5). The TOPO[®] vectors were purchased from Invitrogen (Paisley, UK), pET-32 Xa/LIC from Novagen (Nottingham, UK) and pTYB12 from NEB. Molecular biology ‘kits’ for the purification of PCR amplified and restriction endonuclease digested cDNA were MinElute[™] PCR Purification Kit and QiaQuick[®] PCR purification kits from Qiagen (Crawley, UK). Extraction of DNA from agarose gel slices used the QIAquick[®] Gel Extraction Kit (Qiagen). QIAprep[®] Spin Miniprep Kits, also from Qiagen, were used for isolation and purification of plasmid DNA from bacterial suspensions. Agarose, DNA loading buffer and Hyperladder IV were purchased from Bioline (London, UK).

Protein Purification reagents

Electrocompetent Top10 Cells were purchased from Sigma-Aldrich (Dorset, UK) and BL21 Star DE3 Chemically Competent Cells were purchased from Invitrogen. Liquid broth contained peptone and yeast extract both of which were purchased from Merck (Feltham, UK) and sodium chloride from Fischer Scientific (Leicestershire, UK). Agar was from Oxoid (Basingstoke, UK). Antibiotics used were ampicillin and chloramphenicol and these were purchased from Sigma. IPTG was from Bioline (London, UK). All reagents for SDS-PAGE

gels were purchased from Sigma-Aldrich as were DNase, RNase and lysozyme. SDS protein molecular weight marker was from NEB. Western blotting reagents included Anti-Hsp47 Mouse Monoclonal antibody from Stressgen Bioreagents (Cambridge, UK), Anti-mouse Ig, horseradish peroxidase linked whole antibody (from sheep) GPR purchased from GE Healthcare (Buckinghamshire, UK) and ECL western blotting reagents from Amersham (Buckinghamshire, UK). The Impact Expression system and chitin-affinity beads were purchased from NEB whilst Ni-NTA was from Qiagen. Lysis buffers contained imidazole, triton, β -mercaptoethanol from Sigma-Aldrich and Tris-HCl from Fischer Scientific. Glutathione, HEPES and DTT were purchased from Sigma, NaH_2PO_4 from BDH Laboratory Supplies (Poole, UK), urea from Fischer Scientific and polyethylene glycol 6,000 from Fluka (Buchs, Switzerland). Silver staining used thiosulphate and silver nitrate from Sigma, formaldehyde and sodium carbonate from BDH Laboratory Supplies and acetic acid from Fischer Scientific. Cleavage of fusion tags was attempted with thrombin, enterokinase and Factor Xa, all of which were purchased from Invitrogen.

Protein Characterisation reagents

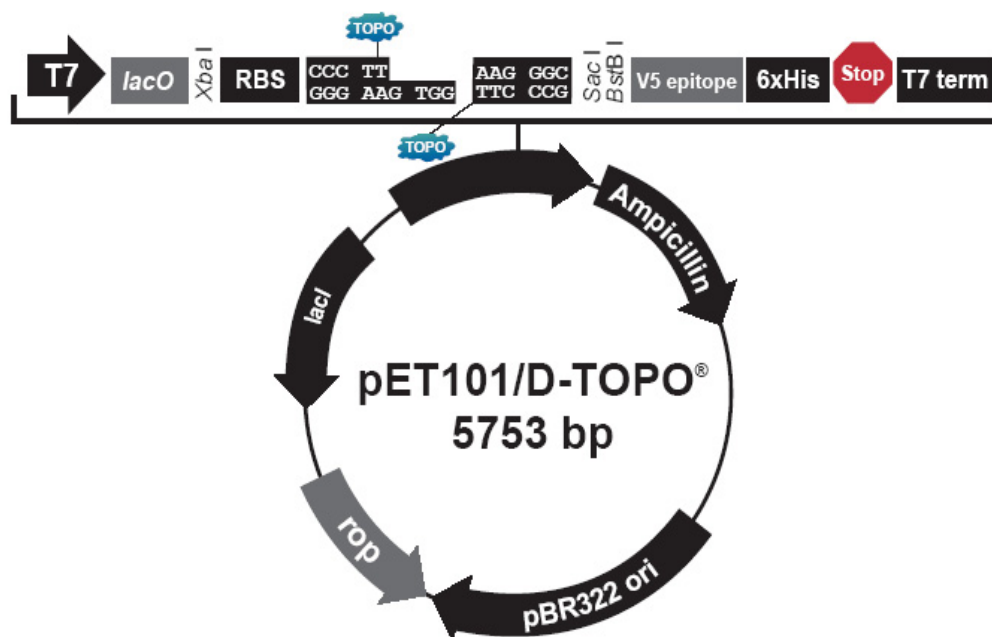
Gel filtration used a Superdex HiLoad 16/60 column from GE Healthcare. Molecular weight markers; blue dextran, ferritin, immunoglobulin G, bovine serum albumin, ovalbumin, myoglobin, ribonuclease A and Vitamin B12 were acquired from Sigma-Aldrich. Phosphoric acid was purchased from Fisons (Surrey, UK). Sephadex resin was from Sigma-Aldrich and the calf Achilles tendons were from a local butcher (Kenilworth, UK).

2.2 Methods

The Hsp47 gene from *Xenopus laevis* (frog), *Danio rerio* (zebrafish) and *Mus musculus* (mouse) was amplified using PCR for cloning into vectors with different characteristics in order to obtain pure, soluble protein. Vectors used were pET101/D-TOPO[®] (Figure 2.1), pET102/D-TOPO[®] (Figure 2.2), pET-32 Xa/LIC (Figure 2.3), PTYB12 (Figure 2.4) and pCR[®]T7/NT-TOPO[®] (Figure 2.5). A summary of the vectors and important features is shown in Table 2.1.

Vector	Antibiotic(s)	Promoter	N-terminal tag	C-terminal tag	Cleavage
pET101/D-TOPO[®]	Ampicillin	T7		6xHis	
pET102/D-TOPO[®]	Ampicillin	T7	HP-thioredoxin	6xHis	Enterokinase cleaves N-terminal tags.
pET-32 Xa/LIC	Ampicillin	T7	HP-thioredoxin S.tag His tag	6xHis	Thrombin cleaves HP-thioredoxin & S.tag. Factor Xa cleaves all N-terminal tags.
PTYB12	Ampicillin Tetracycline	T7	Self-cleavable intein tag		Presence of DTT
pCR[®]T7/NT-TOPO[®]	Ampicillin Chloramphenicol (for expression in pLysS cells)	T7	6xHis		Enterokinase

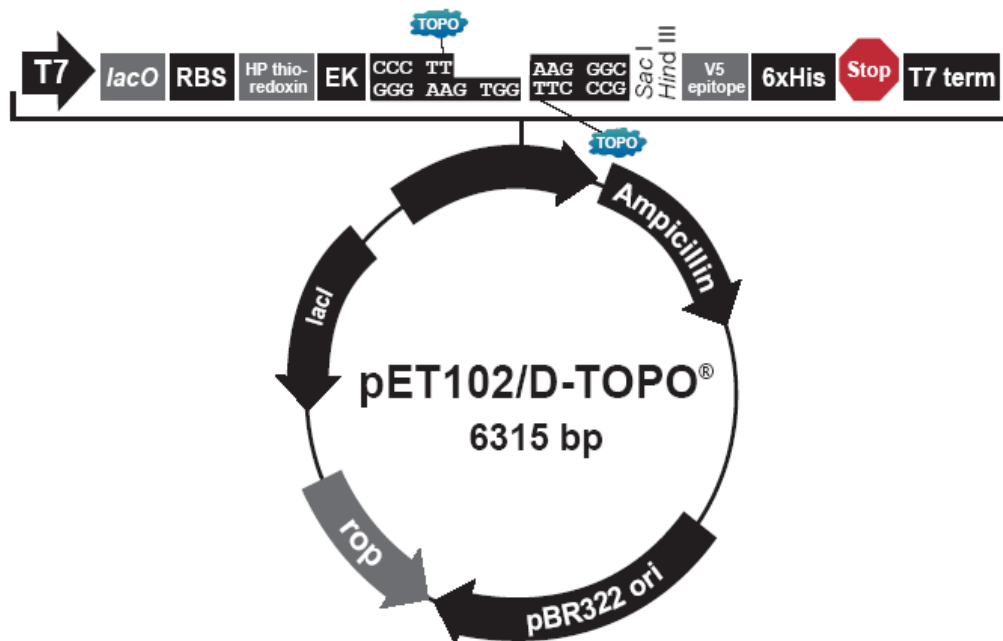
Table 2.1. Summary of vectors used.



T7 promoter: bases 209-225
 T7 promoter priming site: bases 209-228
lac operator (*lacO*): bases 228-252
 Ribosome binding site (RBS): bases 282-288, 292-296
 TOPO[®] cloning site (directional): bases 297-310
 V5 epitope: bases 333-374
 Polyhistidine (6xHis) region: bases 384-401
 T7 reverse priming site: bases 455-474
 T7 transcription termination region: bases 416-544
bla promoter: bases 845-943
 Ampicillin (*bla*) resistance gene (ORF): bases 944-1804
 pBR322 origin: bases 1949-2622
ROP ORF: bases 2990-3181 (complementary strand)
lacI ORF: bases 4493-5584 (complementary strand)

(www.invitrogen.com)

Figure 2.1. Vector map of pET101/D-TOPO[®]. Transcription is initiated from a T7 promoter in cells that produce T7 RNA polymerase. A Lac operator controls this promoter; consequently protein expression is induced on the addition of IPTG. The vector encodes a C-terminal His tag for high affinity purification of the protein on Nickel-NTA resin.



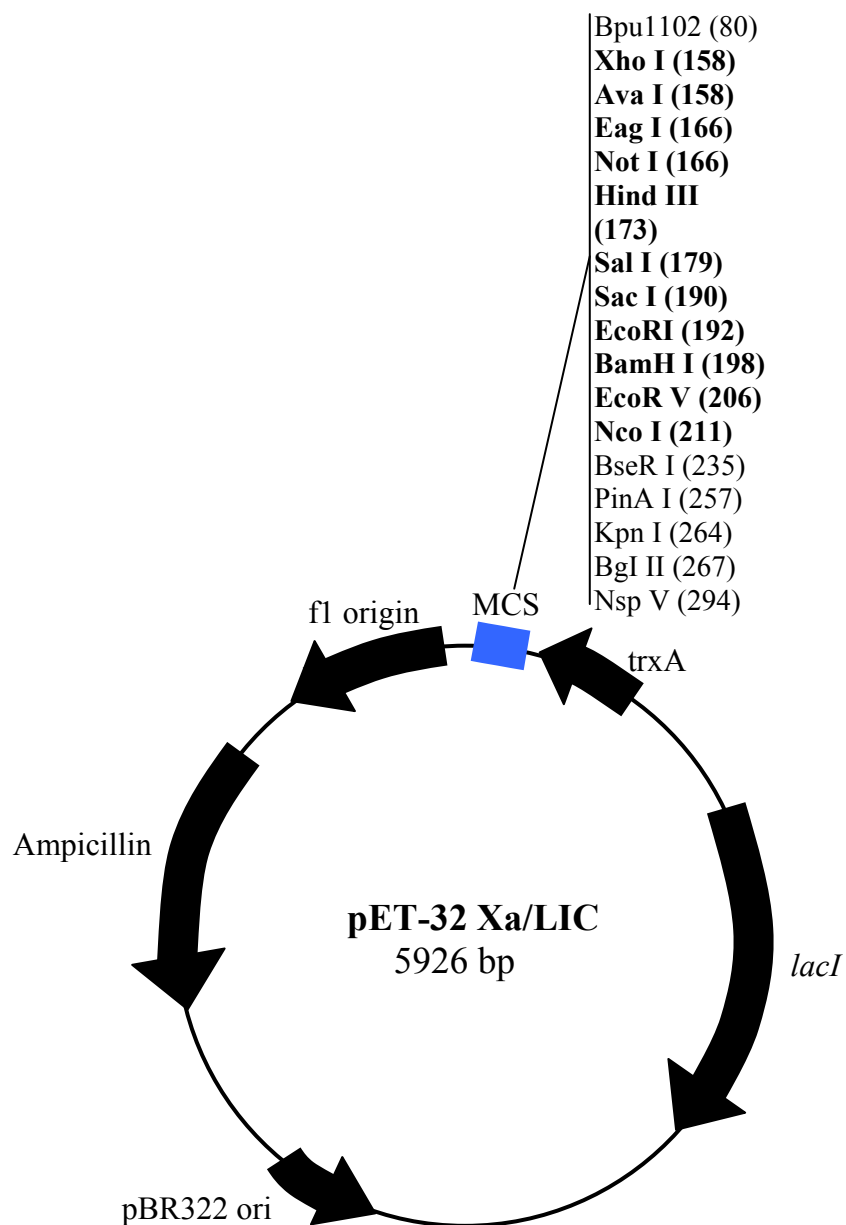
T7 promoter: bases 209-225
 T7 promoter priming site: bases 209-228
lac operator (*lacO*): bases 228-252
 Ribosome binding site (RBS): bases 282-288
 His-patch (HP) thioredoxin ORF: bases 298-627
 TrxFus forward priming site: bases 607-674
 Overhang: bases 675-678
 TOPO[®] recognition site 1: bases 670-674
 V5 epitope: bases 700-741
 Polyhistidine (6xHis) region: bases 751-768
 T7 reverse priming site: bases 822-841
 T7 transcription termination region: bases 783-911
bla promoter: bases 1407-1505
 Ampicillin (*bla*) resistance gene (ORF): bases 1506-2366
 pBR322 origin: bases 2511-3184
ROP ORF: bases 3552-3743 (complementary strand)
lacI ORF: bases 5055-6146 (complementary strand)

(www.invitrogen.com)

Figure 2.2. Vector map of pET102/D-TOPO[®]. Transcription is initiated from a T7 promoter in cells that produce T7 RNA polymerase. A Lac operator controls this promoter; consequently protein expression is induced on the addition of IPTG. The vector encodes a C-terminal His tag for high affinity purification of the protein on Nickel-NTA resin.

This vector also comprises a His tag on the C-terminus. As well as this it also contains a His-Patch thioredoxin (HP-thioredoxin), which is known to enhance the expression of

recombinant proteins, increase translational efficiency and solubility. It is termed 'His-Patch thioredoxin' because in this vector, the thioredoxin has been mutated so that it contains a metal binding domain. The mutation is of glutamate³² and glutamine⁶⁴ to histidine residues. Folding of the His-Patch thioredoxin causes the two histidines to interact with a native histidine, forming a 'patch,' which has a high affinity for divalent ions (such as Ni^{2+}). Thus HP-thioredoxin proteins can be purified using metal chelating resins such as Ni-NTA. The positioning of the ATG initiating sequence, a certain distance from the ribosome binding site (RBS), maximises translation in this vector. There is also an enterokinase (EK) cleavage site on this vector which when cleaved removes the N-terminal His tag from the recombinant fusion protein. Furthermore, the pET101/D-TOPO[®] and pET102/D-TOPO[®] vectors both possess an ampicillin resistant gene.

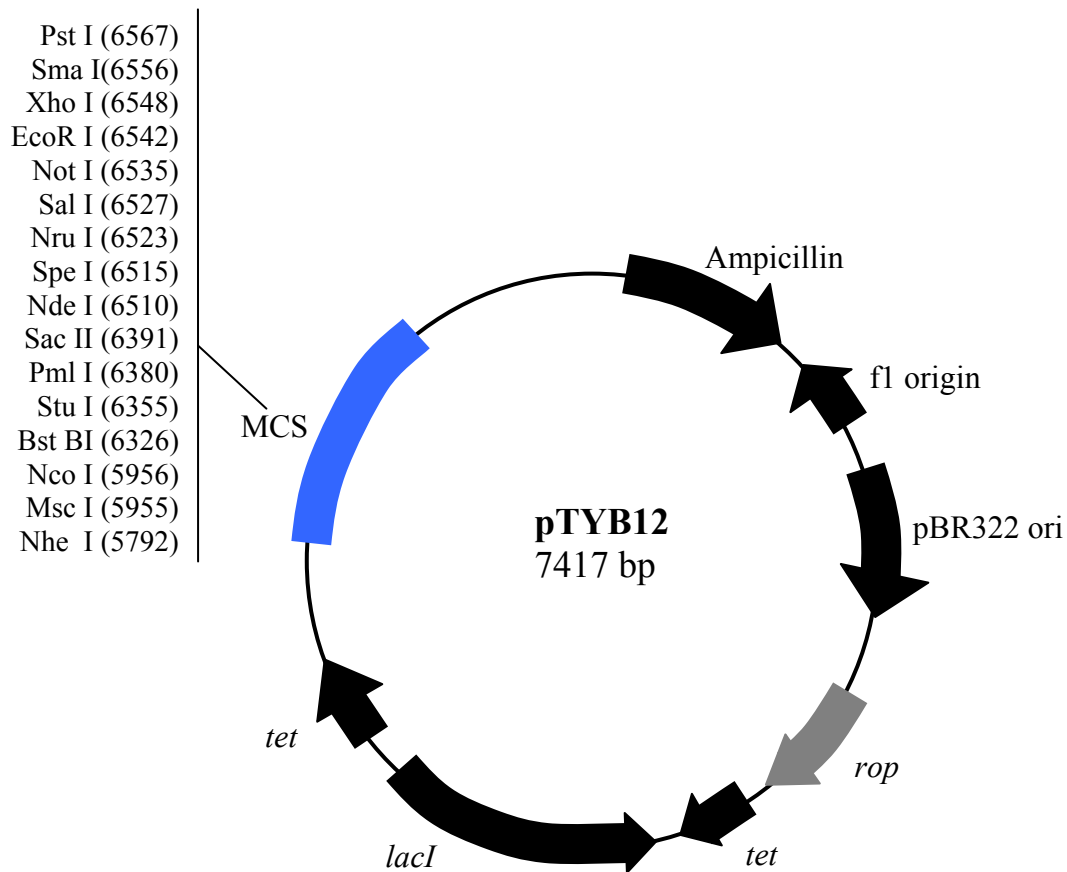


(Adapted from www.emdbiosciences.com)

T7 promoter: bases 790-806
 T7 transcription start: base 789
 HP-Thioredoxin tag coding sequence: bases 392-718
 His tag coding sequence: bases 353-370
 S tag coding sequence: bases 275-319
 Multiple cloning sites (MCS) (*BseR I-Xho I*): bases 158-225
 His tag coding sequence: bases 140-157
 T7 terminator: bases 26-72
lacI coding sequence: bases 1197-2276
 pBR322 origin: base 3710
 Ampicillin (*bla*) resistance gene coding region: bases 4471-5328
 f1 origin: 5460-5915

Figure 2.3. Vector map of pET-32 Xa/LIC.

The pET-32 Xa/LIC vector contains an N-terminal His tag and a HP-thioredoxin, separated by an S tag. The Xa/LIC vector possesses a Factor Xa (fXa) cleavage site, which when cleaved allows the removal of all the vector encoded sequences from the N-terminus of the purified protein. LIC (ligation-induced cloning) enables PCR products to be cloned without the need for digestion or ligation. LIC uses the 3' to 5' exonuclease activity of T4 DNA polymerase to create overhangs in the Xa/LIC vector. The building of 5' extensions into primers can be used to create complementary overhangs in the PCR product. Purification of the PCR product removes dNTPs and the original template plasmid. T4 DNA polymerase, in the presence of dGTP can then be used to generate overhangs that are compatible to the vector. When mixed with the vector the sticky ends form a suitable vector-insert complex that can be transformed without dissociation. Ligases within the *E. coli* then ligate the insert into the vector.



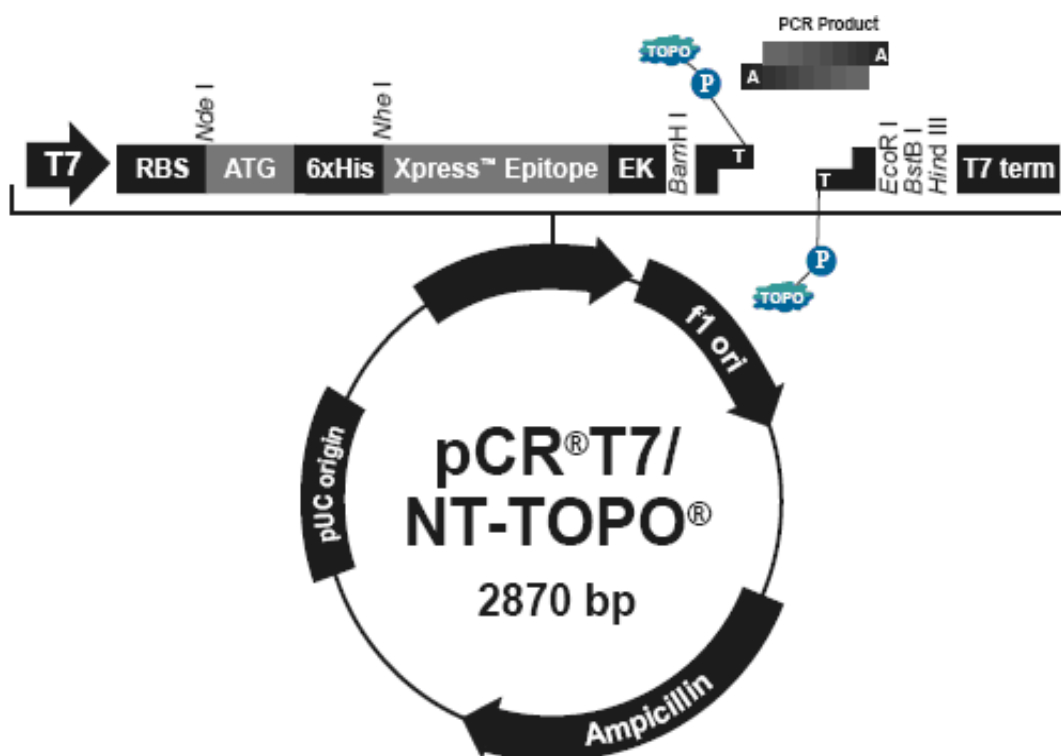
(Adapted from www.addgene.org)

T7 promoter: bases 4840-4858
Ampicillin (*bla*) resistance gene coding region: bases 140-1000
f1 origin: bases 1538-1232
pBR322 ori: bases 1650-2269
rop ORF: bases 2814-2623
Tetracycline (*tet*) resistance gene coding region (312-612): bases 3010-3309
lacI ORF: bases 4462-3371
Tetracycline (*tet*) resistance gene coding region (636-888): bases 4554-4805
lac operator (*lacO*): bases
Ribosome binding site (RBS): bases 4903-4919
T7 terminal primer: bases 6642-6624
T7 transcription termination region: bases 6582-6710

Figure 2.4. Vector map of pTYB12.

pTYB vectors are fusion vectors in which the C- or N- terminal is fused to an intein tag. pTYB12 is an N-terminal fusion vector in which the intein tag is fused to the N-terminus. As with the TOPO[®] vectors, gene expression in pTYB vectors is controlled by the T7 promoter and *lacI* gene. This vector does not contain a thioredoxin or His tag, therefore Hsp47 would

be obtained without the need to cleave unwanted tags as the chitin binding domain (CBD) (described in section 2.2.10) and intein tags are self cleaved and do not require a protease.



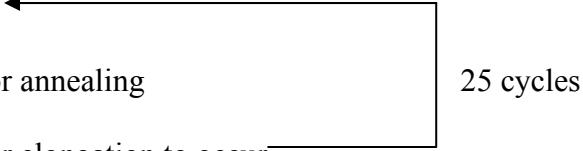
T7 promoter: bases 20-36
T7 promoter priming site: bases 20-39
Ribosome binding site (RBS): bases 87-90
Polyhistidine (6xHis) region: bases 112-129
Xpress™ epitope: bases 169-192
TOPO® Cloning site: bases 204-205
T7 reverse priming site: bases 207-289
T7 transcription termination region: bases 231-360
f1 origin: bases 845-943
Ampicillin resistance gene (ORF): bases 1017-1877
pUC origin: bases 2022-2695

(www.invitrogen.com)

Figure 2.5. Vector map of pCR®T7/NT-TOPO®. Transcription is initiated by the T7 promoter which is controlled by the *lac* operator, thus protein expression is induced by IPTG. The vector also encodes a 6xHis tag, which allows affinity purification using Nickel-NTA resin.

2.2.1 PCR

Primers (described in Chapter 11, Appendix A) were used to perform a PCR to amplify the required gene. A scale up (x 10) of the PCR procedure from Champion™ pET Directional TOPO® Expression Instruction Manual was used. Hsp47 cDNA was used as the template. All components except Taq polymerase were mixed. The Taq polymerase was added at the end as this commences the reaction. Five repeats were done so that the unsuccessful PCR reactions could be discarded and successful ones used for further studies. The PCR cycle was:

1. 95 °C for 5 minutes – this allowed separation of the strands
 2. 95 °C for 30 seconds
 3. x °C for 30 seconds – for annealing
 4. 75 °C for 2 minutes – for elongation to occur
 5. 75 °C for 10 minutes
- 

In Step 3, varying temperatures were used for optimisation of the annealing temperature. This is denoted by 'x' in the PCR cycle above. The time period for the annealing reaction was also varied depending on the primers and vector in the experiment. An annealing temperature of 30 seconds was optimal for most PCRs that were performed.

The PCR products were loaded onto 1% agarose gel (50 ml of 1 x TAE, 0.5 g agarose).

Agarose Gel Electrophoresis

50 x TAE buffer composition was as follows:

Component and final concentration	Amount added per litre
2 M Tris base	242 g
1 M acetate	57.1 ml of glacial acetic acid
100 mM EDTA	200 ml of 0.5 M (pH 8.0)
Water	to make 1 litre

Table 2.2. Composition of 50 x TAE buffer.

The gel contained 0.4 mg/ml ethidium bromide. Ethidium bromide is fluorescent when in DNA so bands on the gel (indicating presence of DNA) monitored the presence of successful PCR product. The gel was run for 40 min at 80 V. 5 µl of HyperLadder I was loaded alongside the samples, this is a molecular weight marker, which is particularly useful for quantifying and determining the size of the DNA.

Hsp47 PCR products were purified using either a QIAquick[®] Gel Extraction Kit or QIAquick[®] PCR Purification kit.

QIAquick[®] Gel Extraction Procedure

A sterile scalpel was used to excise the DNA fragment from the agarose gel. Subsequently, the slice was weighed in a tube and 3 volumes of Buffer QG was added to 1 volume of gel (100 mg ~ 100 µl). The mixture was incubated at 50 °C until the gel slice had completely dissolved. The tube was vortexed every 2-3 minutes during incubation to help dissolve gel. Once the gel slice had completely dissolved, the colour of the mixture was checked and was yellow. A QIAquick spin column was placed in a provided 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick column and centrifuged for 1 minute. The flow-through was discarded and the QIAquick column was placed back in the same collection tube. The column was washed by adding 0.75 ml of Buffer PE and centrifugation for 1 minute. The flow-through was discarded and the QIAquick column centrifuged for an additional 1 minute at 13, 000 rotations per minute (rpm). The QIAquick column was placed into a clean 1.5 ml microcentrifuge tube. DNA was eluted by the addition of 50 µl Buffer EB (10 mM Tris-HCl, pH 8.5) to the centre of the QIAquick membrane followed by centrifugation for 1 minute.

QIAquick PCR Purification Procedure

5 volumes of Buffer PB were added to 1 volume of the PCR sample and mixed. A QIAquick spin column was placed in a provided 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick column and centrifuged for 30-60 seconds. The flow-through was discarded and the QIAquick column placed back into the same tube. To wash, 0.75 ml Buffer PE was added to the QIAquick column and centrifuged for 30-60 seconds. The flow-through was discarded and the QIAquick column placed back in the same tube. The column was centrifuged for an additional 1 minute. The QIAquick column was placed in a clean 1.5 ml microcentrifuge tube. To elute DNA, 50 µl Buffer EB (10 mM Tris-HCl, pH 8.5) were added to the center of the QIAquick membrane and the column centrifuged for 1 minute.

2.2.2 Insertion of Genes into Vectors

Cloning into pET101/D-TOPO[®] and pET102/D-TOPO[®] Vectors

Cloning into the TOPO[®] vectors followed the Invitrogen Champion[™] pET Directional TOPO[®] Expression Kit Version G 2004 protocol. Directional TOPO[®] cloning enables cloning of blunt-ended PCR products in a 5' to 3' orientation directly into an expression vector using a five minute ligation reaction, eliminating the need for subcloning steps. Directional TOPO[®] cloning vectors contain a single-stranded GTGG overhang on the 5' end and a blunt end on the 3' end. The four-nucleotide overhang invades the double-stranded DNA of the PCR product and anneals to the CACC sequence placed in the 5' primer. Subsequently, ligation of the PCR product in the correct orientation is done by Topoisomerase I (Figure 2.6 below).

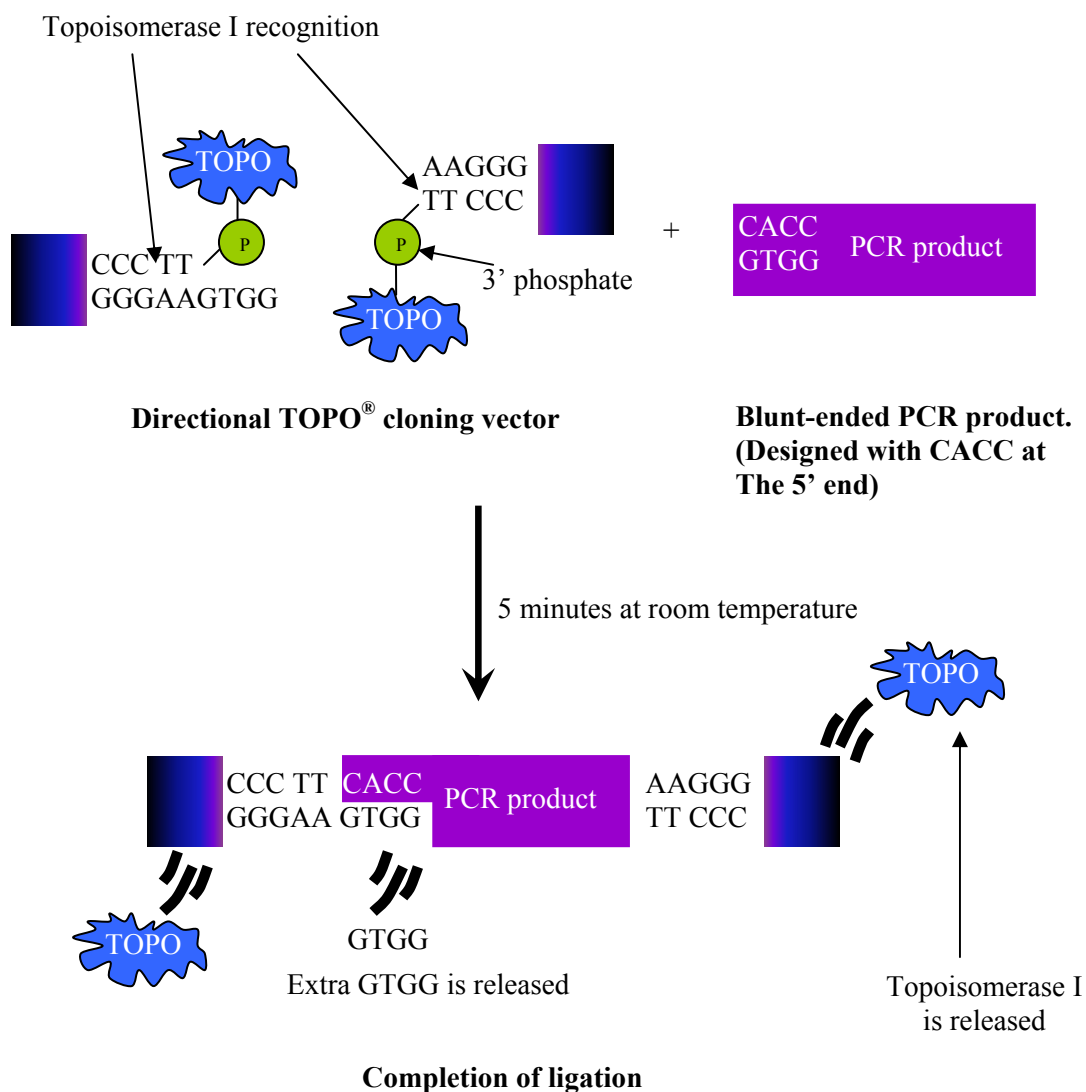


Figure 2.6. The mechanism of Directional TOPO® Cloning.

Cloning into the pET-32 Xa/LIC Vector

This was done using the Novagen Xa/LIC Cloning Kits User Protocol TB184. Ligation independent cloning (LIC) allows directional cloning of PCR products without the requirement for restriction enzyme digestion or ligation reactions. The Xa/LIC vectors possess a Factor Xa cleavage site, which allows vector encoded sequences to be removed from the purified protein. As all the Xa/LIC vectors possess the same Xa/LIC cloning site, Xa/LIC treated target inserts can be annealed to Xa/LIC vectors in a 5-minute reaction.

LIC uses the 3' to 5' exonuclease activity of T4 DNA Polymerase to create specific single stranded overhangs in the Xa/LIC vectors. The building of appropriate 5' extensions into primers creates PCR products with complementary overhangs. Purification of the PCR product to remove dNTPs and subsequent treatment with T4 DNA Polymerase in the presence of dGTP generates specific vector compatible overhangs. Cloning the LIC method is highly efficient as annealing produces only the desired product.

Procedure

Assembly of the components in Table 2.3 below into a sterile microcentrifuge tube and incubation on ice.

Component	Concentration	Volume (μl)
Purified PCR product	20 nM	10
T4 DNA polymerase buffer	10 x	2
dGTP solution	25 mM	2
DTT solution	100 mM	1
T4 DNA polymerase	5000 U/ml	0.4
Sterile Water	-	4.6
Final Volume	-	20

Table 2.3. The reaction mix for treatment of the insert with T4 DNA Polymerase prior to ligation into pET-32 Xa/LIC.

The reaction was started by addition of the enzyme and mixed with the pipette tip. The mixture was incubated at 22 °C for 30 minutes. The enzyme was inactivated by incubation at 75 °C for 20 minutes. The prepared insert could now be annealed into the Xa/LIC vector.

Cloning into the pTYB12 Vector

Cloning directly into pTYB12

Following PCR, the amplified DNA insert along with the pTYB12 vector was digested with *Bsm* I and *EcoR* I restriction enzymes in order to produce compatible sticky ends for the insertion of the Hsp47 gene into the vector.

Cloning into pTYB12 via the shuttle vector pET-32 Xa/LIC

The successfully amplified Hsp47 gene was cloned firstly into the pET-32 Xa/LIC shuttle vector, following the same protocol as above. Subsequently, the insert was cut out of the Xa/LIC vector using restriction enzymes (*EcoR* I and *Nde* I) and cloned into pTYB12, which had previously been cleaved using the same restriction enzymes. This method allowed the sequences at either end of the insert to be known, ensuring ligation into the pTYB12 vector. Following the cutting of the Hsp47 insert from the pET-32 Xa/LIC vector, the DNA was run on a 2% agarose gel and purified using the Novagen Xa/LIC Cloning Kits User Protocol TB184.

Cloning into pCR[®]T7/NT-TOPO[®]

Cloning of the mHsp47 gene into pCR[®]T7/NT-TOPO[®] was carried out by Dr. Taka and Dr. Olerenshaw at Imperial College London. mHsp47 cDNA was amplified by PCR using a 5' primer with a *Hind* III restriction site and a 3' primer with an *Nde* I restriction site, respectively. Subsequently, the Hsp47 gene was cloned into the pCR[®]T7/NT-TOPO[®] vector containing the equivalent restriction sites. Mutations were introduced using a QuickChange Site Directed Mutagenesis Kit (Stratagene). Each construct was sequenced to confirm that the correct mutation had been incorporated. Transformation of the plasmids into *E. coli* BL21 star (DE3) pLysS cells allowed expression of the protein. In addition to the advantages of BL21 star (DE3) competent cells (described in section 2.2.5), the BL21 star (DE3) pLysS strain

produces T7 lysosyme to reduce basal level expression of the gene of interest, thereby preventing leaky protein expression. In addition to the ampicillin resistance conferred by BL21 star (DE3) cells, pLysS strains confer resistance to chloramphenicol (Cam^R).

2.2.3 Ligation

Ligation into the TOPO[®] Vectors

Directional TOPO[®] cloning was used to ligate the PCR product into the appropriate TOPO[®] vectors. A 1:1 molar ratio of PCR Product: TOPO[®] vector was used in the ligation mix (Table 2.4) as this was thought to give the highest cloning efficiency.

Ligation Mix

Ligation Mix Component	Concentration of Stock	Volume for ligation of <i>Danio rerio</i> Hsp47 (μl)	Volume for ligation of <i>Xenopus laevis</i> Hsp47 (μl)
PCR Product	x	0.5	0.4
Salt Solution	1.2 M NaCl 0.06 M MgCl ₂	1.0	1.0
Sterile Water	-	3.5	3.6
TOPO Vector	15-20 ng/μl	1.0	1.0

Table 2.4. Components of ligation mixes for cloning into TOPO[®] Vectors.

Ligation mixes were incubated at room temperature for 5 minutes after being mixed gently. They were then placed on ice.

Ligation into pET-32 Xa/LIC

The LIC (ligation-independent cloning) system involves the treatment of the PCR product with T4 DNA Polymerase to produce compatible sticky ends for the ligation into the Xa/LIC cloning site (Figure 2.7).

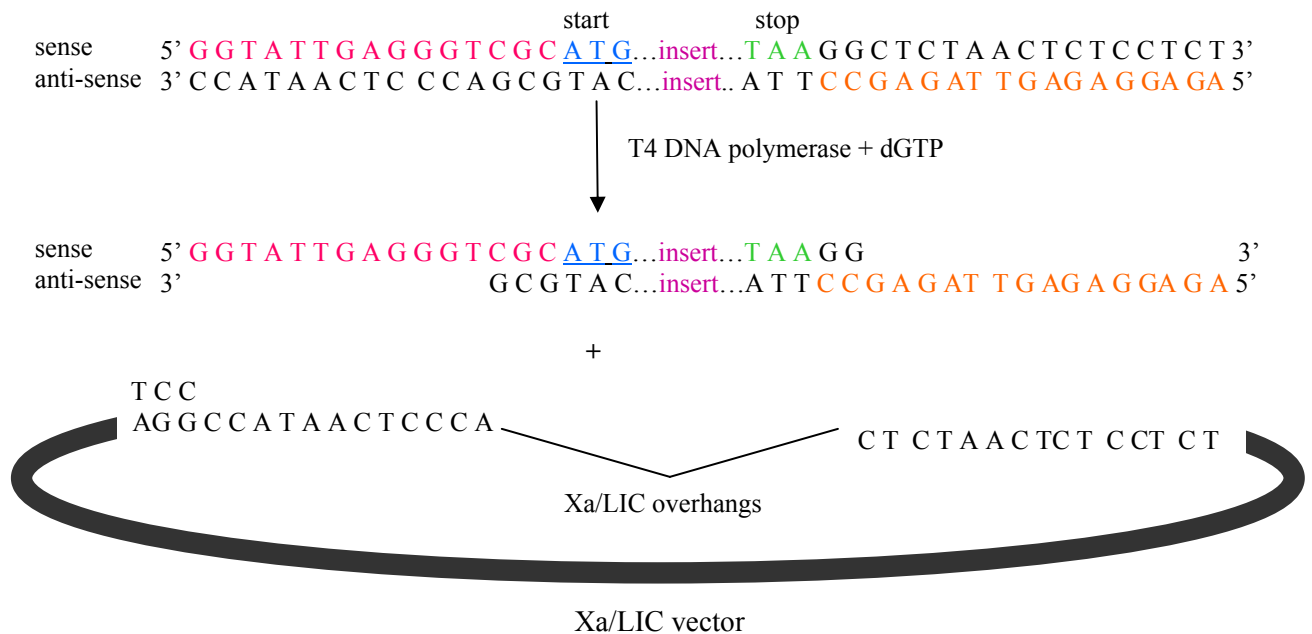


Figure 2.7. The mechanism of Xa/LIC cloning using T4 DNA polymerase. Amplification of the PCR insert with primers including the 5' LIC extensions and subsequent treatment with T4 DNA Polymerase in the presence of dGTP produced overhangs that were compatible to those of the Xa/LIC vector.

Following treatment of the insert with T4 DNA Polymerase (components described in section 2.2.2 above), the vector and the Xa/LIC insert were annealed.

Annealing the Vector and Xa/LIC Insert

A 2:1 molar ratio of PCR Product: Xa/LIC vector (0.01 pmol/μl) was used in the ligation mix to anneal the T4 DNA polymerase treated insert to the pET-32 Xa/LIC vector.

The procedure was as follows:

The following components were assembled in a sterile microcentrifuge tube:

Component	Concentration	Volume (µl)
Xa/LIC Vector	1 µg	1
T4 DNA Polymerase treated Xa/LIC insert	0.02 pmol	2
EDTA	25 mM	1
TOTAL VOLUME	-	4

The mix was incubated at 22 °C for 5 minutes prior to the addition of 1 µl of EDTA (25 mM).

The resulting mix was incubated at 22 °C for a further 5 minutes.

Ligation into pTYB12

Ligation directly into pTYB12

Before the PCR product could be inserted into the pTYB12 vector, both the insert and vector had to be digested with the same restriction enzymes in order to generate complimentary sticky ends for ligation. Digestion mixes (Table 2.5) had a final volume of 60 µl.

Digestion Mix

Component	Stock Concentration	Volume (µl)
Hsp47 insert/pTYB12 vector	40 ng/µl	40
<i>EcoR</i> I restriction enzyme	20, 000 units/ml	3
<i>Nde</i> I restriction enzyme	20, 000 units/ml	3
Buffer at pH 7.5 containing: Tris Acetate, Potassium Acetate, Magnesium Acetate, Dithiothreitol	20 mM 50 mM 10 mM 1 mM	6
Sterile water	-	8

Table 2.5. Components of digestion mix for the production of complementary sticky ends.

The digested inserts and vector were separated on a 1% agarose gel and purified using the QIAquick[®] Gel Extraction Kit. Subsequently, the insert and pTYB12 vector were ligated (mixes shown in Tables 2.6 and 2.7) using Quick[™] and standard T4 ligase.

Quick™ Ligation

50 ng of vector was mixed with a 3-fold molar excess of insert (Table 2.6). The final volume was made up to 10 µl with sterile water. Next, 10 µl of 2 x Quick Ligation Reaction Buffer was added and the reaction mixed. Subsequently 1 µl of Quick T4 DNA Ligase was added and the components mixed thoroughly. This was followed by brief centrifugation and incubation at room temperature for 5 minutes. The reaction was then chilled on ice and used to transform *E. coli*.

Component	Concentration	Volume (µl)
Hsp47 Insert	8 ng/µl	0.9
pTYB12 vector	8 ng/µl	2
Sterile water	-	7.1
FINAL VOLUME	-	10

Table 2.6. Quick ligation mix. The volume of insert and vector required was done by estimating their concentration following clean up.

Overnight Ligation

The components of the ligation mix (Table 2.7) were combined and the reaction was left at 16 °C overnight.

Component	Concentration	Volume (µl)
Reaction Buffer (pH 8.0): Tri-HCl MgCl ₂ NAD Dithiothreitol BSA	30 mM 4 mM 26 µM 1 mM 50 µg/ml	0.5
Ligase	10, 000 units/ml	0.2
pTYB12	8 ng/µl	2
Hsp47 Insert	8 ng/µl	0.3
Sterile Water	-	2
FINAL VOLUME	-	5

Table 2.7. Overnight ligation mixes.

Ligation into pTYB12 via pET-32 Xa/LIC as a shuttle vector

The Hsp47 insert was ligated first into the pET-32 Xa/LIC vector as described above. It was then cut out of the LIC vector using restriction enzymes, identical to those used for cutting pTYB12. This produced complementary sticky ends that would ligate easily. Ligation of the complementary insert and vector followed the overnight ligation procedure described above.

2.2.4 Transformation of One Shot[®] Top10 Chemically Component *E. coli* Cells

The ligation mixes were used to transform One Shot[®] Top10 Chemically Component *E. coli* cells. This makes the cells excellent hosts for the recombinant plasmids and allows the reproduction, characterisation, propagation and maintenance of the plasmid. The Top10 cells do not contain the T7 RNA polymerase so expression of the gene should not occur in the absence of the inducer.

Transformation procedure

3 µl of the TOPO[®] Cloning reaction from **Performing the TOPO[®] Cloning Reaction** were added to a vial of One Shot[®] TOP10 Chemically Competent *E. coli* cells and mixed gently by inversion. The mixture was incubated on ice for between 5 to 30 minutes. The cells were then heat-shocked for 30 seconds at 42 °C without shaking. The tubes were immediately transferred to ice. 250 µl of S.O.C. medium (0.5% Yeast extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM Glucose) were then added to the tubes. The tubes were capped tightly and shaken (200 rpm) in an incubator at 37 °C for an hour. On completion of the transformation, the sample was spread on liquid broth (LB) agar (10 g Tryptone, 5 g Yeast extract, 10 g NaCl and 15 g Agar in 1 liter de-ionised water) plates containing the appropriate antibiotic at the appropriate concentration (see table 2.8 below). The plates were incubated overnight at 37 °C.

Vector	Antibiotic	Concentration (µg/ml)
pET101/D-TOPO [®]	Ampicillin	100
pET102/D-TOPO [®]	Ampicillin	100
pET-32 Xa/LIC	Ampicillin	100
pTYB12	Ampicillin	100
	Tetracycline	20
pCR [®] T7/NT-TOPO [®]	Ampicillin	100
	Chloramphenicol	34

Table 2.8. Table showing the different vectors used in this in this study and the corresponding antibiotic(s) and concentrations.

Growth of 5 ml Cultures and Purification of Plasmid DNA

5 ml cultures of LB medium containing the appropriate antibiotics were inoculated with selected colonies produced on the LB ampicillin plates by the cloning procedures described previously (Section 2.2.4). The cultures were left to grow overnight at 37 °C with shaking at 200 rpm. The following day the cells were harvested by centrifugation at 13,000 rpm using a bench top centrifuge.

Isolation of the plasmid involved following a QIAprep[®] Spin Miniprep Kit (described below).

QIAprep[®] Spin Miniprep Kit

Pelleted bacterial cells were resuspend in 250 µl Buffer P1 and transferred to a microcentrifuge tube. 250 µl Buffer P2 were added and the tube was gently inverted 4-6 times. 350 µl Buffer N3 was added and the tube inverted immediately but gently 4-6 times. The mixture was centrifuged for 10 minutes at 13, 000 rpm in a table-top centrifuge. The supernatant from Step 4 was applied to the QIAprep spin column by decanting or pipetting prior to centrifugation for 30-60 seconds. The QIAprep spin column was washed by adding 0.5 ml Buffer PB and centrifuging for 30-60 seconds. The flow-through was discarded. The QIAprep spin column was washed by adding 0.75 ml Buffer PE and centrifuging for 30-60

seconds. The flow-through was discarded and centrifuged for an additional 1 minute to remove residual wash buffer. The QIAprep column was placed in a clean 1.5 ml microcentrifuge tube. To elute DNA, 50 µl Buffer EB (10 mM Tris-Cl, pH 8.5) were added, left to stand for 1 minute and centrifuged for 1 minute.

The presence of plasmid was confirmed by separating 2 µl samples of the miniprep eluate on a 1% agarose gel. PCRs (where the miniprep eluate is used as the template in a PCR reaction) were also performed to ensure that the gene was present in the plasmid.

The genes within the plasmid were sequenced using ABI PRISM[®] BigDye[™] Terminator v3.0 Ready Reaction Cycle Sequencing Kit.

2.2.5 Transformation of BL21 star (DE3) One Shot Cells

Recombinant plasmids containing the correctly sequenced genes were used to transform BL21 star (DE3) One Shot *E. coli* Cells. These competent cells contain the λ DE3 lysogen which carries the gene for T7 RNA polymerase under the control of the lac UV5 promoter. This allows for high levels of protein expression that is induced by IPTG. These strains carry a mutated *rne* gene which encodes a truncated RNase enzyme that lacks the ability to degrade mRNA, resulting in an increase in mRNA stability. The (DE3) and pLysS strains do not contain the *lon* protease. They are also deficient in the outer membrane protease, OmpT. The lack of these proteases reduces degradation of heterologous proteins expressed in the strains.

One 50 µl vial of BL21 star (DE3) One Shot *E.coli* Cells was used per transformation. 1 µl of the recombinant plasmid was added to a vial, which was incubated on ice for 30 minutes. The cells were then heat shocked at 42 °C for 30 seconds and recovered by the addition of 250 µl

of S.O.C medium. The transformation reaction was incubated at 37 °C for 30 minutes with shaking prior to plating on a LB ampicillin plate, which was incubated at 37 °C overnight.

Successful transformation into BL21 star (DE3) competent cells yields a construct that is ready for protein expression experiments.

2.2.6 Determination of Protein Expression Levels

Colonies from the successful transformants were selected and used to inoculate 5 ml of LB medium containing the required antibiotic(s) (see Table 2.7). These cultures were incubated at 37 °C overnight with shaking at 200 rpm. The following day, 1 ml from the overnight samples was diluted to 3 ml with LB media containing appropriate antibiotics. From the resulting 3 ml, 1 ml was retained as the uninduced sample. 2 µl IPTG (1 M) was added to the remaining 2 ml of culture to induce the protein expression. The samples were then incubated at 37 °C for a further 4 hours. Subsequently, 1 ml of the induced samples was harvested by centrifugation at 13,000 rpm for 1 minute using a desktop microcentrifuge along with the 1 ml uninduced sample retained earlier. The pellet resulting from spin was resuspended in 20 µl of SDS loading buffer. Samples were then boiled for 5 minutes and SDS-PAGE (described in section 2.2.12) was used for the analysis.

2.2.7 Purification of the Protein expressed in TOPO[®] and Xa/LIC vectors

E. coli colonies that had been shown to be expressing protein were used to inoculate 5 mls of LB containing 100 µg/ ml ampicillin. Following confirmation of successfully expressing cells, 4 ml out of the 5 ml overnight culture was used to secondary inoculate 1 litre of LB medium containing ampicillin (100 µg/ ml). The cells were left to grow at 37 °C until an OD_{600nm} of approximately 0.6 was reached, at which point protein expression was induced by the addition

of 1 ml IPTG (1 M). The cells were then incubated overnight at 16 °C with shaking at 200 rpm.

The following day cells were harvested by centrifugation at 6000 x g for 10 minutes using a F10-6 x 500 rotor in a Beckman J2-2 Lite centrifuge. The pellet resulting from the spin was resuspended in 50 ml Tris-HCl buffer (100 mM Tris, 150 mM NaCl), pH 8.0. After the addition of 10 µg/ ml lysozyme, 10 µg/ ml DNase and 5 µg/ ml RNase, the cells were lysed by sonication which involved 20-second pulses followed by 20 second intervals for 4 minutes. The cell lysate produced was then centrifuged at 15,000 rpm for 10 minutes in a Beckman J2-2 Lite Centrifuge. Centrifugation produced a pellet (cell debris) and supernatant (crude lysate).

The presence of a His tag at the N-terminal of the protein enabled the use of nickel-nitrilotriacetic acid (Ni-NTA) metal affinity chromatography (section 2.2.7.1 below) to purify the protein. Prior to the application of the supernatant, the Ni-NTA resin was equilibrated with Tris-HCl buffer (100 mM Tris, 150 mM NaCl), pH 8.0. Once the supernatant had been applied to the resin, supernatant-resin interactions were allowed to occur with occasional turning.

The supernatant-resin mix was centrifuged at 3,200 rpm for 1 minute using a Sorvall RT 6000B centrifuge in a Sorvall H1000B swing out rotor and the supernatant removed. This fraction was defined as flow-through (FT) and would comprise of protein that did not bind the Ni-NTA beads. The column was then washed with 40 ml Tris-HCl buffer, pH 8.0 (100mM Tris, 150 mM NaCl) to remove any unwanted material. After spinning at 3,200 rpm for 1 minute, the supernatant was removed and the step was repeated with a further 40 ml of Tris-HCl buffer. Washing the column with 40 ml of low Imidazole-Tris buffer (20 mM Imidazole,

100 mM Tris, 150 mM NaCl, pH 8.0) eluted any partially bound protein. Centrifugation at 3,200 rpm for 1 minute was followed by removal of the supernatant. This step was repeated. A high concentration of imidazole releases the His tags from the resin, permitting the elution of the His tagged protein. Therefore, 10 ml of high imidazole-containing Tris-HCl buffer (100 mM Tris, 150 mM NaCl, 250 mM imidazole, pH 8.0) was added to the resin in order to elute the protein. The column was centrifuged at 3,200 rpm for 1 minute and the supernatant retained.

Cell debris, crude lysate, flow-through, wash, elution and resin fractions were analysed by SDS-PAGE.

2.2.7.1 Ni-NTA Chromatography

Constructs comprising HP-thioredoxin or histidine tags can be purified using Ni-NTA affinity chromatography. Ni-NTA is a tetradentate chelating adsorbent that is thought to overcome leaching. NTA occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free for interaction with the 6xHis tag (Figure 2.8).

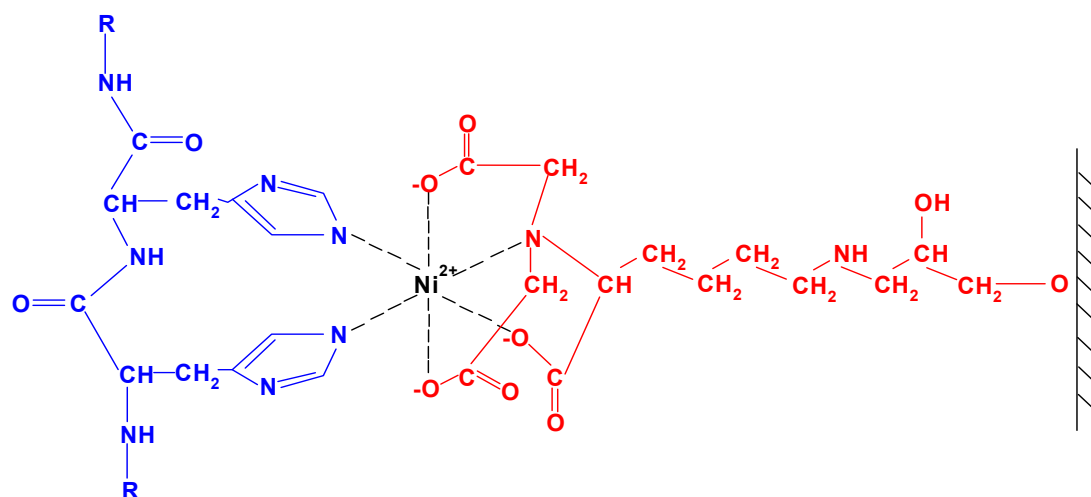


Figure 2.8. The interaction between the 6xHis tag and Ni-NTA.

Non-specific binding is reduced by low concentrations (10-20 mM) of imidazole, which reduces the potential for binding background contaminants. The imidazole rings in the

histidine residues of the 6xHis tag bind to the nickel ions, immobilised by the NTA groups, on the matrix. The binding of imidazole to the nickel ions disrupts the binding of dispersed histidines residues in non-tagged background proteins. Low imidazole concentrations prevents non-specific, low affinity binding of background proteins while still allowing string binding of 6xHis tagged proteins to the Ni-NTA matrix.

Endogenous proteins with histidine residues that interact with the Ni-NTA groups can be washed out of the matrix with stringent wash conditions such as an increase in imidazole concentration (10-50 mM).

A further increase in the imidazole concentration (100-250 mM) instigates the dissociation of 6xHis tagged proteins from the Ni-NTA as they can no longer compete for the binding sites on the Ni-NTA resin.

2.2.8 Purifying Hsp47 from Insoluble Intracellular Aggregates

Unfolding the protein

In some cases Hsp47 was retained in *E. coli* as inclusion bodies. To purify the protein from these inclusion bodies involved unfolding and refolding before purification could be attempted. As before (section 2.2.7), a large culture (1-6 L) *E. coli* was established, the cells were spun, pellet resuspended in 100 mM Tris buffer pH 8.0 containing 150 mM NaCl and sonicated. As the protein was to be purified using the insoluble fraction, a temperature of 37 °C was used even after induction. In the previous culture, the temperature was lowered to prevent aggregation of the protein. After sonication, the sample was spun again at 3,200 rpm using a JA20 (fixed angle) rotor and a Beckmann J2-21 centrifuge. The pellet was resuspended in 100 mM Tris at pH 8.0 containing 0.1% Triton to wash away the membrane,

spun again at 3,200 rpm (again using the JA20 rotor and Beckmann J2-21 centrifuge) and the supernatant resulting from the spin was discarded. The pellet was resuspended in 40 ml buffer containing 100 mM Tris-HCl, 100 mM NaH₂PO₄ and 8 M urea at pH 8.0 (urea was added to unfold the protein). The resulting solution was added to 20 ml of a Ni-NTA resin in a 50 ml Falcon tube for purification and as before, unbound, bound, wash and elution fractions were collected for analysis on SDS-PAGE.

The only difference to the previous purification protocol being the inclusion of urea in the wash and elution buffers. Wash buffer at pH 6.3 contained 8 M urea, 100 mM NaH₂PO₄ and 10 mM Tris-HCl. Elution buffer at pH 4.5 contained 8 M urea, 100 mM NaH₂PO₄ and 10 mM Tris-HCl.

Optimisation of Refolding Hsp47

Refolding of Hsp47 was attempted using dialysis where 1 ml of the eluted protein was placed into dialysis tubing, which acted as a membrane, allowing the passage of water molecules. Four different dialysis tubes containing 1 ml of the protein were set up and each one placed in a buffer solution of different urea concentrations for varied times so that the optimum conditions for refolding could be distinguished. The urea concentrations used were 6 M, 4 M, 2 M and 0 M urea. As mentioned above, the protein solution resulting from the purification contained 8M urea, thus placing it in buffer with different urea concentrations would allow the urea to migrate from a high concentration to a low concentration, leaving pure Hsp47 in the dialysis tubing. The buffer solution consisted of: 100 mM Tris-HCl at pH 8.0

250 mM NaCl

5 mM Glutathione

6 M/4 M/ 2 M/ 0 M urea

Four dialysis experiments were set up:

- (i) At time (t) = 0 hour, the protein (sample 1) was placed in buffer containing 6 M urea. At t = 2 hours, the tube containing the protein was moved into the buffer containing 4 M urea and at t = 4 hours, the protein was placed in 2 M urea. After 6 hours the protein was placed in the 0M urea buffer and left to dialyse overnight.
- (ii) At t = 0 hour the protein (sample 2) was placed in the 4 M urea buffer, it was then moved to the buffer containing 2 M urea at t = 2 hours. Then at t = 4 hours the protein was placed in the 0 M urea buffer and left to dialyse overnight.
- (iii) At t = 0 hour, the protein (sample 3) was placed in 2 M urea and at t = 2 hours, it was placed in the buffer solution of 0 M urea and left to dialyse overnight.
- (iv) The 8 M urea protein (sample 4) solution was placed in 0 M urea buffer solution and left to dialyse until the following day.

The above method is known as step-wise dialysis. The change in urea concentration allowed the urea to slowly diffuse out of the protein, leaving pure folded Hsp47.

The following day the solutions in each of the four dialysis tubes were spun at 16,000 x g to separate the soluble and insoluble fractions. The supernatant included soluble Hsp47 whilst the pellet after spinning contained any remaining insoluble Hsp47. Protein loading buffer was added to each of the samples prior to them being boiled and SDS-PAGE was used to identify whether the unfolding and refolding was successful.

2.2.9 Cleavage of the Thioredoxin Tag

The pET-32 Xa/LIC construct contained both Factor Xa and Thrombin cleavage sites (Figure 2.9).

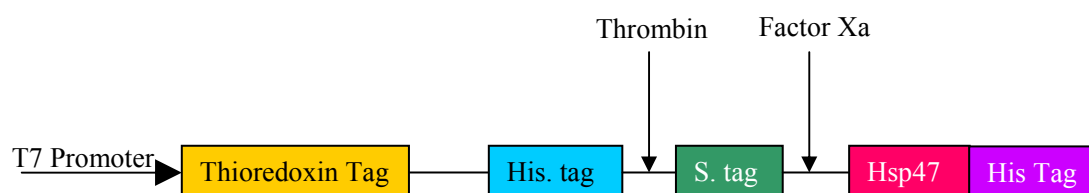


Figure 2.9. Position of Thrombin and Factor Xa cleavage sites on the pET-32 Xa/LIC construct.

An initial cleavage of the fusion tags was attempted using thrombin and factor Xa separately.

Cleavage with Thrombin

Cleavage with thrombin used a stock of 2 mg of thrombin in 1 ml water (37 units/ mg). 0.5 ml of the thrombin solution was added to 0.5 ml of protein solution, which was incubated at 4 °C. At specific time points (0 hours, 0.5 hours, 1 hour, 2 hours, 4 hours and overnight), 100 µl samples were taken and boiled after the addition of 20 µl of SDS protein loading buffer to denature the enzyme and stop the cleavage reaction. Samples were analysed by SDS-PAGE.

Cleavage with Factor Xa (fXa)

Cleavage with fXa involved the addition 10 µl of the enzyme to 200 µl of protein solution at room temperature. Again, samples were taken at regular intervals to monitor the cleavage over time. This would give an indication of the optimum time for the cleavage. As above, the samples were denatured by the addition of SDS loading buffer and boiling and analysed by SDS PAGE.

Cleavage where both Factor Xa and Thrombin sites were present

Cleavage was repeated, in which four different experiments were set up (Table 2.9). The first experiment included a control in which neither Factor Xa or thrombin was added, a second experiment was set up in which cleavage of the protein was with thrombin. The third

experiment involved cleavage of the protein using Factor Xa and a fourth experiment was set up in which a mixture of Factor Xa and thrombin were used to try to cleave the fusion tags.

Component	Concentration	Volume (μl)			
		Experiment 1	Experiment 2	Experiment 3	Experiment 4
Protein	~1 mg/ml	200	200	200	50
Thrombin	50 U	40	0	40	0
Factor Xa	400 U	0	20	20	0
Sterile water	-	20	40	0	10

Table 2.9. Reaction mix components of Thrombin and Factor Xa Cleavage Assays.

The cleavages were left to occur overnight at 4 °C and the samples analysed by SDS-PAGE.

2.2.10 Purification of protein expressed in pTYB12

The Intein Mediated Purification with an Affinity Chitin-binding Tag (IMPACT) expression system was used to express and purify the protein in *E. coli*. IMPACT uses the self-cleavage of an intein (a protein splicing element) (Figure 2.10) to separate the target protein from the affinity tag. IMPACT allows the purification of a native recombinant protein tag without the use of a protease. A target protein is fused to a self-cleavable intein tag in which a chitin-binding domain allows affinity purification of the fusion precursor on a chitin column.

The absence of any other fusion tags makes this advantageous for cloning into as it allows single-column purification and release of the fusion tag without the use of proteases.

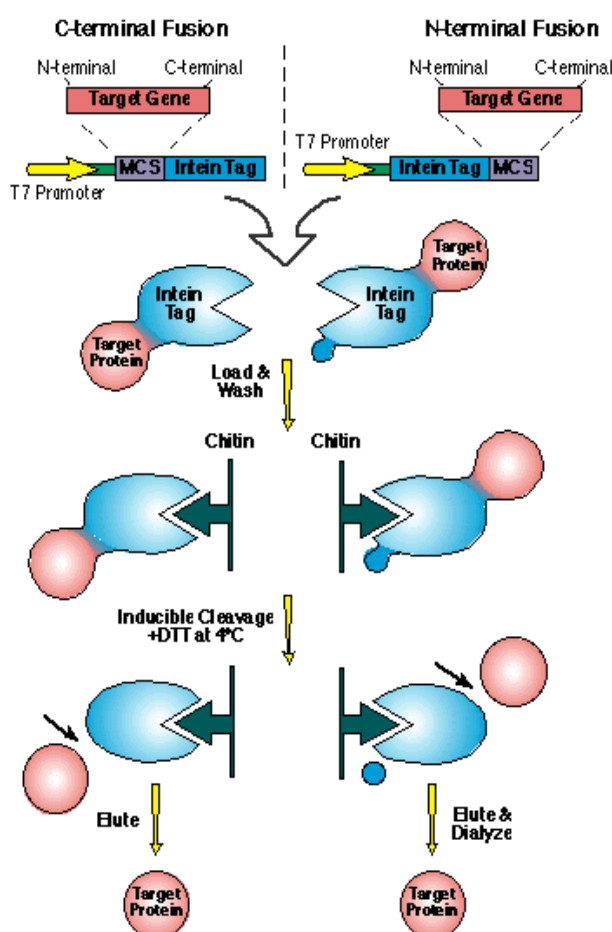


Figure 2.10. Purification of protein expressed by pTYB12. The Hsp47 gene was cloned into the multi-cloning sites (MCS). A chitin binding domain and an intein tag are encoded upstream of the cloned gene and these features enhance the purification steps. When using pTYB12 vector system, the target protein is expressed with an intein tag and can be purified by a chitin column. The intein tag binds target protein to the column. Cleavage of intein tag, induced by adding DTT, will release the desired protein.

Following confirmation that cells were successfully expressing Hsp47, a 5 ml overnight culture was used to secondary inoculate 1 litre of LB medium containing ampicillin (100 µg/ml). The cells were left to grow at 37 °C until an OD_{600nm} of approximately 0.6 was reached, at which point expression was induced by the addition of 1 ml IPTG (1 M). The cells were then incubated overnight at 16 °C with shaking at 200 rpm. Cells were harvested by centrifugation at 6,000 x g for 10 minutes using a F10-6 x 500 rotor in a Beckman J2-2 Lite centrifuge. The pellet resulting from the spin was resuspended in 50 ml cell lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 0.1% Triton). 50 µl DNase and 50 µl RNase were added to the suspension, which was then incubated at 20 °C with shaking at 200 rpm for 30 minutes. Lysozyme was not added to the suspension as it is known to bind and digest chitin. Cell lysis was by French Press rather than sonication. The resulting crude lysate was centrifuged at 20,000 x g for 30 minutes to pellet cell debris. The supernatant (cell extract) was collected

and the pellet discarded. 40 µl of the cell extract was retained for analysis by SDS-PAGE. The clarified cell extract was loaded onto a chitin column at a flow rate of 1 ml/minute. The flow-through was collected for analysis by SDS-PAGE. The chitin column was then washed with 10 bed volumes of column buffer (20 mM HEPES, 500 mM NaCl, 0.1% Triton). Due to the high affinity of the Chitin Binding Domain (CBD) for the chitin beads, a higher flow rate of 2 ml/minute and stringent wash conditions were used. The target protein is released from the chitin column when the chitin-bound intein tag undergoes self-cleavage in the presence of DTT. Therefore induction of on-column cleavage was done by quickly flushing the column with 3 bed volumes of Cleavage Buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 50 mM DTT). This allowed even distribution of thiols throughout the column. Following the quick flush the column flow was stopped and the column left at 4 °C overnight. The target protein was eluted off the column using Cleavage Buffer without DTT. 1 ml fractions were collected for 3 column volumes and fractions were analysed by SDS-PAGE. The cleavage efficiency was assessed by removing some chitin resin from the column for analysis by SDS-PAGE.

2.2.11 Purification of mHsp47 in pCR[®]T7/NT-TOPO[®] (adapted from [87])

Colonies expressing Hsp47 were used for primary inoculation of 5 ml LB containing the appropriate antibiotics (100 µg/ml ampicillin and 34 µg/ml chloramphenicol). The culture was incubated overnight at 37 °C with shaking at 200 rpm. Subsequently, the culture was used for secondary inoculation of 1 litre LB media containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. The culture was grown at 37 °C until an OD_{600nm} of 0.6 was reached, whereupon protein expression was induced by the addition of 1 mM IPTG. Consequently, the temperature was reduced to 20 °C and the cells incubated overnight at 200 rpm. The cells were harvested by centrifugation at 6,000 x g for 15 minutes at 4 °C using a F10-6 x 500 rotor in a Beckman J2-2 Lite centrifuge. The resultant pellets were resuspended in 50 ml of 50 mM Tris-HCl, pH 7.5, containing 0.5 M NaCl, 0.02% Triton-X-100 and 3 mM β-mercaptoethanol.

10 µg/ml lysozyme, 10 µg/ml DNase and 5 µg/ml RNase was added to the resuspension prior to cell lysis using a French press at 11400 psi. Cell debris was removed by centrifugation at 20,000 x g for 30 minutes at 4 °C in a Beckman Lite centrifuge. The resulting cell lysate (approximately 50 ml) was loaded onto a 15 ml gelatin-agarose column using a Pharmacia ActaPrime HPLC. The column had been pre-equilibrated with 3 column volumes of 50 mM Tris-HCl, pH 7.5, containing 0.5 M NaCl. The flow-through was collected for analysis using SDS-PAGE. The column was washed with 50 mM Tris-HCl, pH 7.5, containing 0.3 M NaCl to remove any impure or weakly binding protein. This was continued until the absorbance reached base levels and the wash fraction was collected for SDS-PAGE analysis. A 0.3 M phosphate buffer pH gradient (from pH 9.0 to pH 3.0) was used to elute the bound Hsp47 from the gelatin beads. Elution fractions of 1 ml were collected for analysis and data on the pH, absorbance and salt concentration was collected through-out the purification procedure. Following elution, the pH of the protein was restored by the addition of 50 mM Tris-HCl, pH 8.0 to each of the elution fractions. Fractions were analysed SDS-PAGE and those containing Hsp47 were dialysed against 50 mM Tris-HCl, pH 8.0 overnight. Fractions were then snap-frozen in liquid nitrogen prior to storage at -70 °C.

2.2.12 Detection of Hsp47

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

The SDS-polyacrylamide gels comprised of 10% resolving (30% acrylamide, 1.5 M Tris-HCl (pH 8.8), 10% SDS, 10% ammonium persulphate and 0.04% TEMED) and 5% stacking (30% acrylamide, 1.0 M Tris-HCl (pH 6.8), 10% SDS, 10% APS and 0.04% TEMED) gels. Samples for analysis were boiled for 10 minutes at 100 °C in 5 x SDS loading buffer. The procedure followed was that of Roskams and Rodgers 2002 [184]. Gels were run for 40 minutes at 200 volts (V) and 400 milliamps (mA) in SDS running buffer (25 mM Tris-HCl, 192 mM Glycine and 0.1% SDS) using a Biorad mini protean III apparatus. New England

Biolabs Broad Range Pre-stained markers were used as molecular weight markers. Coomassie Brilliant Blue (CBB) staining was used for visualisation of protein bands.

Visualisation by CBB

The CBB stain was made by dissolving 0.25 g of R-250 CBB in 90 ml of methanol: water (1:1 v/v) and 10 ml glacial acetic acid. The solution was filtered through a Whatman #1 filter paper to remove particulate matter. The gels were stained for a minimum of 4 hours in 5 volumes of coomassie prior to destaining. Destain contained 5:4:1 ratio of methanol: water: acetic acid and was continuously changed until distinct bands were visible on the gels.

Visualisation by Western Blot

This is a more sensitive technique (can detect ng of protein) for the detection of protein in a gel. Once the gel was run in the normal way (described above), it was transferred to nitrocellulose using a Biorad transfer apparatus (see manufacturers instructions). This time the transfer was run for 60 minutes at 80 V. Ponceau stain (2 g of Ponceau S, 30 g of trichloroacetic acid and 30 g of sulfosalicyclic acid in 100 ml of water) was then added to the nitrocellulose for a couple of seconds to visualise the transferred bands. The nitrocellulose blot was then washed for 60 minutes with PBST (8 mg/ml NaCl, 0.2 mg/ml KCl, 1.44 mg/ml $\text{Na}_2\text{HPO}_4\text{O}$, 0.24 mg/ml $\text{KH}_2\text{PO}_4\text{O}$ and 0.5 $\mu\text{l/ml}$ tween) + 5% milk, this removed any stain that was not due to presence of protein. The blot was then washed for another 60 minutes, this time with PBST + 0.5% milk + 1:10000 anti-mouse Hsp47. The anti-mouse Hsp47 is an Hsp47-specific antibody that should bind to any Hsp47 present on the gel. The gel was then washed with 1 x PBST for five minutes. This was repeated three times, to remove any unbound antibody. It was then washed for a further 60 minutes with PBST + 0.5% milk + 1:1000 anti mouse-HRP (horse radish peroxidase) conjugate. This binds to the anti-Hsp47 antibody that was used to wash the gel with previously. This antibody catalyses the

conversion of substrate to products, giving off light in the process. It is this light that is used to visualise the presence of protein. Before visualisation, however, the gel was washed three times with 1 x PBST for 5 minutes to remove any unbound antibody. A light sensitive film was used to photograph the gel.

Visualisation by Silver Stain

The silver stain method is much more sensitive than the normally used coomassie blue staining technique. Therefore minimal amounts of protein can be detected. The silver stain method required 4 solutions:

Starter solution – 37.5 mg/ 500 ml thiosulphate

Solution I – 0.3 g AgNO_3 + 67 μl of 38% formaldehyde in 80 ml water

Solution II – 10 g Na_2CO_3 + 113 μl of 38% formaldehyde in 80 ml water

Stopper solution – 10% acetic acid

Gloves were worn at all times to reduce contamination from human protein. The silver stain procedure started with washing the gel with water. This was repeated three times. Starter solution was added for 1 minute. The gel was washed with water three times. Solution I was added and left for 20 minutes with shaking. The gel was washed three times with water. Solution II was added and the bands visualised. It took about 15 minutes for the bands to become clearly visible. Solution II was then poured off and stopper solution added.

Protein Concentration Determination

The absorbance of the eluted protein was measured at 280 nm using a spectrophotometer. A 10 mm quartz cuvette was used and absorbance spectra from 220 nm to 300 nm were collected.

Determining protein concentration using the Beer-Lambert Law requires the use of the following equation: $A_{280} = \epsilon Cl$

A_{280} = Absorbance at 280 nm

ϵ = Molar extinction coefficient ($M^{-1} \text{ cm}^{-1}$)

C = Protein concentration (M)

l = Pathlength (cm)

Each protein has a unique extinction coefficient that depends on the number of tryptophan and tyrosine residues and the number of disulphide bonds within the protein. mHsp47 has a molar extinction coefficient of $42400 \text{ M}^{-1} \text{ cm}^{-1}$.

Conversion of the calculated concentration in M to mg/ml can be done by multiplying the concentration (M) by the molecular weight of the protein. In the case of Hsp47 it is 47,000 Da.

2.2.13 Characterisation of Purified Hsp47

2.2.13.1 Circular Dichroism

CD Spectra

CD spectroscopy measures the difference in absorbance of left- and right- handed circularly polarised light by chiral molecules. Chiral molecules can exist as two isomers that have low degrees of symmetry. Proteins consist of many chiral subunits thus producing CD signals. The CD signals that arise from proteins are a result of cysteine, tryptophan, tyrosine and phenylalanine residues and peptide bonds. The CD spectra between 260-180 nm is specific for peptide bond conformation and allows identification of different secondary structures, such as alpha helices, beta sheets, random coils and turns.

CD measurements were performed on a Jasco J815 spectropolarimeter using a thermostatically controlled 0.2 cm pathlength quartz cuvette. Far UV spectra were recorded from 190 to 300 nm at a constant temperature of 25 °C. Each spectrum was an average of 16 accumulations and the signal was averaged over 15 seconds.

Thermal melts

CD measurements were performed on a Jasco J815 spectropolarimeter using a thermostatically controlled 0.2 cm pathlength quartz cuvette. The change in dichroic signal was monitored at a wavelength of 222 nm while the temperature was increased from 20 to 90 °C at a constant rate of 1 °C/min.

As Tris-HCl is an unsuitable buffer for CD measurements, samples were dialysed against the elution phosphate buffer, pH 8.0 and all CD measurements were performed immediately after protein elution from the column. A buffer only background CD trace was measured at the beginning of each CD spectra measurement and this was removed from the trace of each sample.

2.2.13.2 Gel Filtration Chromatography

Gel filtration chromatography (also known as size exclusion chromatography) separates molecules according to size using porous particles. Small molecules can enter the pores of the particles so have a longer path before they are eluted. Larger molecules however, cannot enter the porous particles, thus have a shorter path length and shorter transit time. Size exclusion columns are generally used to determine the molecular weight (M_r) of a molecule.

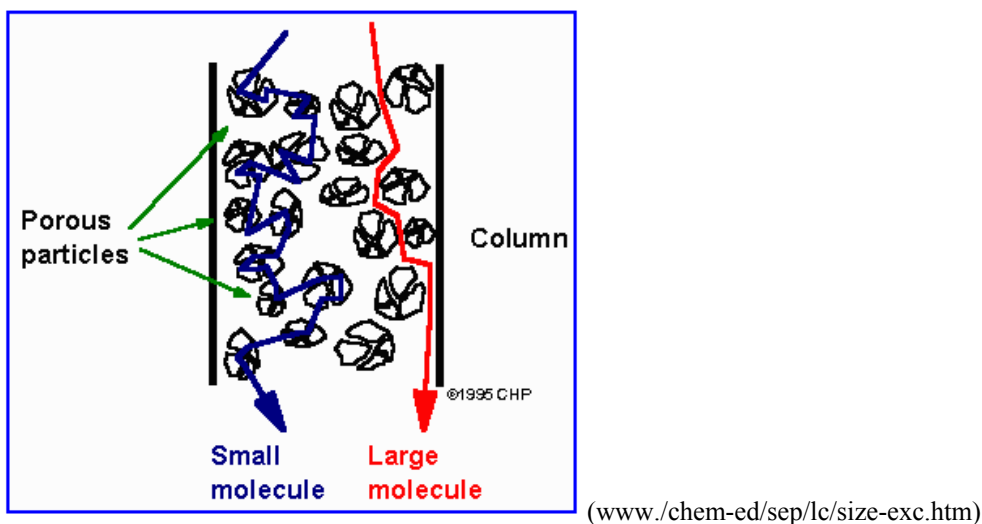


Figure 2.11. A schematic diagram of a size exclusion chromatography column.

Molecules that are smaller than the size of the pore are retained and elute last whereas larger molecules cannot enter the porous particles and are eluted first.

Gel filtration experiments were carried out using a Superdex 200 HiLoad 16/60 column with a bed volume of 120 ml. The flow rate was 1 ml per minute and the pressure limit was 0.3 MPa. The column was equilibrated with Tris-HCl buffer (pH 8.0) prior to the loading of 2 ml of protein onto the column. The absorbance of the column eluate was recorded at 280 nm. A solution of molecular weight markers, comprising of blue dextran (2000 kDa), ferritin (440 kDa), immunoglobulin G (160 kDa), bovine serum albumin (67 kDa), ovalbumin (45 kDa), myoglobin (16.7 kDa), ribonuclease A (13 kDa) and vitamin B₁₂ (1.355 kDa) was also run through the column. Data from these MW markers was used to plot a calibration curve, which was then used to perform linear regression in order to establish the MW of the elution peaks.

2.2.13.3 Analytical Ultracentrifugation (AUC)

This technique allows characterisation of key sample properties of biomolecules in their native state under biologically relevant solution conditions. The experiments are performed in

free solution, thus there are no complications caused by interactions with matrices or surfaces that can obscure interpretation.

Sedimentation velocity experiments allow interpretation of the entire time-course of sedimentation of a sample. It is a hydrodynamic technique that provides information about the shape and molecular mass of the macromolecules in addition to their size distribution. An initially uniform solution is placed in a cell and a moving boundary is formed on application of a strong centrifugal force which causes rapid sedimentation of solute towards the cell bottom. This results in a depletion of solute near the meniscus and the formation of a sharp boundary between the depleted region and the uniform concentration of sedimenting solute (the plateau). A series of scans are recorded at regular intervals to determine the rate of movement and broadening of the boundary as a function of time. The rate of movement of the boundary is measured to determine sedimentation coefficients, which depend directly on the mass of the particles in the sample.

Following elution of protein from the gelatin column, all samples were dialysed against 50 mM Tris-HCl, pH 8.0 prior to analysis by AUC. Sedimentation velocity measurements of Hsp47 samples were made using a Beckman (Paolo Alto, US) proteomelab XLI analytical ultracentrifuge. Samples were run using two sector cells in an Xti rotor. Samples were centrifuged at 40,000 rpm for 8 hours at 4 °C. The protein content in the cell was measured by continuously monitoring the protein absorbance at 280 nm. SEDFIT (sedfitsedphat.nibib.nih.gov) was used to analyse the data and obtain a sedimentation distribution. All results were corrected to the viscosity and density of Hsp47.

2.2.13.4 pH Titrations

pH titrations were carried out on a Jasco J815 connected to an autotitrator. A stirred, thermostatically controlled 1 cm quartz cuvette was used. All mHsp47 constructs were studied at a final protein concentration of 0.15 mg/ml in phosphate buffer, pH 8.0 at 25 °C. The pH was then progressively decreased to approximately pH 4.0 by the addition of 2.5 µl of 1.52 M phosphoric acid. The change in dichroic signal was measured at 222 nm with an emission slit width of 3 nm. Spectra for each variant were measured before and after the titration to identify any structural variation following a reduction in the pH.

2.2.13.5 Collagen Fibrillisation Assay

Collagen used in this study was isolated from fresh calf Achilles tendons. First, the tendons were rinsed with acetone to remove the fatty deposits. Subsequently, the collagen was extracted by stirring with 0.5 M acetic acid at 4 °C for 48 hours and then salting out with sodium chloride. The sample was further dialysed against deionised water at 4 °C for 48 hours. The collagen was then freeze-dried and kept in the refrigerator.

Before use, collagen from calf Achilles tendons was dissolved in 75 mM sodium citrate pH 3.6-4.0 to a final concentration of 1 mg/ml. This solution was brought to a pH of 7.0 and to a final protein concentration of 0.2 mg/ml by the addition of 20 mM sodium phosphate buffer, pH 7.4, containing 50 mM NaCl. The protocol used for collagen fiber formation was similar to that described by Williams and coworkers [185]. Turbidity measurements were made at 313 nm in a spectrophotometer over a period of 320 minutes at a constant temperature of 34 °C, which was maintained by a controlled water bath. A 1 cm quartz cuvette was used.

2.2.13.6 Collagen Binding Assay

A collagen-binding assay was performed to determine whether *Xenopus laevis* and *Danio rerio* Hsp47 bound to collagen immobilised on agarose beads. A control resin, containing no collagen was used to determine non-specific binding.

Preparation of resin

The collagen resin was prepared by spinning 500 µl of resin at 14,000 x g in a centrifuge for 1 minute. The supernatant was discarded and the pellet resuspended in 1 ml Tris-HCl at pH 8.0; this was repeated three times. Finally, the resulting beads were resuspended in 500 µl of Tris pH 8.0 to complete the preparation of the collagen resin.

The control resin was prepared by resuspending sephadex in Tris-HCl buffer at pH 8.0. The solution was spun until all the sephadex had resuspended and could now be used for the binding assay.

The binding assay procedure

0.5 ml of the Hsp47 was thawed and spun at 14,000 x g for 15 minutes in order to remove any aggregated insoluble protein. The supernatant resulting from the spin was transferred to a fresh tube and stored on ice. A mix containing 50 µl Hsp47, 50 µl collagen agarose beads and 50 µl Tris-HCl pH 8.0 buffer (100 mM Tris-HCl, 150 mM NaCl) was incubated on ice for 5 minutes whilst turning. It was then spun for 1 minute at 14,000 x g, the supernatant was retained in a microcentrifuge tube for analysis on SDS-PAGE whilst the beads were resuspended in 1 ml Tris buffer (100 mM Tris-HCl, 150 mM NaCl, pH 8.0). The samples were then spun at 14,000 x g for 1 minute and the supernatant discarded. The beads were resuspended in the Tris buffer and spun twice more. After the final spin, the supernatant was

discarded and 50 µl of loading buffer added to the resin remaining in the microcentrifuge tube and the previously retained supernatant sample. The samples were then boiled and run on SDS-PAGE.

The same procedure was used to perform the binding assay using the control resin, however instead of collagen-agarose beads; the pre-prepared sephadex resin was used.

2.2.13.7 Electron Microscopy

Transmission electron microscopy (TEM) was used to visualise collagen samples pre- and post-fibrillation. This allowed the comparison of the surface features and the size and shape of collagen fibres before and after fibrillation and furthermore, in the absence and presence of Hsp47.

TEM permits determination of the topography and morphology of a sample. It involves transmitting a beam of electrons through an ultra thin specimen using a positive electrical potential. Focussing of the beam is by a magnetic lens. Interaction of electrons with the specimen forms an image that is magnified onto an imaging device, such a fluorescent screen. Thus, the interactions and effects are detected and transformed into an image.

Negative staining was used as a contrast enhancement method that allows the specimen to appear translucent against a dark background.

Negative Staining Protocol

A carbon coated grid was discharged just before use to increase hydrophilicity. 5 µl of sample was placed on the grid and left to dry for 30 seconds. The grid was blotted with filter paper. 30 µl of 2% uranyl acetate stain were pipetted onto the grid and immediately blotted using

filter paper. This was repeated for a further four drops of uranyl acetate ensuring that drying between drops was prevented. A further 30 μ l of uranyl acetate were pipetted onto the grid and this time was allowed to soak for 20 seconds prior to blotting. This was repeated for a further 30 μ l of uranyl acetate. The grid was allowed to dry for a few minutes before imaging.

Imaging used a JEOL 2011 microscope running at 200 KV with a La6B filament. A defocus of 3-6 microns was used to improve the visual performance.

CHAPTER 3: Purification and Characterisation of *Xenopus laevis* and *Danio rerio* Hsp47

3.1 Introduction

Work done previously on Hsp47 has focused mainly on murine Hsp47. However, this form is relatively insoluble and unstable in solution making it difficult to work with. Therefore, an alternative strategy using *Xenopus laevis* and *Danio rerio* Hsp47 was developed. Consequently, the *Danio rerio* and *Xenopus laevis* gene for Hsp47 was used to try to purify larger quantities of protein that would allow more detailed structural studies from which molecular information could be elucidated. This has the advantage of maintaining function while altering DNA sequence. It was hoped that these new sequences would express better and/or produce a more stable protein.

Earlier studies used the *E. coli* strain DH1 with λ hind⁻ prophage for the cloning and overexpression of the mature mouse Hsp47 [186]. The same group attempted to clone the cDNA of Hsp47 into the HB101 strain with pRH1 or pCNR1 plasmids. This proved unsuccessful as it yielded low levels of expression [187]. To overcome this problem, the pKS26 vector, possessing a P_L promoter for high expression levels, was selected. This plasmid allowed overexpression of Hsp47 to 40-50% of the total protein in the cell. Following the transformation of DH1 cells, expression of recombinant Hsp47 involved growing of the plasmid overnight at 37 °C in LB containing ampicillin. This overnight culture was used to inoculate a culture at 1% (v/v). The cells were heat induced at 42 °C, after reaching an OD_{600nm} of 0.5-0.8, by increasing the temperature in the incubator. The cells were grown at this temperature for a further 1 hour and 30 minutes. Western blot analysis using a monoclonal antibody to Hsp47 confirmed expression of the protein [187]. Hsp47 studies in our lab however, have found that Hsp47 is a rather unstable and fragile protein and such an

increase in temperature results in the majority of the protein being insoluble and present in inclusion bodies or aggregates. Furthermore, even a reduction in temperature following induction with IPTG resulted in very small amounts of soluble protein.

It was therefore decided that the Hsp47 gene from different organisms would be used to attempt to obtain larger amounts of pure, soluble protein. As the sequence for both *Xenopus laevis* and *Danio rerio* Hsp47 is slightly different to that of *Mus musculus* Hsp47, these organisms were considered favourable for the above objective. Amino acid sequences from Genbank were used to calculate the sequence identities between various organisms. *Xenopus laevis* has 73% and 70% sequence homology with *Mus musculus* and *Danio rerio*, respectively. Most of the sequence identity was focused at the centre of the sequence and at a few carboxyl terminal regions. This sequence identity is in part due to three key features that are conserved between Hsp47 from most species; a hydrophobic leader sequence, two potential glycosylation sites and the ER-retention signal.

Furthermore, *Xenopus laevis* is a poikilothermic vertebrate and *Danio rerio* is an ectothermic vertebrate in comparison to the mammalian endotherm, *Mus musculus*. As these organisms differ in the controlling of their body temperature it was thought that there may also be a difference in the stability and solubility of Hsp47 between ectothermic and endothermic organisms.

Some of the work detailed in this chapter was done in our lab prior to the start of this study. It is important to present all the work that was carried out in order to tell the complete story. Those experiments and results that did not involve me have been clearly stated.

This chapter will detail the outcome of cloning, expression and purification of *Xenopus laevis* and *Danio rerio* Hsp47 for each of the vectors used (for vector maps see Figures 2.1-2.4).

3.2 Results

3.2.1 Cloning into the vectors

The gene for Hsp47 from *Xenopus laevis* and *Danio rerio* was cloned into four different vectors, each possessing different tags that should aid purification (described in section 2.1).

3.2.1.1 Cloning the *Xenopus laevis* Hsp47 gene into pET102/D-TOPO[®]

PCR was carried out to introduce specific terminal sequences into the gene to allow cloning into the D-TOPO[®] vector (results shown in Figure 3.1 below). Subsequently, samples that showed the correct amplimer (1, 300 base pairs (bp)) were combined and purified to remove any impurities from the amplified genes.

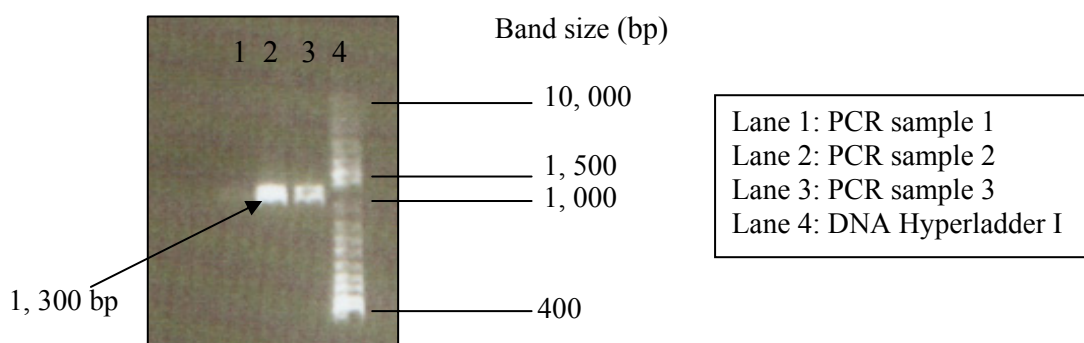


Figure 3.1. PCR result for *Xenopus laevis* Hsp47 gene. Lanes 1-3 showed bands at ~1, 300 bp, the size of the Hsp47 gene, illustrating successful PCR reactions.

Following purification, the amplimer was ligated into the pET102/D-TOPO[®] and transformed into Top10 cells. Colonies resulting from the overnight incubation of the transformation mix on LB agar plates (supplemented with 100 mg/ml ampicillin) were used to inoculate 5 ml LB comprising 100 mg/ml ampicillin. The presence of the correct size of plasmid DNA was

checked by performing mini-preps on the transformants (results not shown). Samples that revealed a band at 7615 bp (6315 bp for vector + 1300 for Hsp47 insert) were identified as successfully cloned constructs and were used to transform BL21 Star (DE3) One Shot Cells. The presence of insert was confirmed by DNA sequencing.

3.2.1.2 Cloning the *Xenopus laevis* Hsp47 gene into pET-32 Xa/LIC

Using the same protocol as for the pET102/D-TOPO[®] vector amplification of the Hsp47 gene was performed so that it could be used for insertion into the pET-32 Xa/LIC vector. Results from the PCR are illustrated in Figure 3.2 below.

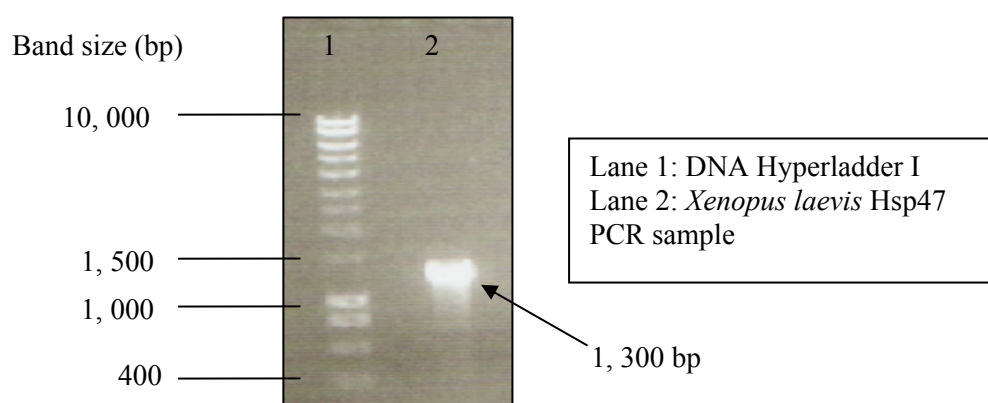


Figure 3.2. Results for PCR of *Xenopus laevis* Hsp47 gene. Successful amplification of the protein gene was illustrated by a single band at 1,300 bp.

Following purification of the DNA product, the target insert was treated with T4 DNA polymerase before annealing the vector and the Xa/LIC insert. Analysis of the mini-prep (shown in Figure 3.3 below) revealed that the pET-32 Xa/LIC vector had successfully taken up the Hsp47 gene.

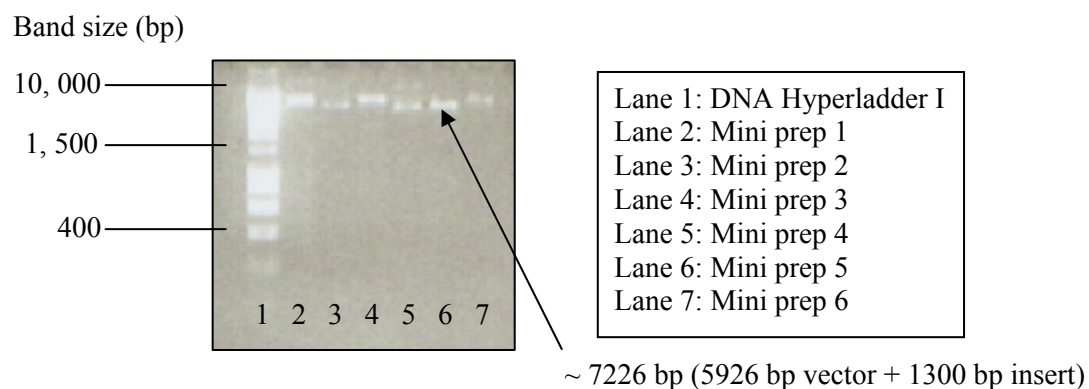


Figure 3.3. Results for mini-preps of pET-32 Xa/LIC containing *Xenopus laevis* Hsp47 insert.

The presence of bands at ~7226 bp indicates insertion of the gene into the vector.

One of the above plasmids was used to transform Top10 *E. coli* cells.

3.2.1.3 Cloning into pTYB12

The gene for *Xenopus laevis* Hsp47 and *Danio rerio* Hsp47 was amplified by PCR. The annealing time was increased to 45 seconds as the previously used annealing time of 30 seconds did not produce any PCR product. An annealing temperature of 40 °C in the PCR cycle produced successful gene amplification (Figure 3.4) for both the *Danio rerio* and *Xenopus laevis* Hsp47.

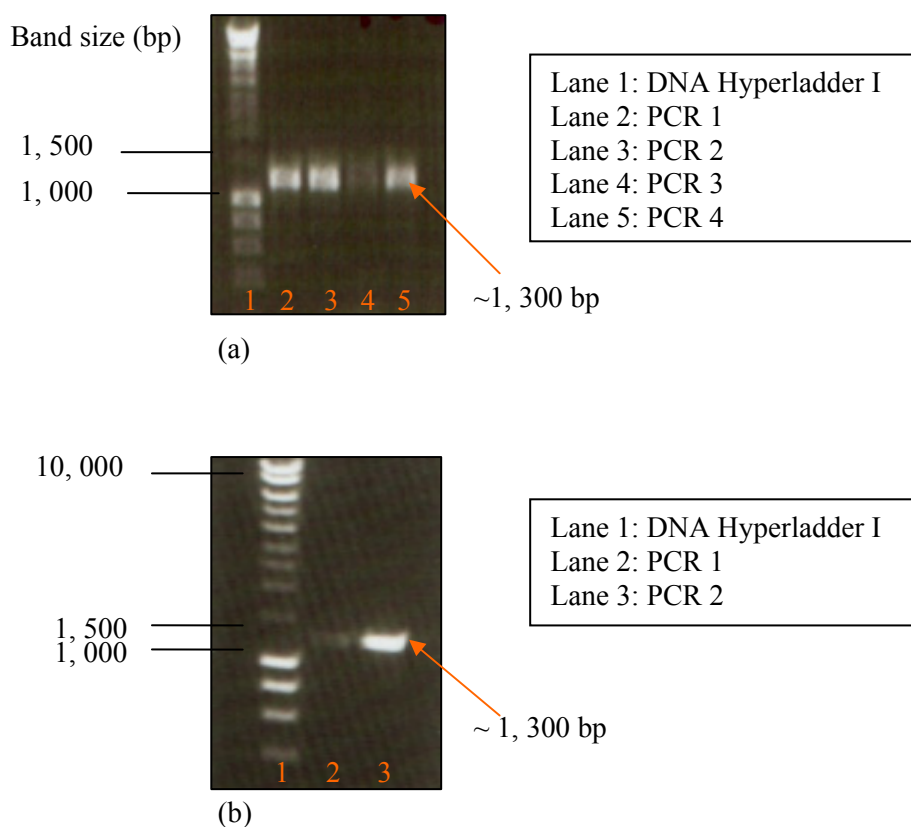


Figure 3.4. A 1% agarose gel of PCR results for (a) *Danio rerio* Hsp47 and (b) *Xenopus laevis* Hsp47.

The successful PCR reactions for each species were pooled, run on a gel and the band excised using QIAprep® MinElute Gel Extraction Kit. Presence of purified PCR product was checked by running the samples on a 1% agarose gel (results not shown). The purified *Danio rerio* Hsp47 gene was cloned into pET101/D-TOPO®, pET102/D-TOPO®, pET-32 Xa/LIC and pTYB12 and *Xenopus laevis* Hsp47 was inserted into pET102/D-TOPO®, pET-32 Xa/LIC and pTYB12. Consequently, successfully clones were used to transform BL21 star (DE3) One Shot Cells prior to protein expression determination (method described in section 2.8).

Previous to this study, the Hsp47 gene from *Danio rerio* had been successfully cloned into pET102/D-TOPO® and pET-32 Xa/LIC and these constructs were used for the characterisation of *Danio rerio*.

3.2.2 Expression of the protein from the Vectors

3.2.2.1 Expression of Hsp47 from *Xenopus laevis*

from pET102/D-TOPO[®]

The optimum expression conditions for *Xenopus laevis* Hsp47 from pET102/D-TOPO[®] were distinguished by growing expression cultures at different temperatures. Samples were taken at 1 hour intervals for 5 hours and the turbidity measured. These samples were analysed by SDS-PAGE.

Analysis of the samples using SDS-PAGE and coomassie brilliant blue staining showed that no Hsp47 was present in the culture grown at 15 °C, as there were no distinct bands for Hsp47 (Figure 3.5).

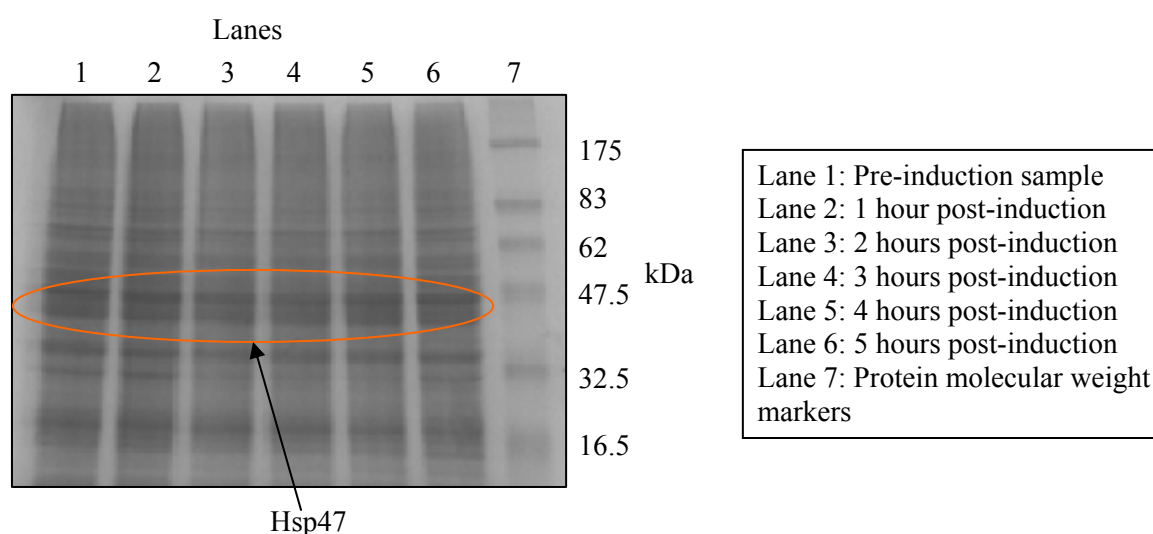


Figure 3.5. SDS-PAGE of *Xenopus laevis* Hsp47 expression at 15 °C. The lack of a distinct 47 kDa band following induction implied that Hsp47 was not being expressed.

An incubation temperature of 20 °C produced results (Figure 3.6) different to those given by an incubation temperature of 15 °C. Protein expression occurred but it was not enhanced by induction with IPTG. Western blot analysis was used to confirm that the protein was Hsp47.

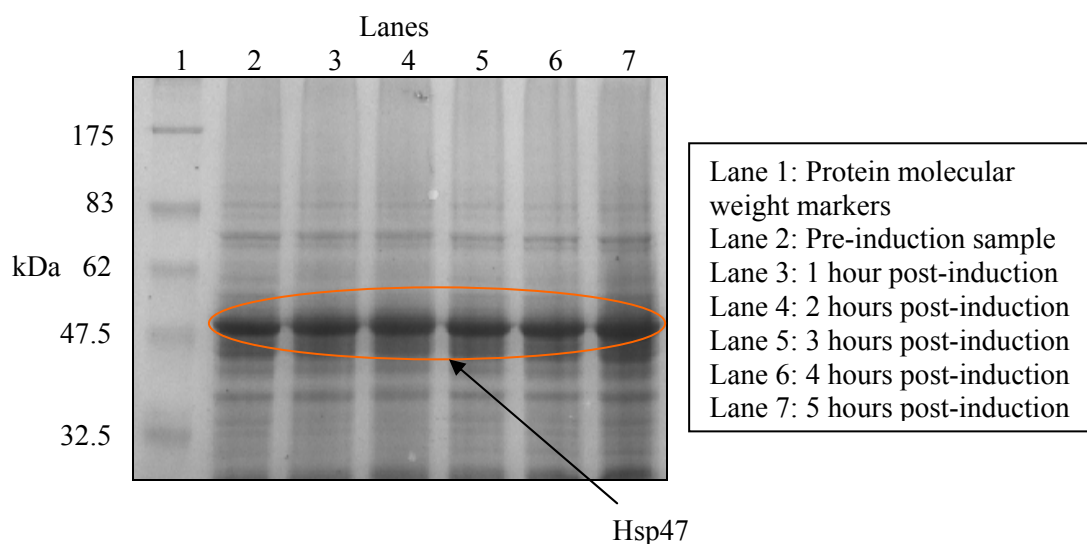


Figure 3.6. SDS-PAGE of *Xenopus laevis* Hsp47 expression at 20 °C. The gel illustrates that the intensity of the bands corresponding to Hsp47 is very similar pre- and post-induction. This implies that protein expression was not enhanced by induction with IPTG.

Subsequently, the protein culture was incubated at 25 °C. The turbidity of the culture was monitored by taking a 1 ml sample from the culture and measuring its absorbance. Once an OD_{600nm} of 0.6 was reached, the culture was induced with IPTG. The absorbance was monitored thereafter until a plateau was reached, indicating that cells could be harvested. The absorbance measurements were used to plot a growth curve (Figure 3.7) to illustrate the absorbance after induction.

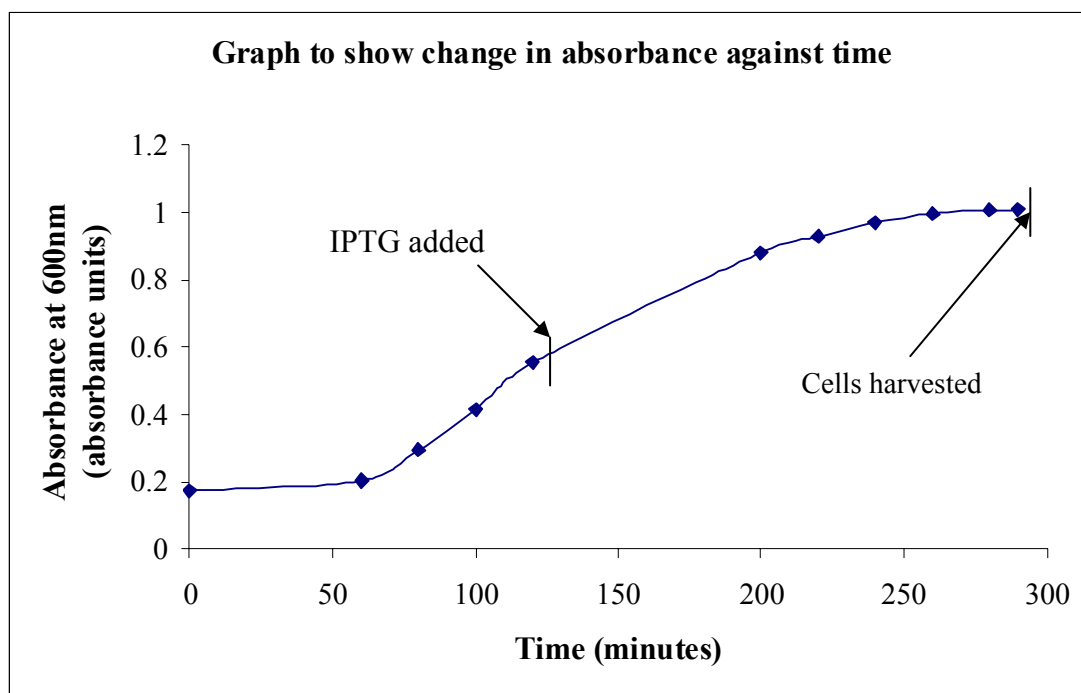


Figure 3.7. A growth curve of a protein culture at 25 °C. There was a steady increase in absorbance that continued after induction, followed by a deceleration phase.

Analysis of the gel for the culture at 25 °C demonstrated that there was no Hsp47 present in the pre-induced sample (as expected) and the amount of Hsp47 produced post-induction gradually increased with time (Figure 3.8).

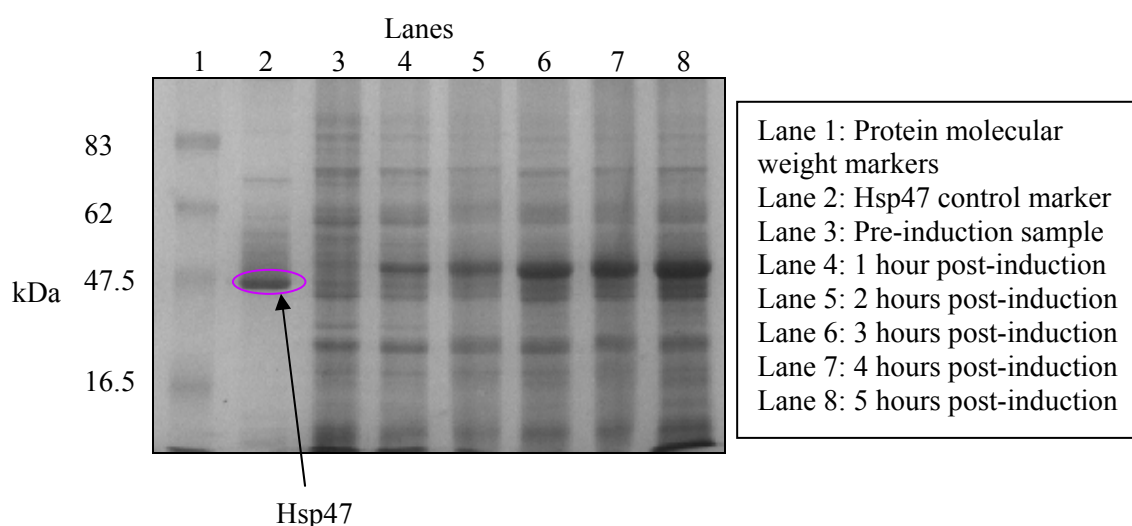


Figure 3.8. SDS-PAGE analysis of *Xenopus laevis* Hsp47 expression at an incubation

temperature of 25 °C. The results reveal that the expression of the protein ~47 kDa increased with time.

The size of the major protein band in the post-induced samples (lanes 4-8) represents Hsp47 with a HP-thioredoxin tag. A Western blot (not shown) confirmed that the bands represented Hsp47. The Hsp47 band became more distinct at 3 hours, indicating an increase in protein expression; therefore this was decided as the optimum time for induction.

Once the optimum incubation conditions had been established these conditions were used to check the expression of Hsp47 *Xenopus laevis* in pET102/D-TOPO[®]. SDS-PAGE (Figure 3.9) revealed that a large band appeared for all the 5 ml cultures that were induced, confirming that over expression of the protein was taking place.

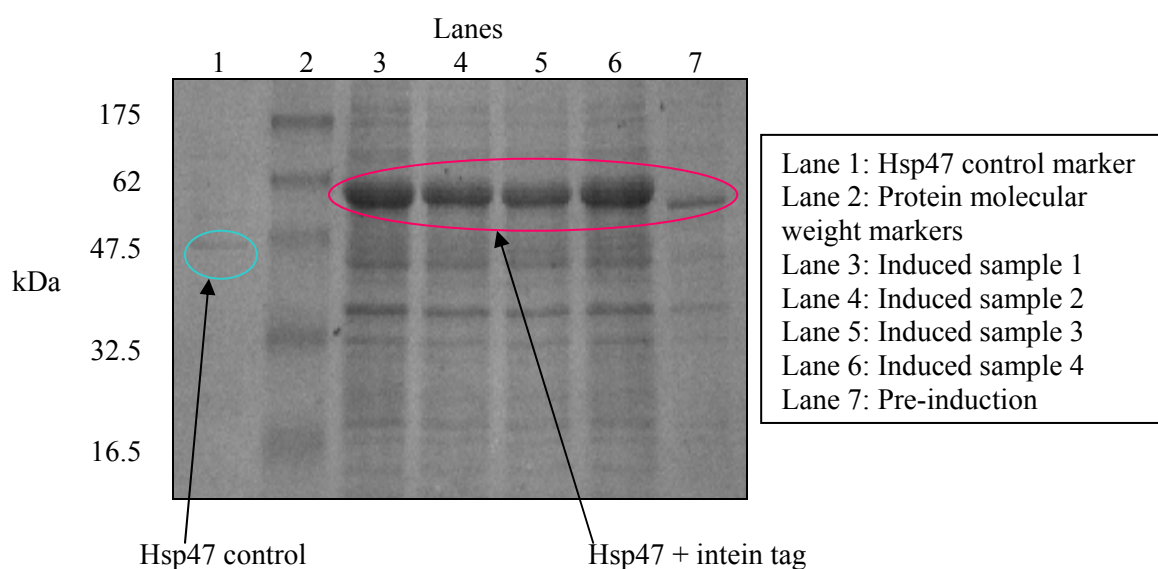


Figure 3.9. SDS-PAGE of *Xenopus laevis* Hsp47 expression by pET102/D-TOPO[®]. The bands at ~62 kDa represented Hsp47 with an attached 16 kDa HP-thioredoxin tag. Induced samples 1-4 are repeats of the same expression experiment to ensure reproducibility of protein expression.

The uninduced and induced samples contained a protein with a molecular weight of approximately 62 kDa. Presence of protein in the uninduced fraction is due to leaky expression. This was consistent with the combined molecular weight of Hsp47 and the HP-thioredoxin tag (16 kDa) in the vector. The additional mass reduced protein mobility, resulting in slower movement along the gel. Due to the large size of the band, it was

estimated that approximately 70% of the protein being expressed was Hsp47. The enhanced protein expression was thought to be due to the presence of the HP thioredoxin.

from pET-32 Xa/LIC

The attempted cleavage of the thioredoxin using enterokinase was unsuccessful and so the gene was cloned into pET-32 Xa/LIC. This vector comprises a thrombin and factor Xa sites for the cleavage of the thioredoxin tag. It was thought that the presence of two possible cleavage sites would increase the chance of successfully removing the fusion tag.

Successful recombinants of vector containing insert were used to transform BL21 star (DE3) One Shot Cells. Presence of multiple colonies indicated that the transformation was successful.

As before, overnight cultures were grown and induced, cells were harvested and centrifuged. SDS-PAGE (Figure 3.10) was used to analyse the samples.

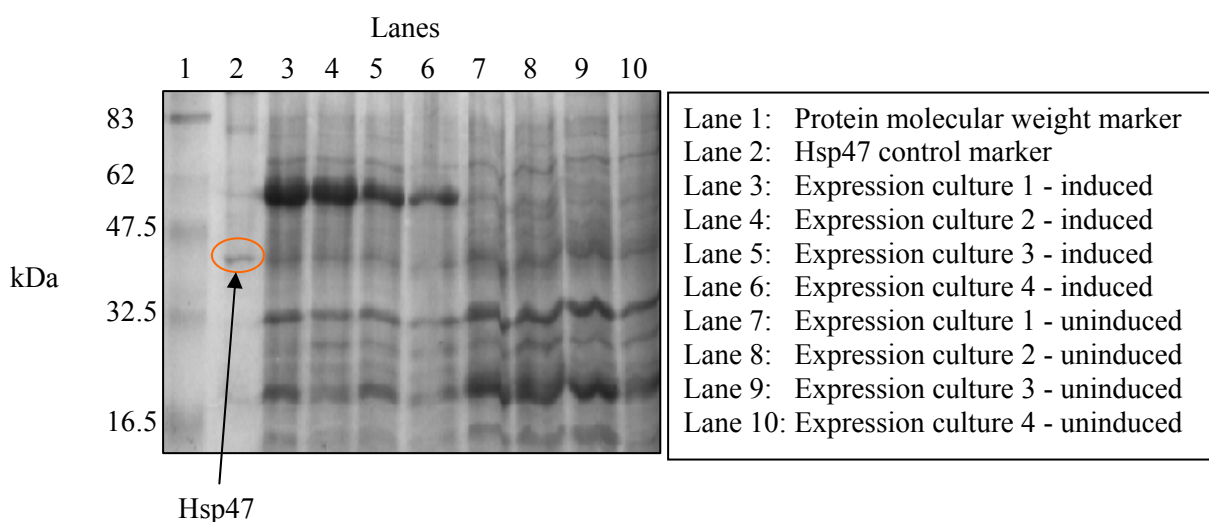


Figure 3.10. SDS-PAGE of *Xenopus laevis* Hsp47 expression in pET-32 Xa/LIC. The protein bands at ~62 kDa represent Hsp47 with a thioredoxin tag. These bands are not present in the uninduced samples (lanes 7-10) but are clearly visible in the induced samples (lanes 3-6). Expression cultures 1-4 represent repeats that were performed under similar conditions to ensure reproducibility.

SDS-PAGE analysis revealed that the induced samples showed successful expression of Hsp47 whereas the uninduced samples did not show this expression. A Western blot confirmed that the over-expressed bands corresponded to Hsp47. The increased molecular weight was due to the presence of the HP-thioredoxin in the construct.

from pTYB12

5 ml overnight cultures of colonies from the transformations were set up. The following day a 1 ml sample was diluted to 3 ml and left to grow until an OD_{600nm} of approximately 0.6 was reached. 1 mM IPTG was then added to induce protein expression and the culture was induced for a further 4 hours at 37 °C. After this period the cells were harvested by centrifugation and analysed by SDS-PAGE (Figure 3.11).

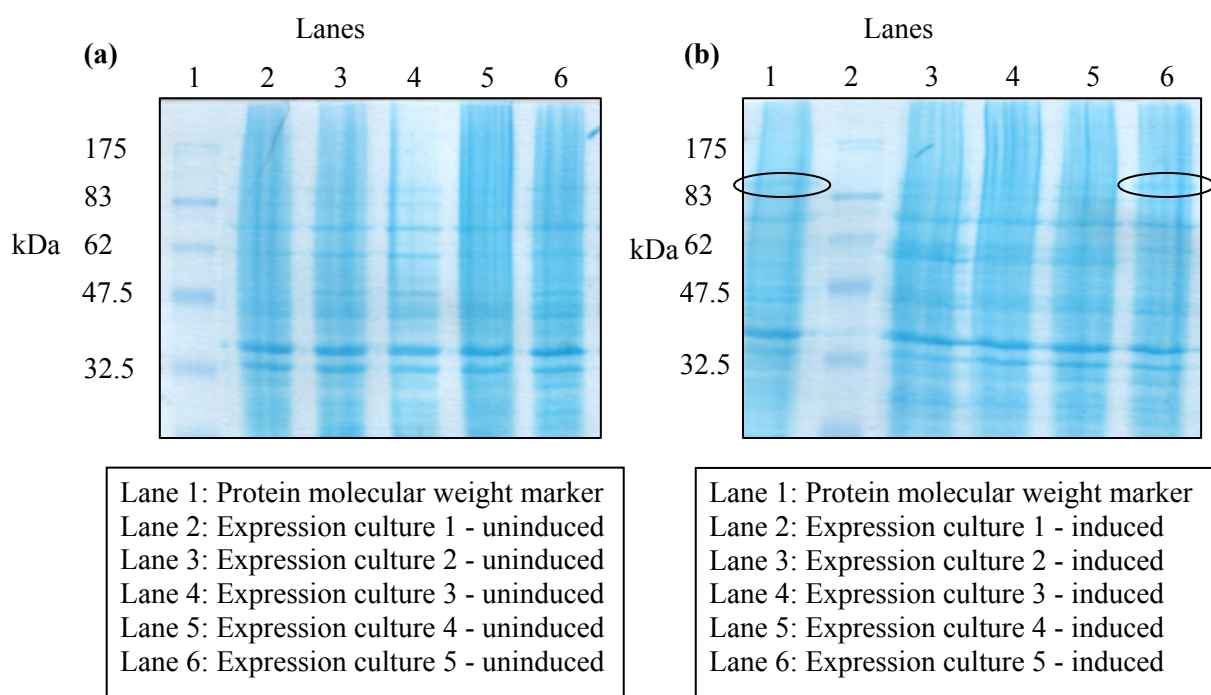


Figure 3.11. SDS-PAGE of *Xenopus laevis* Hsp47 expression by pTYB12. The molecular weight of Hsp47 with an intein tag (55 kDa) was 102 kDa. (a) The uninduced samples did not show distinct bands at ~100 kDa. However, expression cultures 2, 3 and 5 (lanes 3, 4 and 6, respectively) revealed faint bands at this position. (b) The induced fractions were not distinctly different from the uninduced fractions. Closer examination of the gels revealed that the induced fraction of expression culture 1 produced a band at ~100 kDa, which was absent in the uninduced fraction. The ~100 kDa band in

expression culture 5 was more concentrated than the ~100 kDa band of the uninduced fraction of expression culture 5. Expression cultures 1-5 represent experiments that were performed under similar conditions to monitor the reproducibility of protein expression.

The gels demonstrated a very subtle change in protein concentration between the uninduced and induced fractions. Only a very slight difference could be seen between the pre- and post-induced samples of expression cultures 1 and 5. A Western blot was performed and a band was seen at the 47 kDa position, confirming that the protein band did in fact correspond to Hsp47.

In spite of low basal and induced expression levels, cultures that showed successful expression of Hsp47 will be used for protein purification as low induced expression levels are a common factor in protein expression.

In summary, the results from the expression experiments demonstrated that *Xenopus laevis* Hsp47 was successfully expressed in pET102/D-TOPO[®], pET-32 Xa/LIC and pTYB12 and the protein was soluble in all constructs. The expression of the protein cloned into pET101/D-TOPO[®] however was unsuccessful.

3.2.2.2 Expression of Danio rerio Hsp47

from pET101/D-TOPO[®]

The recombinant plasmid was used to transform BL21 star (DE3) One Shot Cells (as described in section 2.7). Subsequently the expression of Hsp47 in the transformed cells to was tested (as described in section 2.9). These cultures however, failed to express protein and so were not studied further.

from pET102/D-TOPO[®]

As before, transformation of the BL21 star (DE3) One Shot Cells was followed by assessing the expression level of the protein. The results of the SDS-PAGE showed that thioredoxin tagged Hsp47 was being expressed.

from pET-32 Xa/LIC

The Hsp47 gene from *Danio rerio* was successfully cloned into the pET-32 Xa/LIC vector and the construct used to transform BL21 star (DE3) One Shot Cells. Successful transformations enabled expression levels to be determined. Samples analysed by SDS-PAGE revealed that expression of the protein was taking place after induction by IPTG.

Cloning of *Danio rerio* Hsp47 into pET101/D-TOPO[®], pET102/D-TOPO[®], pET-32 Xa/LIC were performed in our lab prior to this study but did not involve me. From this point onwards, all work, whether *Xenopus laevis* or *Danio rerio* Hsp47 was performed by me.

from pTYB12

Expression of Hsp47 was checked following successful transformation into BL21 star (DE3) One Shot Cells. This involved inoculation of a culture of LB containing ampicillin followed by induction of protein expression with IPTG (see section 2.8). Samples were collected pre- and post- induction to determine whether Hsp47 was being produced.

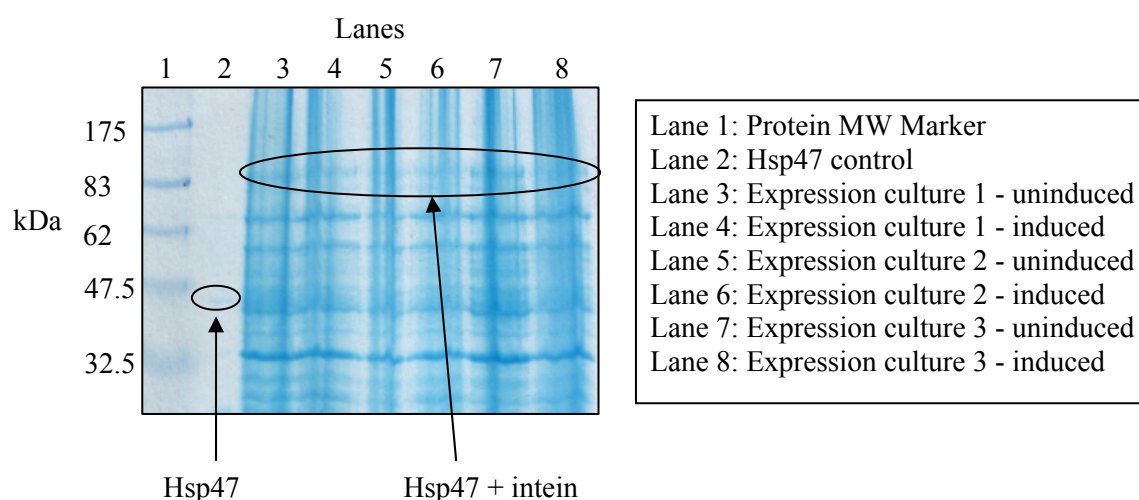


Figure 3.12. SDS-PAGE of *Danio rerio* Hsp47 expression in pTYB12. Lanes 3, 5 and 7 show uninduced fractions whilst lanes 4, 6 and 8 represent induced fractions. Expression cultures 1-3 are repeats of the expression experiment under similar conditions to check the reproducibility of protein expression under the specified conditions.

The SDS-PAGE results revealed that the pre- and post- induction samples showed similar levels of protein expression as the bands that appear in the induced fractions (lanes 4, 6 and 8) were also present in the uninduced fractions (lanes 3, 5 and 7). The size of Hsp47 with the intein tag (55 kDa) is approximately 100 kDa. A Western blot was carried out to determine whether the bands at ~100 kDa represented Hsp47 (Figure 3.13). The results however were negative, showing no bands with a molecular weight of 100 kDa, suggesting that Hsp47 was not being produced by this construct.

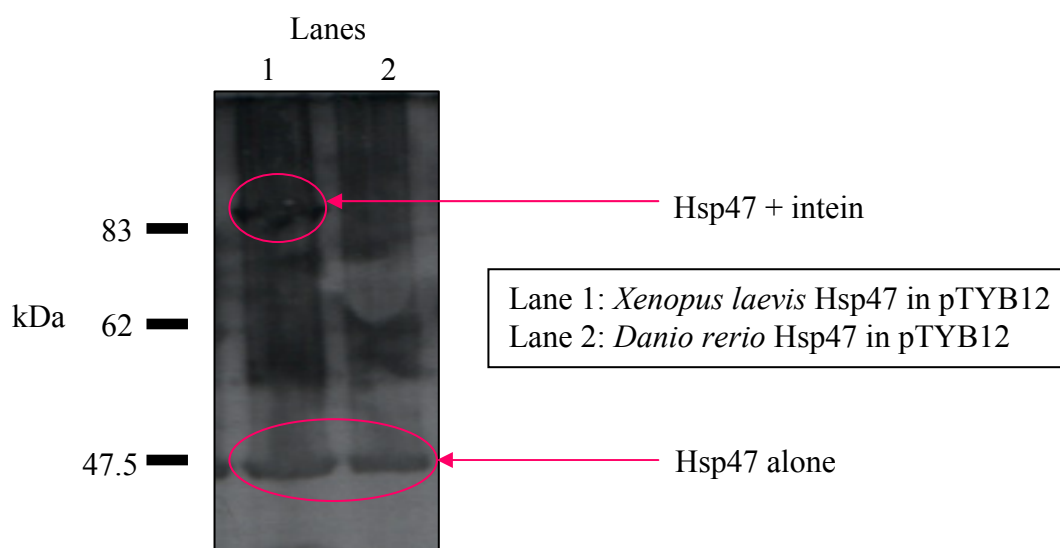


Figure 3.13. A Western blot of *Danio rerio* and *Xenopus laevis* Hsp47 in pTYB12. The bands at 47 kDa in lanes 1 and 2 denote that both *Xenopus laevis* and *Danio rerio* expressed Hsp47 without the intein tag. Only *Xenopus laevis* however, produced a band at ~100 kDa, representing Hsp47 with attached intein tag (lane 1).

The absence of a band at ~100 kDa (Hsp47 + intein) and the presence of a band at 47 kDa (Hsp47 - intein) in the *Danio rerio* sample suggested that the construct was not expressing intein tagged Hsp47 but was expressing non-tagged Hsp47.

3.2.3 Protein Purification

Purification of Hsp47 from pET102/D-TOPO[®] and pET-32 Xa/LIC involved Ni-NTA immobilised on beads whilst protein produced by the pTYB12 construct involved purification using chitin resin.

3.2.3.1 Purification of Hsp47 from *Xenopus laevis* in pET102/D-TOPO[®]

As the construct cloned into the pET102/D-TOPO[®] vector resulted in substantial expression of protein, it was used to attempt to purify the Hsp47. The presence of both HP-thioredoxin and a C-terminal His tag allowed purification using Ni-NTA affinity chromatography.

Samples were collected throughout protein purification for analysis on SDS-PAGE (Figure 3.14).

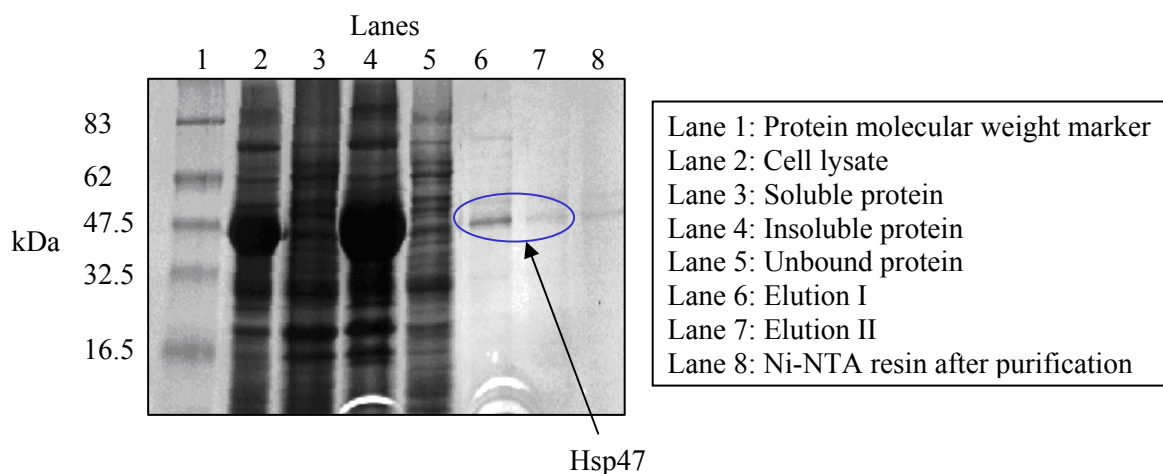


Figure 3.14. SDS-PAGE for purification of *Xenopus laevis* Hsp47 in pET102/D-TOPO[®] construct. The presence of a very intense band at ~47 kDa band in lane 2 suggested that the majority of the protein being produced was Hsp47. The protein in lanes 3 and 4 implied that of the protein being produced, most of this was insoluble. Lanes 6 and 7 demonstrated pure, soluble protein.

The results revealed that although this construct produced large amounts of Hsp47, the majority of this was insoluble (Lane 4) and was probably in inclusion bodies. The lack of a 47 kDa distinct band in lane 5 implied that the majority of the soluble Hsp47 bound to the Ni-NTA for purification via metal affinity chromatography. The bands at 47 kDa in lanes 6 and 7 (elution I and II, respectively) illustrated that the purification of the little soluble Hsp47 was successful and pure, soluble *Xenopus laevis* Hsp47 was obtained. The fractions of pure protein were stored at -70 °C for structural and biochemical analysis.

A Western blot (Figure 3.15) was carried out to confirm the findings of the SDS-PAGE. As this technique is more sensitive and used antibodies specific for Hsp47, any protein that showed up was essentially Hsp47.

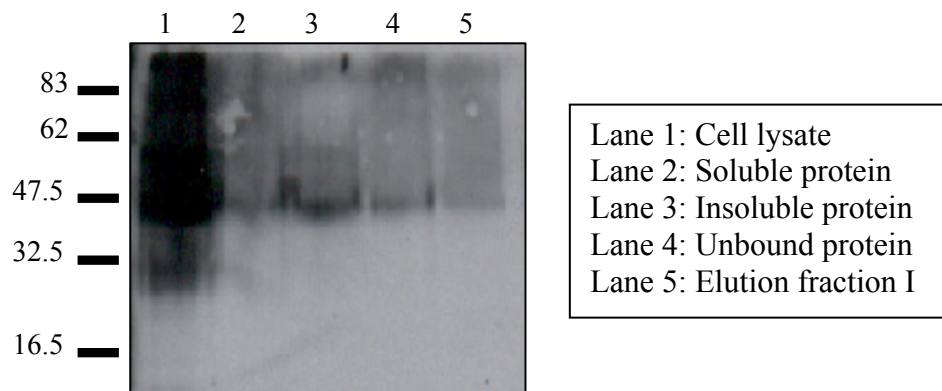


Figure 3.15. A Western blot of purification of *Xenopus laevis* Hsp47 in pET102/D-TOPO[®] construct. An over-expressed band at 47 kDa in the cell lysate fraction (lane 1) confirmed that the majority of the protein being produced was Hsp47. Lanes 2 and 3 suggest that there was no soluble Hsp47 present but a small amount of insoluble protein was produced. The absence of a 47 kDa band in lanes 5 implied that no pure Hsp47 was eluted.

The above Western blot was not very clear as there was a lot of background noise and the antibody did not seem to react well to the protein. A monoclonal anti-Hsp47 mouse primary antibody was used for all Western blots. It may be that the mouse antibody does not react well to Hsp47 from other organisms, resulting in reduced sensitivity and therefore compromised results.

The results of the Western blot (Figure 3.15) told a different story from the results of the SDS-PAGE (Figure 3.14). Similarly, both the SDS-PAGE and Western blot showed that there was a large amount of Hsp47 being produced. However, whereas the SDS-PAGE illustrated that there was an elution band present at 47 kDa the Western blot did not show a band in this fraction, suggesting that this was not the case. Also, the amount of insoluble Hsp47 was considerably lower in the Western blot in comparison to the SDS-PAGE. Furthermore, the Western blot did not reveal a distinct band in the elution lane (Figure 3.15, lane 5); implying that purification of soluble *Xenopus laevis* Hsp47 using Ni-NTA was unsuccessful.

Purification of Hsp47 in the **insoluble fraction** using pET102/D-TOPO[®]

Purification of the protein in the insoluble fraction was attempted in order to increase the yields of pure Hsp47.

Unfolding the Hsp47

In order to obtain a larger amount of soluble *Xenopus laevis* Hsp47, the insoluble form of the protein was unfolded using 8 M urea. It was thought that the His tag may be buried within the insoluble protein, unable to interact with the Ni²⁺ ions of the Ni-NTA resin, thereby preventing protein purification.

Unfolding of the protein involved resuspension of the cell debris (insoluble protein) fraction in a buffer comprising of 8 M urea. The wash and elution buffers used during protein purification by Ni-NTA also contained urea (see section 2.2.8 for a more detailed method).

SDS-PAGE analysis was used to determine protein solubility following protein unfolding.

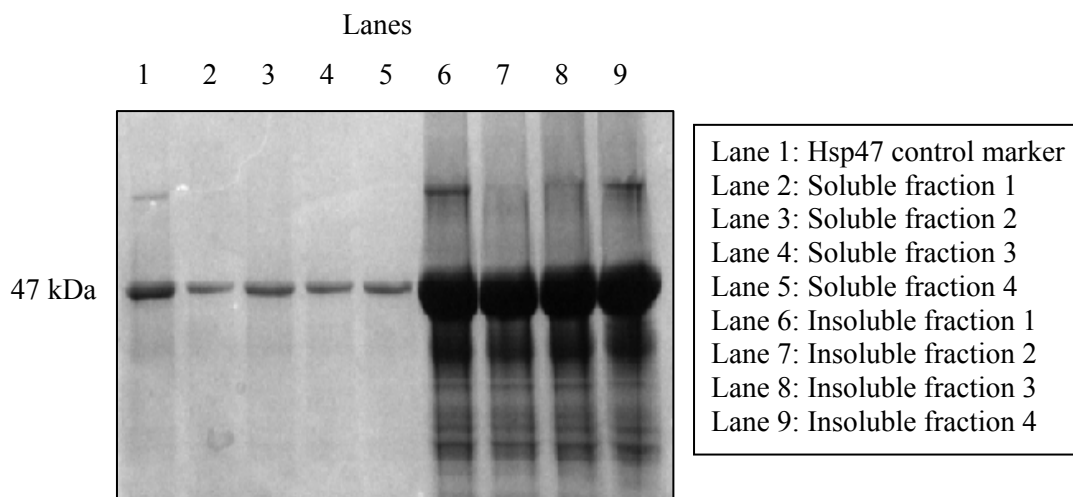


Figure 3.16. SDS-PAGE of unfolding of insoluble *Xenopus laevis* Hsp47. The presence of bands at 47 kDa in lanes 2-5 suggested that some of the insoluble protein had been unfolded and was now soluble. The large bands at 47 kDa in lanes 6-9 illustrate that the majority of the protein was still

insoluble, however. Fractions 1-4 represent repeats of the unfolding experiment from different expression cultures.

The results from the unfolding experiment illustrated that the use of urea in the purification buffers converted some of the insoluble material into a soluble form.

Refolding Hsp47

Dialysis was used to remove the urea from the unfolded protein and consequently allow protein refolding. Four experiments were set up to determine the refolding conditions. Each experiment consisted of 1 ml of unfolded protein containing 8 M urea being placed into buffers of differing urea concentrations for various durations. The protein to be refolded was placed in dialysis tubing with a molecular weight cut off of 12,000-14, 000 Da prior to dialysis.

The experiments were set up as follows:

EXPERIMENT	UREA CONCENTRATION IN BUFFER			
	6M	4M	2M	0M
1	2 hours →	2 hours →	2 hours →	Overnight →
2	4 hours →	2 hours →	Overnight →	
3	6 hours →	Overnight →		
4	Overnight →			

Experiment 1, for example, started with dialysis of the protein in 6 M urea for two hours, followed by dialysis in 4 M urea for two hours and then 2 two hours in 2 M urea before dialysing it in water overnight.

SDS-PAGE was used to analyse the protein samples after the dialysis was complete.

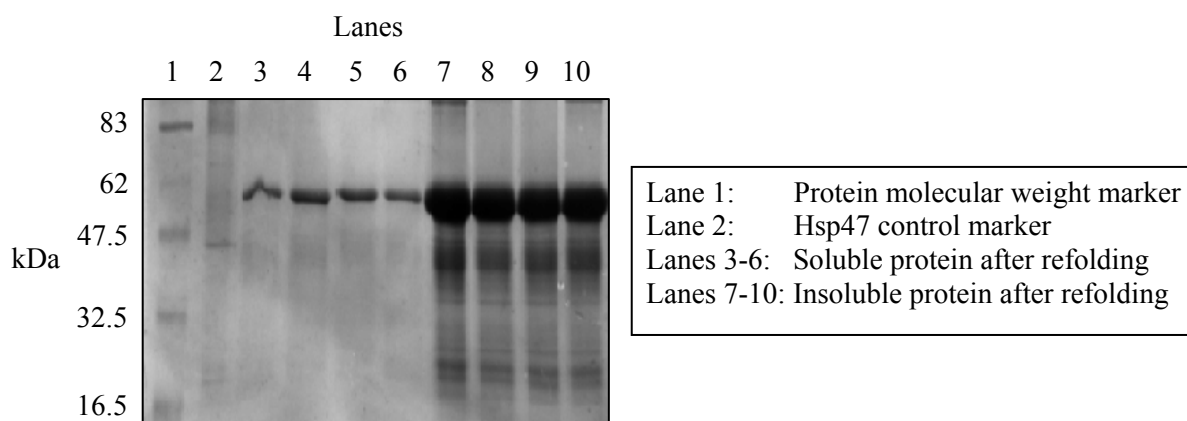


Figure 3.17. SDS-PAGE of refolding of insoluble *Xenopus laevis* Hsp47. The presence of bands at ~63 kDa in lanes 3-6 indicates that soluble protein (thioredoxin-tagged Hsp47) was present after the refolding experiment. However, the higher intensity bands in lanes 7-10 illustrate that the majority of the protein was still insoluble.

in pET-32 Xa/LIC

Hsp47 produced from this construct was also purified by metal affinity chromatography using Ni-NTA. As with the pET102/D-TOPO[®] vector, fractions were collected during the purification procedure for analysis by SDS-PAGE.

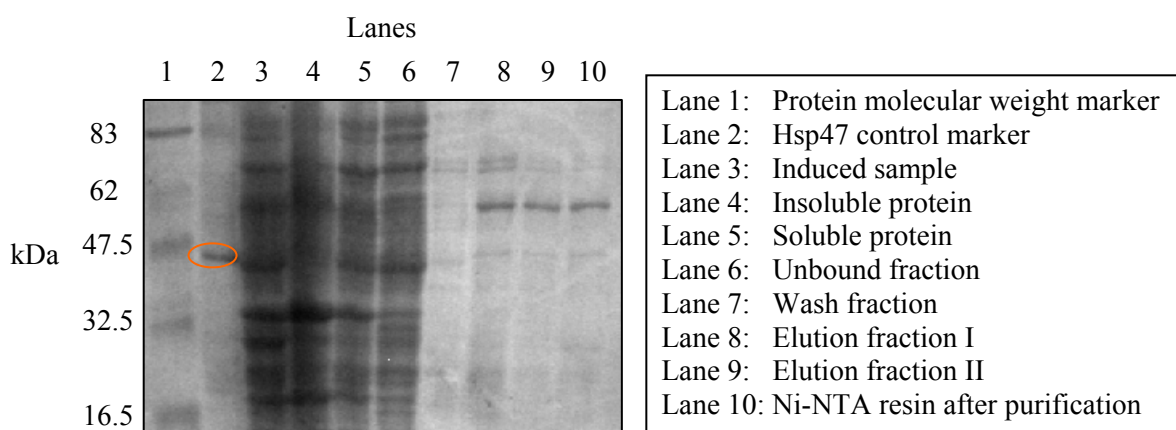


Figure 3.18. SDS-PAGE for purification of *Xenopus laevis* Hsp47 in pET-32 Xa/LIC construct. Presence of a band in the induced fraction (lane 3) at 47 kDa suggests that Hsp47 was being produced.

The bands in lanes 8 and 9 at ~63 kDa correspond to thioredoxin tagged Hsp47, suggesting that pure Hsp47 was obtained.

Analysis of the SDS-PAGE suggested that although the protein expression was small in comparison to the pET102/D-TOPO[®] vector, the majority of the protein was soluble and bound to the Ni-NTA resin. This can be seen in Figure 3.18 as there is no distinct band at ~63 kDa in lane 7. Some of the protein bound more tightly as it remained on the Ni-NTA even after protein elution with 250 mM imidazole (lane 10).

in pTYB12

Protein produced by the pTYB12 construct was purified using a chitin column. A chitin binding domain on the self-cleavable intein tag fused to Hsp47 allowed affinity purification of the protein. DTT was used for on-column cleavage of the intein tag from the recombinant protein (method described in section 2.10). SDS-PAGE was used to analyse the samples collected throughout the purification procedure.

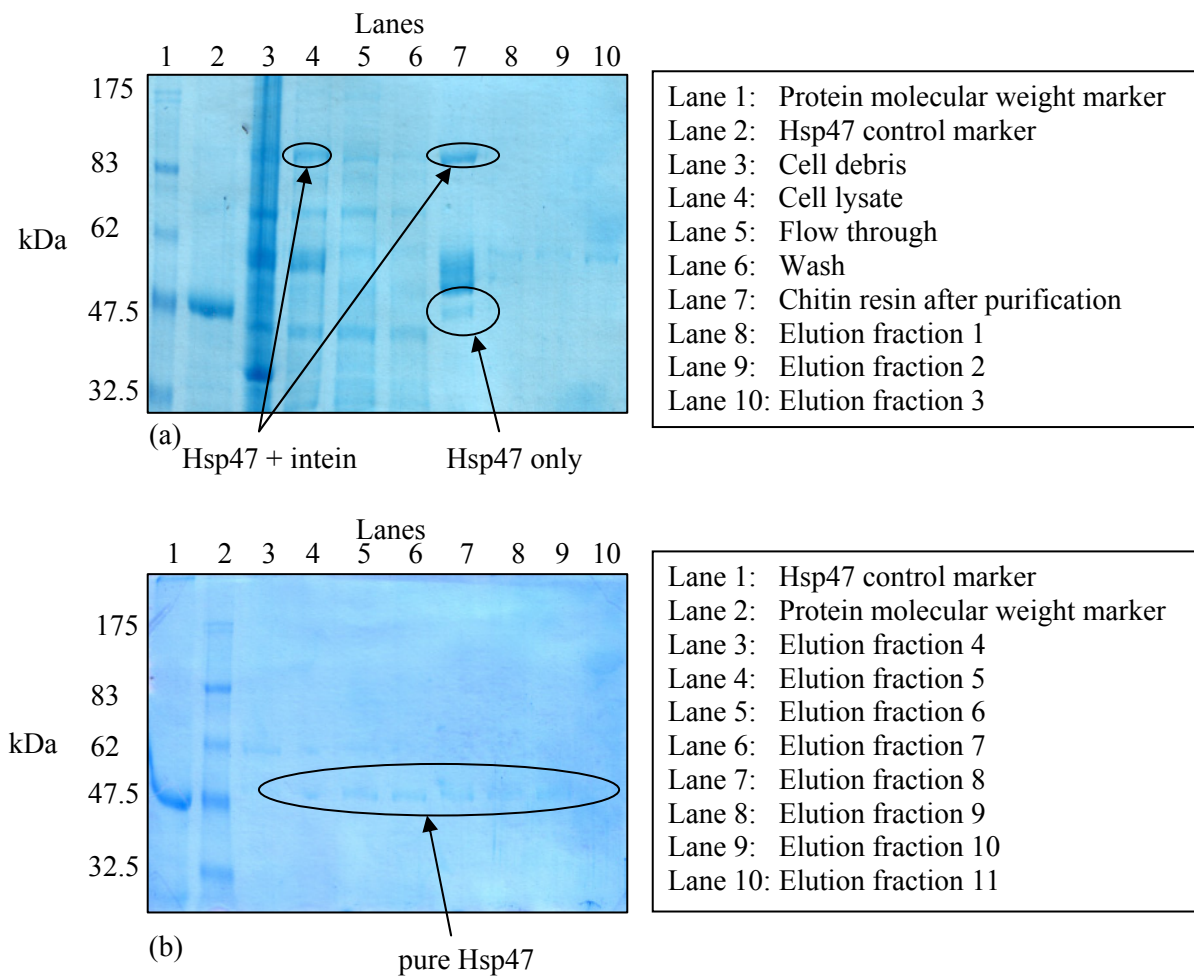


Figure 3.19. SDS-PAGE for purification of *Xenopus laevis* Hsp47 in pET-32 Xa/LIC construct.

(a) The ~102 kDa band in lane 4 represents Hsp47 with an intein fusion tag (55 kDa). Lane 5 illustrates that there was not much protein present in the flow-through fraction, suggesting that most of the protein bound to the chitin column. However, the intensity of the band at 102 kDa in lane 7 implies that the majority of the bound protein did not elute from the column. (b) Lanes 5-8 show faint bands at 47 kDa, corresponding to Hsp47 without the intein fusion tag. The low intensity of the bands, however suggest that there was a small amount of intein-free material after cleavage. Elutions fractions 1-11 are fractions collected during the on-column cleavage with DTT.

The protein produced by this construct was soluble as illustrated by the 102 kDa band in the cell lysate fraction (Figure 3.19(a), lane 4). The absence of distinct bands in the flow-through and wash fractions (Figure 3.19(a), lanes 5 and 6, respectively) implied that most of the expressed Hsp47 bound to the chitin column. However, much of this was not released upon

cleavage of the intein and remained bound to the column (Figure 3.19(a), lane 7). Nevertheless, there was some cleaved Hsp47 eluted from the column (Figure 3.19(b), lanes 5-8) that could be used for further analysis.

*3.2.3.2 Purification of Hsp47 from *Danio rerio**

in pTYB12

Purification using this vector was not successful as the protein did not elute from the chitin beads. However the protein preparation produced soluble protein which could be used for further attempts to purify this protein. A possible explanation for the unsuccessful purification may be due to the lack of intein tagged Hsp47 that was produced. Figure 3.13 illustrated that there was no *Danio rerio* intein-Hsp47 being expressed. Without the intein, protein purification on chitin resin would not be feasible due to the absence of the CBD.

3.2.4 Fusion tag cleavage

Cleavage of fusion tags was necessary because collagen-binding assays with fusion tagged Hsp47 showed no binding between protein and its substrate. It was therefore hypothesised that the His and/or thioredoxin tags may be interfering with the interaction of Hsp47 to collagen.

3.2.4.1 Cleaving the Thioredoxin tag from the pET102/D-TOPO[®] construct

The pET102/D-TOPO[®] vector contains an enterokinase (EK) cleavage site (illustrated in Figure 3.20) hence EKMax[™] was used to attempt to cleave the thioredoxin from the pET102/D-TOPO[®] construct.

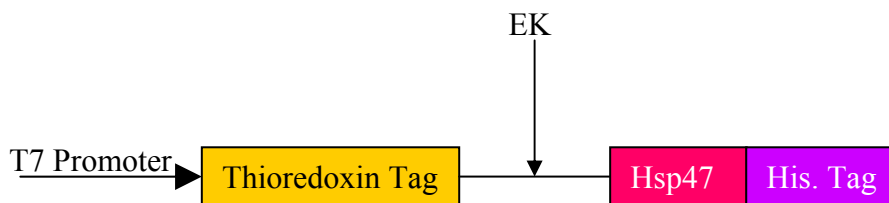
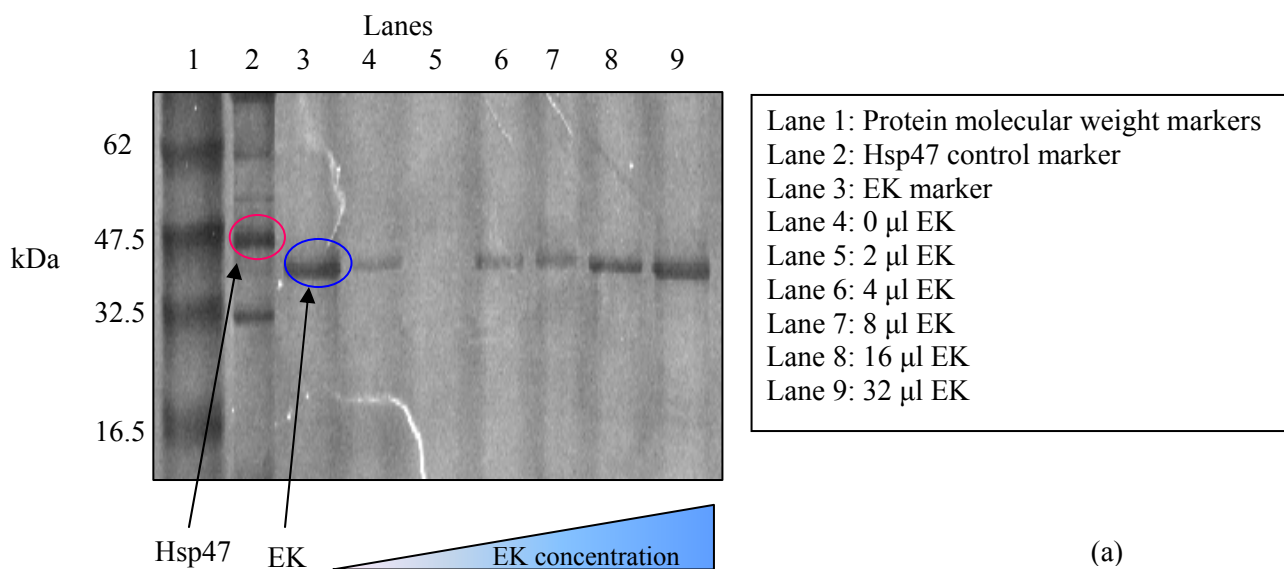


Figure 3.20. Illustrating the position of the EK cleavage site on the pET102/D-TOPO[®] vector.

Xenopus laevis

Two cleavage assays were run at 37 °C to determine optimum cleavage conditions. Varying EK concentrations and cleavage incubation times were tried in order to cleave the thioredoxin tag from the protein. The concentration of EK ranged from 0 µg/ml to 12.8 µg/ml, and two cleavage incubation periods of 6 and 24 hours were used.

Due to the minimal amount of pure protein obtained from this vector, detection of the protein following the cleavage with EK required the use of the silver stain method for detection of protein (Figures 3.21).



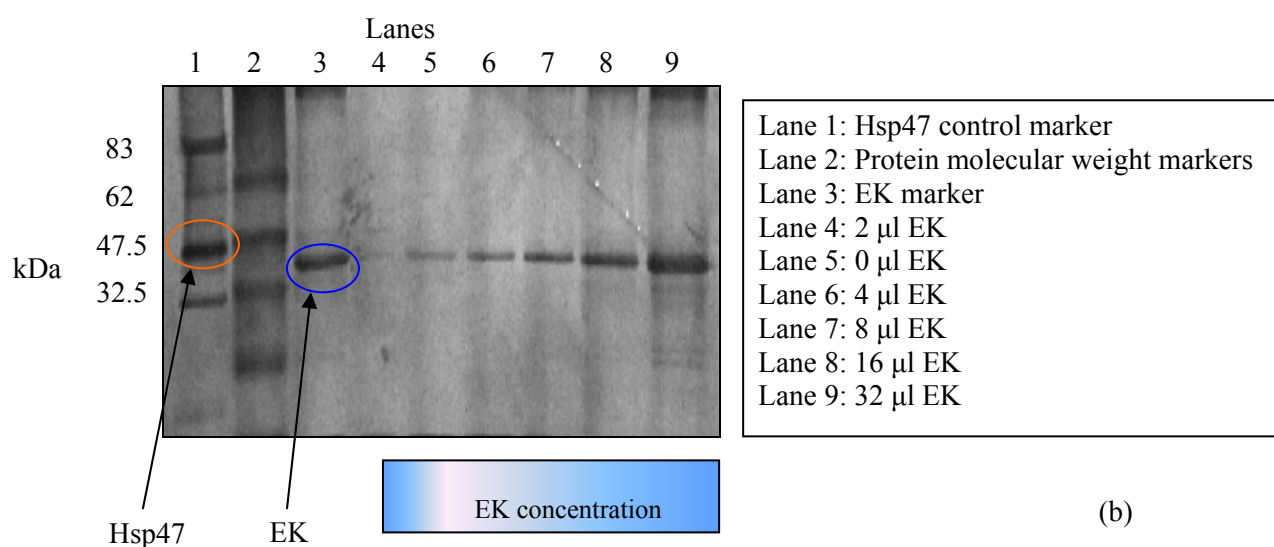


Figure 3.21. Result of digestion of thioredoxin tagged *Xenopus laevis* Hsp47 with enterokinase.

(a) Samples digested for 6 hours with differing amounts of protease. The absence of bands at 47 kDa in lanes 4-9 suggest that there was no successfully cleaved Hsp47 present. (b) Samples digested overnight with differing amounts of enterokinase. Again, no Hsp47 was visible in any of the fractions representing digested protein, implying that that cleavage was unsuccessful.

EKMaxTM is a clone of the catalytic subunit of enterokinase and has a molecular weight of 43 kDa. The results from the 6 hour and overnight digestions revealed that there was no Hsp47 present in any of the fractions but EK was present in all the fractions. This suggested that there was not enough Hsp47 in comparison to EK or it may have been that there was non-specific cleavage by the protease, resulting in complete digestion of Hsp47.

Danio rerio

As with *Xenopus laevis*, various concentrations of EK were used to try to cleave the thioredoxin tag. Two cleavage assays were run at 37 °C, using EK concentrations ranging from 0-12.8 µg/ml and reaction times of 6 and 24 hours. The result was digestion of the entire protein. This indicated that either the enzyme concentration was too high or the incubation period was too long. Consequently, the EK concentration was lowered 100 fold and the

cleavage incubation time was reduced to 3 hours. However this also proved to be ineffective as the SDS-PAGE illustrated one band corresponding to uncleaved thioredoxin tagged Hsp47. Another attempt at cleaving the fusion tag involved keeping the time for cleavage at 3 hours and varying the enzyme concentration between from the highest concentration where complete digestion of the protein occurred to the lowest concentration where no cleavage took place. The results revealed that no cleavage occurred at enzyme concentrations lower than 8 µg/ml and non-specific cleavage occurred at EK concentrations of 8 µg/ml and above. These results demonstrate that the failure to cleave the thioredoxin could be due to the tag being buried in the protein structure and not fully exposed.

On-column cleavage where the protein was cleaved whilst on the Ni-NTA column was attempted. This involved incubation of the column with 8 µg/ml of EK at 37 °C. Elution of the protein required 50 mM Tris-HCl buffer, pH 8.0 containing 250 mM imidazole. This attempt also proved to be unsuccessful.

The failure to cleave the thioredoxin tag on all attempts indicated that the EK was not recognising the cleavage site.

As with *Xenopus laevis*, all cleavage attempts were unsuccessful so it was decided that the uncleaved thioredoxin tagged protein would be used for protein characterisation studies.

3.2.4.2 Cleaving the Thioredoxin tag from the pET-32 Xa/LIC construct

This construct contained both Factor Xa (fXa) and thrombin cleavage sites.

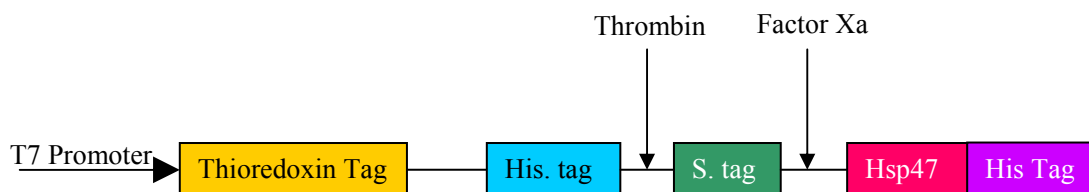


Figure 3.22. Protease cleavage sites on pET-32 Xa/LIC. Cleavage with thrombin cleaves the N-terminal thioredoxin and His tag. Factor Xa cleaves also cleaves the N-terminal S tag in addition to the thioredoxin and His tag.

The ability of these proteases to cleave the fusion tag was tested simultaneously. Experiments were set up so that the each individual protease and a combination of the two proteases were tested for their ability to cleave the thioredoxin tag.

Xenopus laevis

The initial protein cleavages, one with factor Xa alone and another with only thrombin were inconclusive as it was unclear whether the cleavages were successful. Therefore a second set of cleavage experiments (described in section 2.9.2) were set up to achieve a more definitive result. The results for both experiments are shown below.

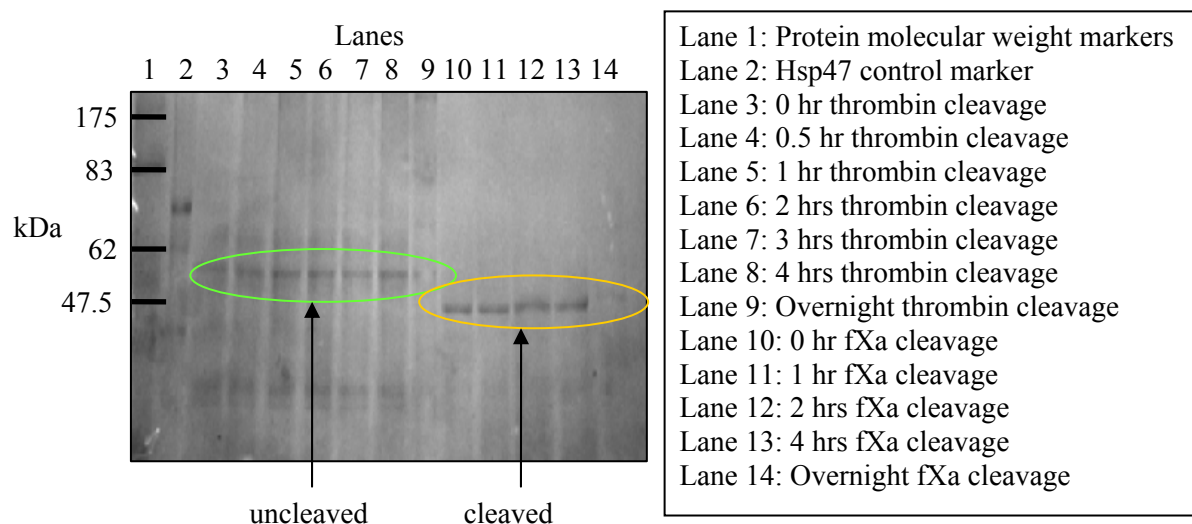


Figure 3.23. SDS-PAGE of attempted cleavage of thioredoxin tag from *Xenopus laevis* Hsp47 from pET-32 Xa/LIC using thrombin and factor Xa. The absence of a 47 kDa band in the fractions representing cleavage with thrombin (lanes 3-9) suggests that it was unsuccessful. Lanes 10-14, however, representing cleavage attempts with factor Xa, revealed bands at 47 kDa implying that cleavage of the thioredoxin tag with this protease was successful.

The attempted cleavage of the thioredoxin tag was unsuccessful even when left overnight. This is evident from Figure 3.23 as the bands at ~59 kDa (representing Hsp47 with attached thioredoxin tag) were present in all the fractions, even when the cleavage was left to occur overnight. Furthermore, there were no bands at 47 kDa in any of the fractions, confirming the absence of tag-free Hsp47. The cleavage of the thioredoxin tag with factor Xa, on the other hand was successful when left from 0-4 hours (Figure 3.23, lanes 10-13, respectively). The presence of a 47 kDa band in the fraction representing 0 hour cleavage time suggested that cleavage of the tag occurred immediately after the protease was added to the protein solution.

The second cleavage attempt involved four experiments; the first was a control in which no proteases were added, the second was an attempted with thrombin only, the third was

cleavage with factor Xa alone and the fourth cleavage attempt was with both thrombin and factor Xa together. The experiments showed variable success.

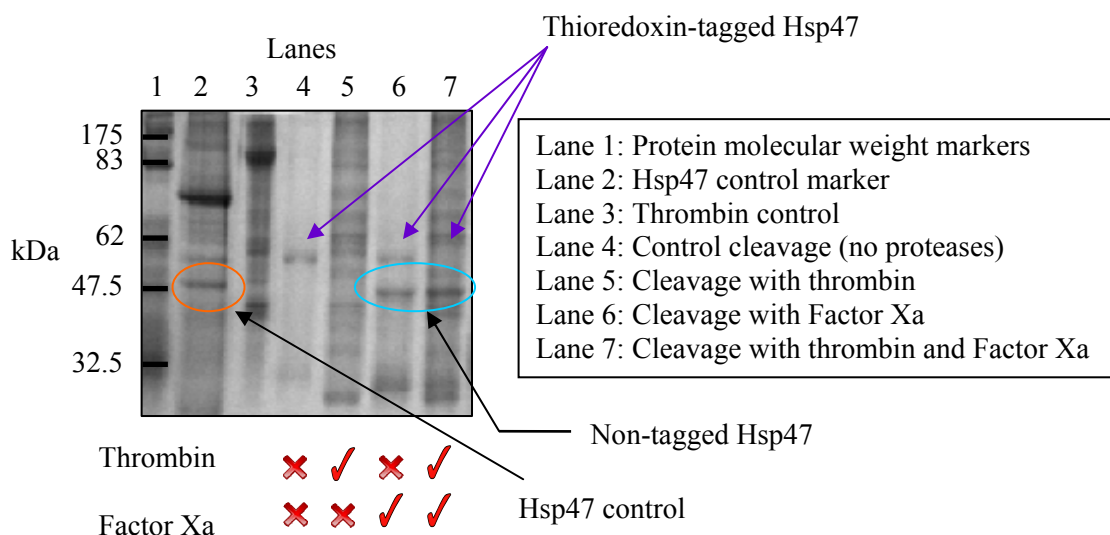


Figure 3.24. SDS-PAGE results for second attempted cleavage of the thioredoxin tagged *Xenopus laevis* Hsp47 from pET-32 Xa/LIC. The absence of a 47 kDa band in lane 5 implies that cleavage with thrombin was unsuccessful. Cleavage with factor Xa alone and thrombin + factor Xa together (lanes 6 and 7, respectively) revealed a band at 47 kDa, implying that cleavage of the thioredoxin tag was successful.

As previously, the attempted cleavage with thrombin was unsuccessful as it did not yield a band that corresponded to non-tagged Hsp47. The cleavage with factor Xa (Figure 3.24, lane 6) revealed a band at 47 kDa, suggesting that cleavage of the thioredoxin tag with this protease was successful. The cleavage with thrombin and factor Xa also showed a band at 47 kDa; however this was most probably due to the factor Xa rather than the thrombin as all previous cleavage attempts with thrombin were unsuccessful. As well as the 47 kDa band, lanes 6 and 7 also illustrated a band at ~59 kDa (corresponding thioredoxin tagged Hsp47) which implied that the some tagged protein was also present.

Danio rerio

SDS-PAGE revealed that digestion with thrombin resulted in non-specific cleavage; however the cleavage with factor Xa was successful, forming *Danio rerio* Hsp47 without an N-terminal fusion tag.

3.2.5 Collagen Binding Assays

The activity of the protein was determined using collagen binding assays. It was thought that active Hsp47 should bind to collagen whilst inactive protein should not.

The assay involved thawing 0.5 ml of frozen pure Hsp47 and spinning at 14, 000 x g to remove any aggregated insoluble protein. 50 µl of the resulting supernatant was combined with 50 µl collagen agarose beads along with 50 µl Tris buffer (100 mM Tris-HCl, 150 mM NaCl, pH 8.0) and the mixture incubated on ice for 30 minutes with turning. Any bound protein was removed after spinning again at 14, 000 x g for 5 minutes. The column was then washed and a sample of the resin beads was retained for SDS-PAGE.

3.2.5.1 Collagen Binding properties of Hsp47 in pET102/D-TOPO[®]

Xenopus laevis

The results for the collagen binding assay (shown in Figure 3.25) in this vector showed that the protein did not bind to either the control or collagen resin.

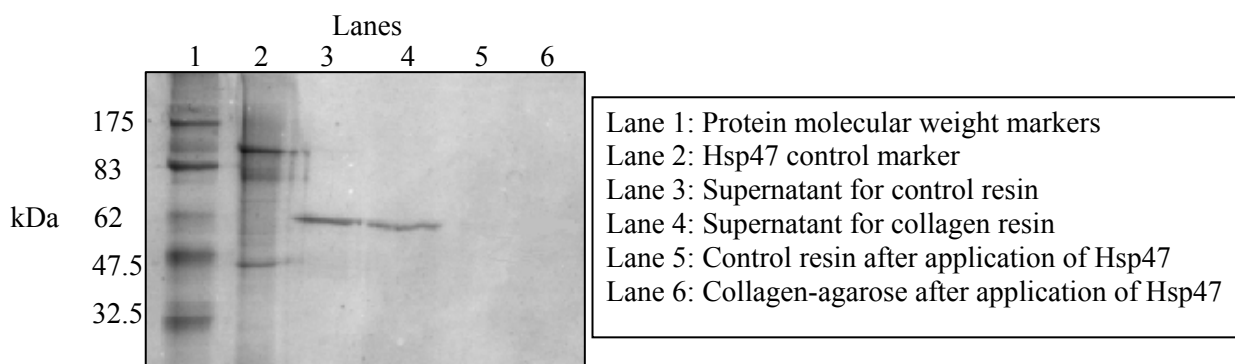


Figure 3.25. SDS-PAGE of collagen binding assay for *Xenopus laevis* Hsp47 from pET102/D-TOPO® construct. The presence of ~63 kDa bands in lanes 3 and 4 suggest that the HP thioredoxin tagged protein did not bind to either the control or collagen-agarose resin. This is confirmed by the absence of any protein in the fractions corresponding to resin samples (lanes 5 and 6).

The result for the collagen binding assay demonstrated that the uncleaved Hsp47 produced by the pET102/D-TOPO® construct did not bind to either the control or collagen-agarose resin. This could be due to the HP-thioredoxin tag interfering with the binding of the protein to the collagen-agarose beads.

Danio rerio

The thioredoxin tagged Hsp47 had equal binding affinity for the control resin and the collagen-agarose beads. This indicates that the protein did not show specificity for collagen. This may be due to the N-terminal thioredoxin tag interfering with the collagen binding properties of Hsp47.

3.2.5.2 Collagen Binding properties of Hsp47 in pET-32 Xa/LIC

Xenopus laevis

The results (Figure 3.26) indicated again that there was no binding of the protein to either the collagen or control (sephadex) resin.

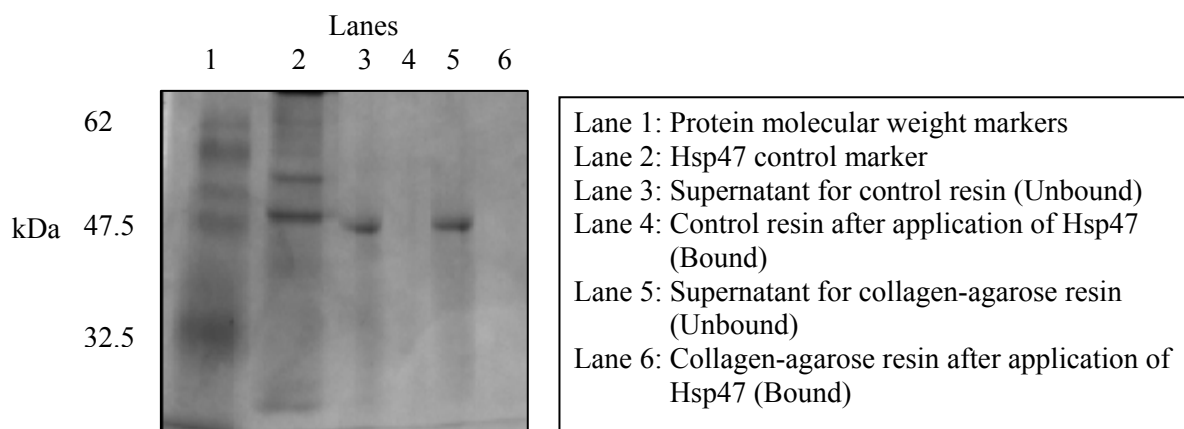


Figure 3.26. SDS-PAGE of collagen binding assay for *Xenopus laevis* Hsp47 from pET-32 Xa/LIC construct. The presence of bands at 47 kDa in both the supernatant fractions (lanes 3 and 5) suggested that the protein did not bind to either of the resins.

As with the protein produced by the pET102/D-TOPO[®] construct, the Hsp47 from the pET-32 Xa/LIC construct did not bind to either the control (sephadex) resin or the collagen-agarose resin, suggesting that the Hsp47 did not bind specifically or non-specifically.

The inability of the protein to bind to either of the resins indicated that the C-terminal His tag must be interfering with the binding of the Hsp47 to collagen.

Danio rerio

Although the N-terminal thioredoxin had been successfully cleaved, the protein failed to bind specifically to collagen. This indicated that the C-terminal His tag must be interfering with the collagen binding properties of Hsp47.

3.2.6 Characterisation of the Protein

Biophysical analyses were used for the determination of secondary structure composition and stability of the protein. Furthermore, the molecular weight and therefore oligomeric state of the protein could be calculated.

As described in the previous section, the protein failed to bind collagen. Therefore CD was considered a useful technique to determine whether the lack of substrate binding was due to the fold or stability of the protein.

3.2.6.1 Using *Xenopus laevis* Thioredoxin tagged Hsp47 from pET-32 Xa/LIC

CD Results

This involved two experiments that allowed the conformation of the protein to be determined. The first involved determination of secondary structure composition by measuring the change in CD signal over a wavelength range. The different secondary structures are identifiable according to the peaks produced between 180-260nm. Minima at 208 and 222 nm denote the presence of α -helices and a minimum at 216 nm represents β -sheet presence. Folded proteins also produce a positive signal between 190-200 nm.

Xenopus laevis Hsp47 produced a spectrum (Figure 3.27) that represented a protein that was neither fully folded nor unfolded, but a protein that was partly folded. The minimum at 216 nm, representing β -sheet was not evident and of the minima representing α -helical presence, only the minimum at 208 nm was clearly visible. Moreover, the signal at 200 nm was at -10 mdeg, implying that was not fully folded.

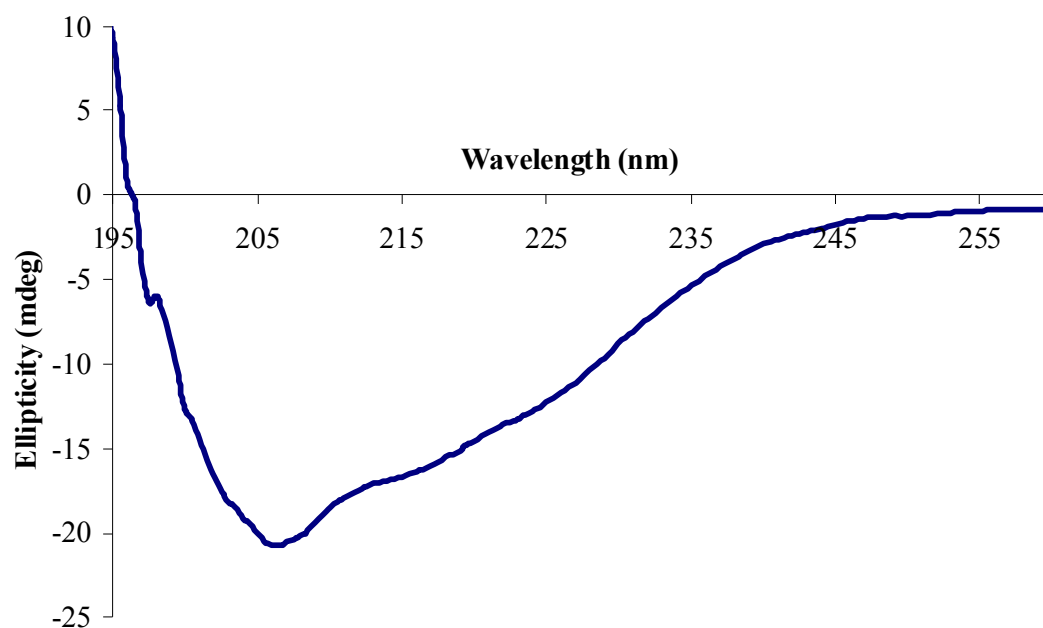


Figure 3.27. CD spectrum of purified thioredoxin tagged *Xenopus laevis* Hsp47. The spectrum does not represent a fully folded protein due to the negative signal at 200 nm. A minimum at 208 nm denotes the presence of α -helices and a minimum at 216 nm implies presence of β -sheet character.

The second experiment measured the effect of temperature on the protein conformation (Figure 3.28) to distinguish its melting point which would in turn allow determination of protein stability. It also allowed identification of whether the protein was in the stressed or relaxed state. The active monomer form of Hsp47 is known to be mesostable (unfolds at approximately 80 °C) whereas the oligomeric (inactive) forms unfold at greater than 100 °C.

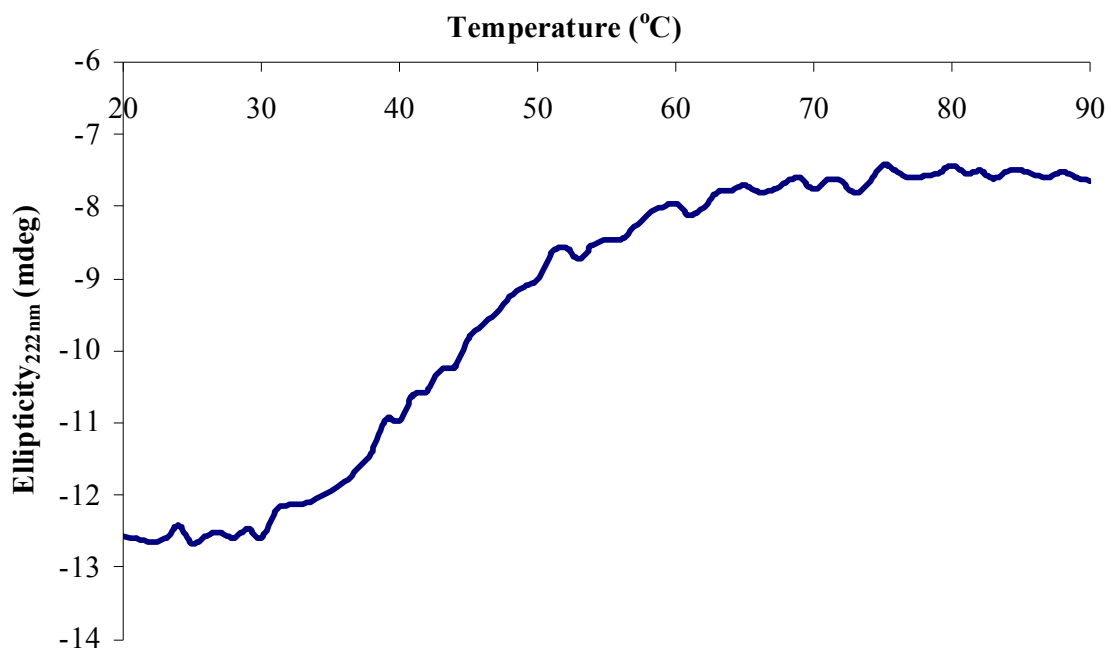


Figure 3.28. Thermal denaturation curve (at 222 nm) for thioredoxin tagged *Xenopus laevis* Hsp47 from pET102/D-TOPO[®] construct. There is a transition with a midpoint of 42 °C. This gradual loss of structure implies the unfolding of material.

Although slow denaturation of protein was evident from heating of the sample, the transition observed was not sharp. Hsp47 usually unfolds with a sharp transition, which represents a distinct loss of secondary structure rather than a gradual decline in the secondary structure composition.

Taken together, the results from the circular dichroism experiments suggest that neither the spectrum nor the denaturation curve were characteristic of Hsp47. It was therefore thought that the protein was largely unfolded, which may have been due to the interference of the thioredoxin tag still fused to the Hsp47. The small transition observed during the denaturation experiment may be due to the unfolding of the thioredoxin tag rather than Hsp47 as there was not very much folded material present at the beginning of the stability experiment (as shown by Figure 3.27).

Size Exclusion Chromatography Results

This was used to determine the approximate mass of the purified protein. The position at which the Hsp47 was eluted from the column was compared to the positions at which proteins of known mass were eluted and the molecular masses of the known proteins used to calculate a probable mass of the Hsp47 purified in this project. Figure 3.29 shows the position at which the purified Hsp47 eluted in the column.

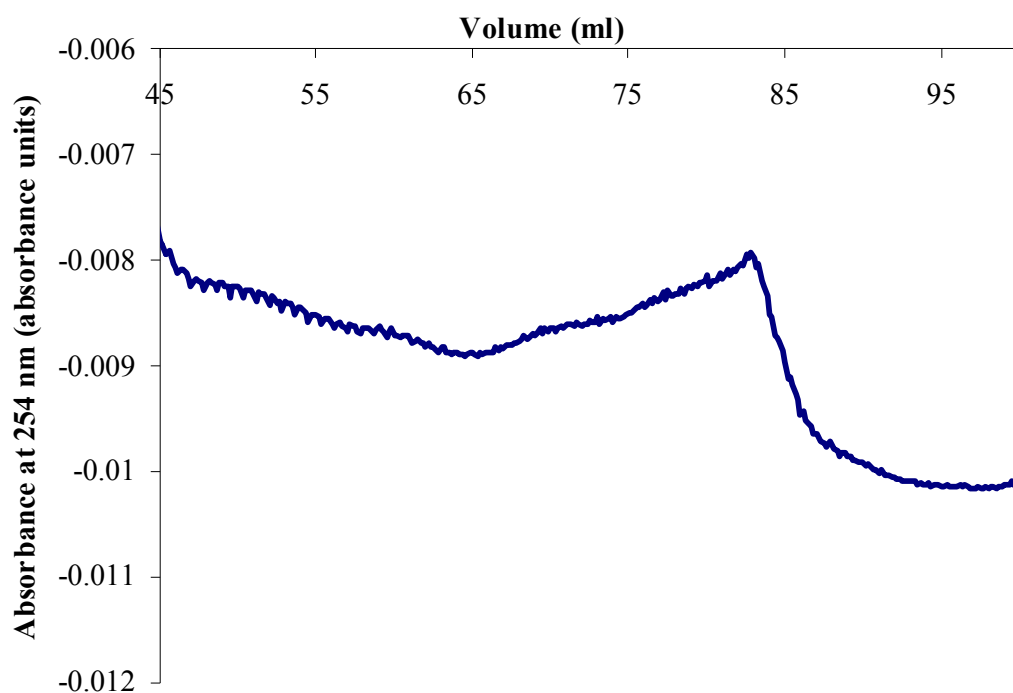


Figure 3.29. Graph to show elution of *Xenopus laevis* Hsp47 by size exclusion chromatography.

The peak at 83.35 ml signifies the elution of Hsp47.

A western blot (data not shown) was used to confirm that the protein eluted was Hsp47.

A chromatogram for proteins of known masses was used to determine the size of the *Xenopus laevis* Hsp47.

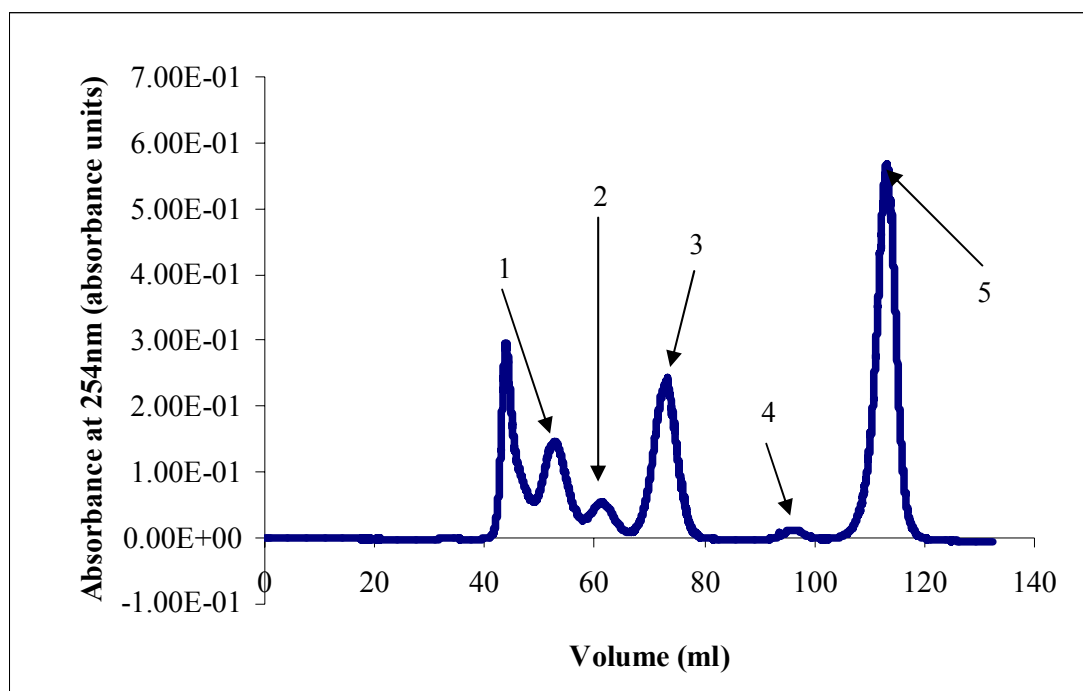


Figure 3.30. Elution positions of proteins with known masses. Peak one is due to Ferritin- a 440 kDa protein, peak two is Ig G- a 160 kDa protein, BSA (67 kDa) produces the third peak, ribonuclease A (13 kDa) is responsible for the fourth peak and the fifth peak corresponds to Vitamin B12- which is a 1.355 kDa protein.

The positions at which each of the proteins eluted were used to plot a graph of volume vs. \log_{10} mass (Figure 3.31) to determine the approximate mass of the purified Hsp47.

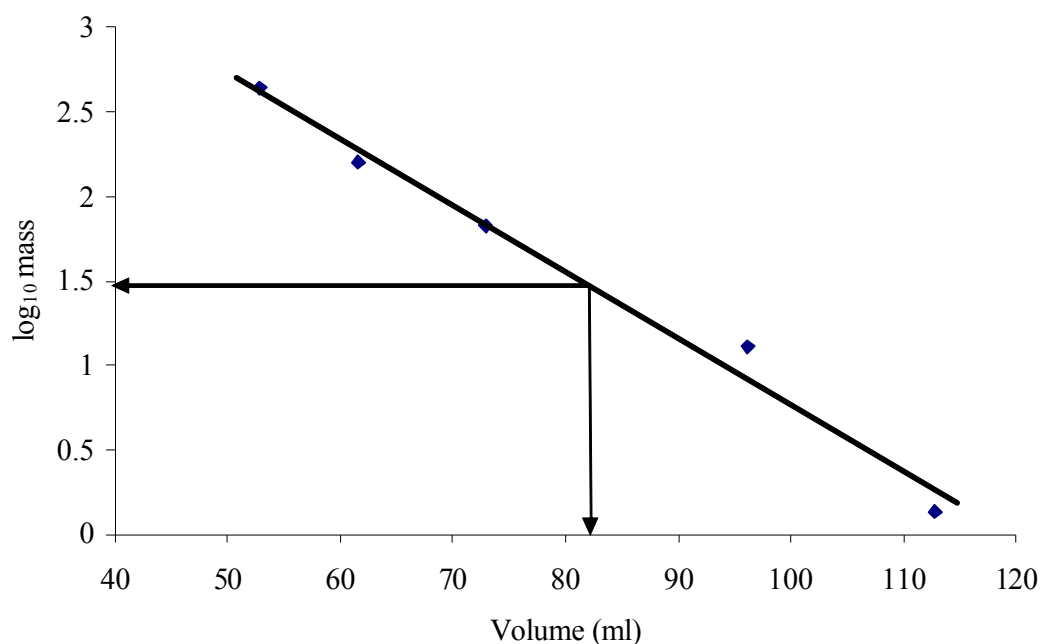


Figure 3.31. Determination of the mass of *Xenopus laevis* Hsp47. The masses of the known proteins were logged and used to plot a straight-line graph that could be used to get a log mass value for the Hsp47 protein. Figure 3.28 showed that the Hsp47 eluted at 83.35ml. This value was used to identify the point on the line that corresponded to the protein. Its log₁₀ mass value could then be identified by looking at the y-axis value for the point on the line. The log₁₀ mass is approximately 1.48.

The antilog of $1.48 = 30.2$, thus the mass of the purified Hsp47 was close to 30.2 kDa. This is lower than the known mass for Hsp47 (47 kDa) and the difference was probably due to the interactions between the protein and the gel filtration column. This was thought to be the case as SDS-PAGE of *Xenopus laevis* Hsp47 in pET102/D-TOPO[®] illustrated a band at 47 kDa (Figure 3.14). However, we identified that the purified protein was a monomer as the size of the protein clearly suggests that it could not be a dimer or otherwise. This was consistent with the results from the CD experiments, which identified the protein might be metastable. The interaction of the protein with the column may have been due to the partially folded state of the Hsp47, as suggested by the spectrum and denaturation analyses (Figures 3.27 and 3.28, respectively).

3.2.6.2 Using *Danio rerio* Thioredoxin tagged Hsp47 from pET102/D-TOPO[®]

Size exclusion chromatography of *Danio rerio* Hsp47 was carried out in our lab prior to this study. The results from size exclusion chromatography showed that the protein was present in the hexameric form whilst the results from CD showed that the protein comprised of both α -helices and β -sheets. There was very little structural change during the transition from 20 °C to 80 °C, suggesting that *Danio rerio* Hsp47 existed in a hyperstable conformation.

Organism	Vector	Protein Purification		Cleavage of His tags				Collagen binding activity	CD	
		Soluble	Insoluble	EK	FXa	Thrombin	Thr + FXa		Spectra	Melt
<i>Xenopus laevis</i>	pET101/D-TOPO [®]	No expression		-	-	-	-	-	-	-
	pET102/D-TOPO [®]	**	**	✗	-	-	-	✗	-	-
	pET-32 Xa/LIC	*	*	-	✓	✗	✓	✗	unfolded	hyperstable
	PTYB12	**	*	Construct contained self-cleavable intein tag				-	-	-
<i>Danio rerio</i>	pET101/D-TOPO [®]	No expression		-	-	-	-	-	-	-
	pET102/D-TOPO [®]	**	**	✗	-	-	-	Non-specific ✓	folded	metastable
	pET-32 Xa/LIC	*	**	-	✓	✗	-	✗	-	-
	PTYB12	**	*	Construct contained self-cleavable intein tag				-	-	-

Table 3.1. Summary of results for *Xenopus laevis* and *Danio rerio* Hsp47.

3.3 Discussion

Heat shock protein 47 is a 47 kDa collagen specific molecular chaperone and is a member of the serpin superfamily of proteins. The majority of studies on Hsp47 have focused mainly on mouse Hsp47. However, this form of the chaperone is relatively insoluble and unstable in solution, making it difficult to work with.

The genes for *Danio rerio* and *Xenopus laevis* Hsp47 were cloned into four vectors. Subsequently the protein was expressed and purified using various techniques. Following successful purification, cleavage of fusion tags was performed and subsequently the protein was characterised by CD and size exclusion chromatography. Activity was determined by collagen binding assays.

Thus far, this research has focused on the *Danio rerio* and *Xenopus laevis* forms of Hsp47. Very little data has been published regarding these forms of the protein and it was hoped that they would be easier to work with. So far in this project, *Danio rerio* and *Xenopus laevis* Hsp47 have been successfully cloned, expressed and purified. The thioredoxin tagged form has also been characterized using collagen-binding assays, CD and size exclusion chromatography.

Cloning of the Hsp47 Gene

The cloning of *Danio rerio* Hsp47 into pET101/D-TOPO[®] was successful as the sequencing results indicated that the gene had successfully inserted into the vector, however over-expression of the protein did not occur. *E. coli* use the majority of their protein synthesis machinery to over-express a protein, thus it is not surprising that expression problems arise [188]. One possible explanation for the lack of over-expression could be due to translational problems that are a result of *Danio rerio* having a different codon bias to *E. coli*. This occurs

because *E. coli* cells have inadequate tRNA to recognise particular codons on the eukaryotic gene. This can cause amino acid substitutions and frameshifts, halting translation and subsequently preventing over-expression of the protein. There are ways in which this problem can be overcome. One such method is the mutation of the codon that is disrupting the expression to a codon that is favoured by *E. coli*. Another method is to reduce the temperature during the growth of the culture so that protein expression is slowed and the chance of aggregation is reduced. The cloning results of *Danio rerio* Hsp47 into pET101/D-TOPO[®] suggested that this vector was unsuccessful as an expression vector for *Xenopus laevis*.

Danio rerio and *Xenopus laevis* Hsp47 were successfully cloned into pET102/D-TOPO[®]. The recombinant plasmid was transformed into BL21 Star One Shot *E. coli* cells; subsequently soluble thioredoxin-Hsp47 was obtained for both *Danio rerio* and *Xenopus laevis*. The production of soluble protein was enhanced by the presence of the HP-thioredoxin tag at the N-terminus. The presence of a 6xHis tag at the C-terminus enabled the purification of the fusion protein using Ni-NTA. Cleavage of the thioredoxin tag using EK was unsuccessful albeit that the vector pET102/D-TOPO[®] encodes a cleavage site for this protease. This is a common problem as it has been encountered by groups working on other proteins [189]. It is thought that the EK may have been contaminated by other proteases, resulting in the non-specific cleavages that were seen. A way of overcoming this problem may be the enterokinase EKMax. This protease is produced by over-expression in yeast cells and so undergoes post-translational modifications meaning that it can be expressed without risk of protease contamination. This form of the EK is thought to be very specific; therefore the likelihood of fusion tag cleavage is greater.

Both *Danio rerio* and *Xenopus laevis* Hsp47 was successfully cloned into pET-32 Xa/LIC. This vector possesses an N-terminal thioredoxin fusion tag at its N-terminal and a His tag at

its C-terminal. Cleavage of the N-terminus thioredoxin can be done using Thrombin or Factor Xa as this vector has cleavage sites for both these proteases. Transformation of the recombinant plasmid resulted in the production of soluble protein that could be purified using Ni-NTA resin. The thioredoxin tag was successfully cleaved using Factor Xa. Cleavage of the thioredoxin tag by Factor Xa resulted in the removal of vector encoded sequences from the purified protein leaving an N-terminal free of fusion tags.

The cloning of the *Danio rerio* and *Xenopus laevis* Hsp47 gene into pTYB12 was also a success. Following transformation into BL21 Star One Shot Cells, soluble *Xenopus laevis* Hsp47 was obtained. *Danio rerio* Hsp47 has proved more difficult to attain and this will be the primary aim of future work. As of yet only a small amount of pure *Xenopus laevis* Hsp47 has been acquired using the chitin resin and it is hoped that this can be increased once the optimum on-column cleavage conditions have been established.

To conclude, it can be said that the pET101/D-TOPO[®] vector was not suitable for the cloning of the Hsp47 gene as expression could not be induced. Although the gene was successfully cloned into pET102/D-TOPO[®], cleavage of the thioredoxin tag with EK resulted in non-specific cleavage of the entire protein. Cloning into the pET-32 Xa/LIC was highly successful as it yielded pure, soluble protein from which successful thioredoxin fusion tag cleavage was enabled using Factor Xa. Cloning into the pTYB12 vector was also successful but produced only small amounts of soluble *Danio rerio* and *Xenopus laevis* Hsp47.

Molecular Characterisation by Collagen Binding Assay

Both the thioredoxin tagged (cloned into pET102/D-TOPO[®]) and non-thioredoxin tagged (cloned into pET-32 Xa/LIC) form of the protein failed to bind collagen. Therefore it seems likely that C-terminus fusion tags interfere with the collagen-binding site of *Danio rerio* and

Xenopus laevis Hsp47. This implied that the collagen-binding site is likely to be located near the C-terminus of the protein. Another possible explanation for the lack of Hsp47-collagen binding could be that the incompletely folded protein was inactive.

Molecular Characterisation by CD and Size Exclusion Chromatography

Hsp47 belongs to the serpin superfamily of proteins. A serpin moves from a metastable, native conformation to a hyperstable, cleaved conformation when the cleaved RCL inserts itself as a β -strand into the A-sheet (Figure 3.28). CD can be used to observe the changes in the stability of the protein, which can in turn identify the conformation of a serpin [152].

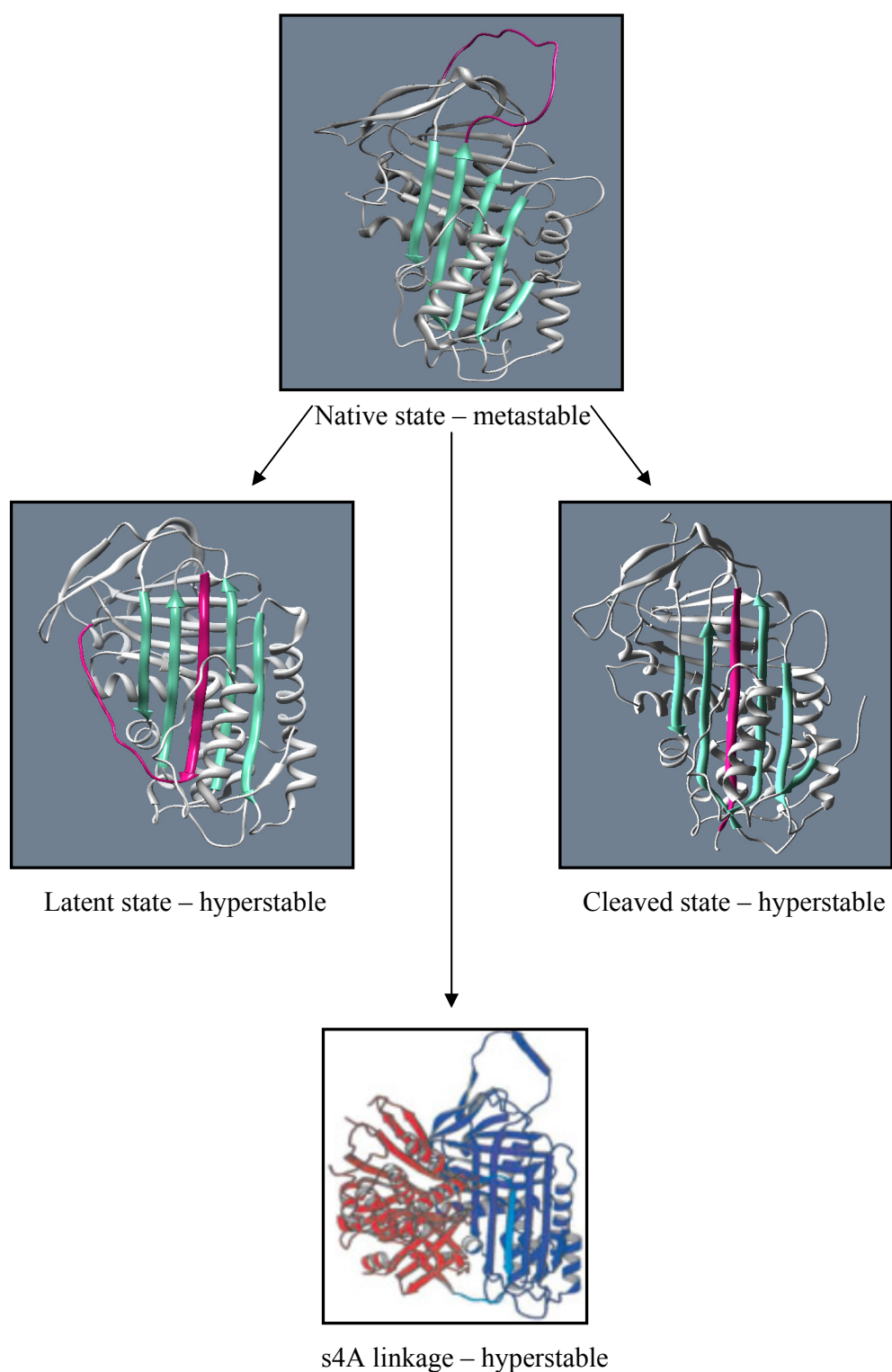


Figure 3.32. Possible serpin conformations. Denaturation of the native metastable form occurs at temperatures between 40 °C and 70 °C. The cleaved form is hyperstable and denatures at temperatures exceeding 100 °C. The latent and polymerized forms are also hyperstable and denaturation occurs at temperatures between 70 °C and over 100 °C. (Adapted from Gettins, 2002)

The CD results for *Xenopus laevis* Hsp47 showed that the protein (cloned into the pET102/D-TOPO[®] vector) had a midpoint melting temperature of 42 °C, indicating that the protein was in the monomeric metastable form. The CD spectrum for the protein suggested the presence of α -helices and β -sheets; this is characteristic of members of the serpin protein superfamily. Results of the size exclusion chromatography were consistent with those obtained by CD, identifying the protein as being monomeric.

The results from this study implied that purified *Xenopus laevis* Hsp47 existed in a partially folded, monomeric conformation that denatured upon heating to 80 °C. Due to the gradual loss of secondary structure during an increase in temperature and a CD spectrum denoting a partially folded protein, it was hypothesized that the thioredoxin tag prevented the correct folding of the protein. Furthermore, the prolonged transition upon denaturation was thought to be due to the unfolding of the thioredoxin rather than Hsp47. The results from the size exclusion experiment also strengthened this proposition as the interaction between Hsp47 and the column could have been a result of residues that are buried within the folded protein being exposed in the partially folded conformation, thereby forming unwanted contacts.

Analysis of the CD results for *Danio rerio* Hsp47 in our lab suggested that the protein exists in a hyperstable conformation, which withstood denaturation when heated to 80 °C. These results indicate that self-insertion of the RCL has occurred. As Hsp47 does not possess protease inhibitory properties, cleavage by the RCL is unlikely. Therefore it seems likely that RCL insertion has occurred without cleavage; forming the latent form. Another possibility is that the protein had polymerised, where the RCL of one serpin inserts itself into the A-sheet of another serpin. The latent and polymerised forms of serpin conformations have been shown to have similar stabilities; both denaturing at temperatures between 70 °C and above 100 °C. Results from the size exclusion chromatography indicate that *Danio rerio* Hsp47 existed as a

trimer. This did not come as a surprise as mouse Hsp47 has also been shown to exist as trimer and it is thought that this may be significant to the association of Hsp47 to collagen [87].

CHAPTER 4: Purification and Characterisation of Wild Type (WT) *Mus musculus* Hsp47

4.1 Introduction

Most of the previous work on WT Hsp47 has used the mouse form of the protein due to facile manipulation of the gene into vectors and non-problematic uptake of host bacterial strain cells. Furthermore, protein solubility and inconsistent protein expression have been a major issue in Hsp47 purification studies.

As shown in Chapter 3, *Xenopus laevis* and *Danio rerio* Hsp47 were difficult to purify and it was proving problematic to produce Hsp47 free of any C-terminal tags. Therefore, a more reliable source of the protein was required in order to study the effect of pH on the structure and function of Hsp47. As the gene for *Mus musculus* Hsp47 had been successfully cloned previously and generally, mHsp47 had been shown to express a reliable, soluble form of the protein, it was decided that mHsp47 would be used for studying the effect of pH on the secondary and tertiary structure of Hsp47. WT mHsp47 contains five tryptophan residues and fourteen histidine residues. As explained in section 1.8, the binding and release of Hsp47-collagen is pH dependent, occurring at a pH that is extremely close to the pK_a of histidine residues. It is therefore widely believed that the protonation/deprotonation of histidine residues are involved in triggering the pH-dependent conformational change of Hsp47. El-Thaher and coworkers [97] have studied this pH dependency and found that mutation of certain histidine residues altered the collagen binding activity of mHsp47.

Examination of the sequences of Hsp47 from all species revealed eleven histidines that are conserved in over 90% of Hsp47 sequences (Figure 4.1). Two of these histidines (H302 and

H335) are conserved across the serpin superfamily, making them less likely to be involved in the activity of the chaperone.

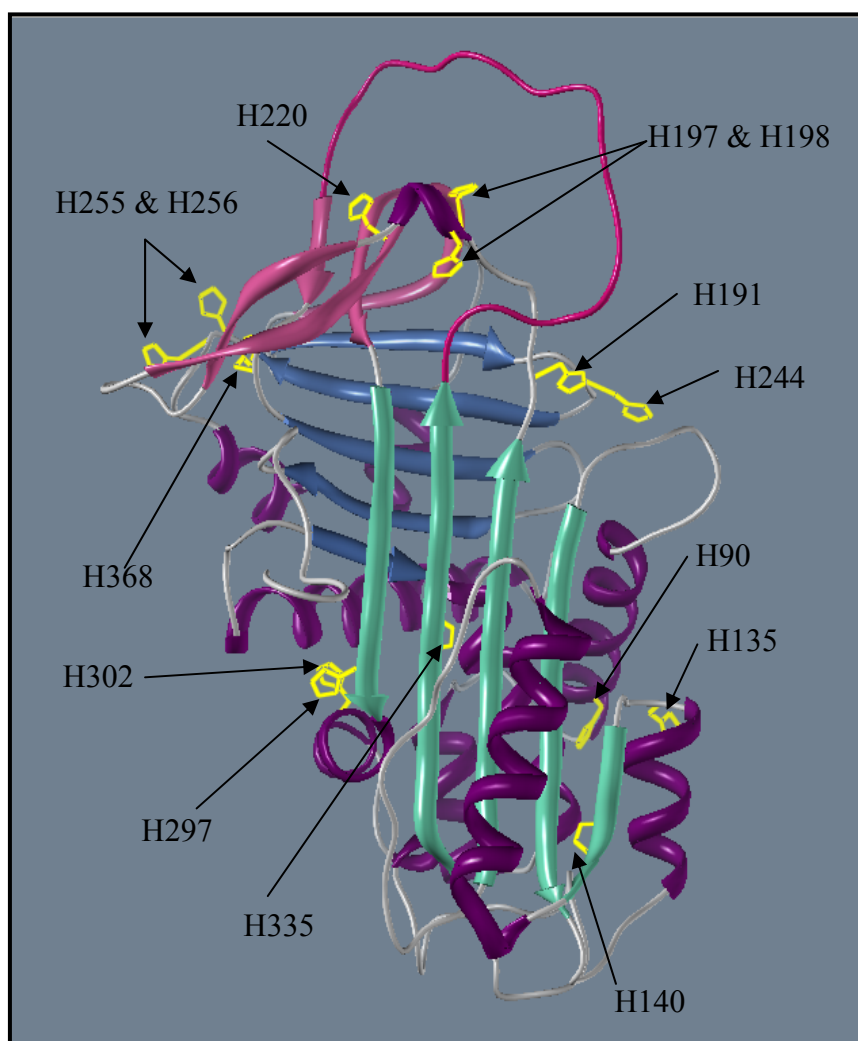


Figure 4.1. Model Hsp47 structure illustrating positioning of histidine residues.

The previous construct utilised a heat shock promoter to attempt to purify the untagged protein using a collagen affinity purification method. This is not ideal for mutant proteins that are thermally unstable and/or do not exhibit collagen binding. Therefore it was important to develop a construct that does not rely on collagen affinity (for example, using an added hexahistidine tag) and that used a different induction system (such as an IPTG inducible promoter).

The gene for mHsp47 was cloned into the pCR[®]T7/NT-TOPO[®] vector by Dr. Taka and Dr. Olerenshaw at Imperial College. Expression of the protein was achieved by transformation into BL21 star (DE3) One Shot pLysS competent cells.

This chapter details the analysis of this new construct. *Mus musculus* Hsp47 was overexpressed in *E. coli* and purified using collagen affinity chromatography. A pH gradient of 0.3 M phosphate pH 9.0 to 3.0 was used to elute the protein. The absorbance (at 280 nm), pH and salt concentration were recorded throughout the purification in order to establish when the protein eluted and the conditions under which this occurred. To probe the pH-dependent interaction of Hsp47 with collagen further, the secondary structure and stability of the protein was examined using CD spectroscopy. AUC was used to determine the approximate molecular weight and hence oligomeric state of the protein. The effect of pH on secondary structure was studied by monitoring changes in CD signal as protein solutions were titrated from pH 8.0 to 2.5. The activity of the eluted Hsp47 was established by observing its effect on collagen fibrillisation.

Previous studies on the pH-dependent binding and release of Hsp47 from its substrate collagen by El-Thaher and coworkers [97] have shown that the interaction is abolished below pH 6.3. In order to elute the protein, El-Thaher and coworkers lowered the pH to 5.3 using Tris-HCl. The elution method used in this study differed from this in that a pH gradient was used so that the exact elution pH could be determined. Another difference was that in this study, the salt concentration was monitored to investigate its effect on the interaction of Hsp47 and collagen. Furthermore, the collagen-affinity chromatography method has an advantage in that any protein that is eluted upon reduction of the pH below 6.3 would be active as it remained bound to the collagen during the wash and high pH elution gradient. A disadvantage of the affinity purification as opposed to the gel filtration method used by El-

Thaher and coworkers was that the oligomeric state of the protein could not be determined immediately. However, in this study this was overcome by using AUC which provided information on the molecular weight as well as the polymeric nature of proteins.

Collagen fibril formation

The effect of the protein on collagen fibrillisation was investigated by measuring the effect of Hsp47 on the turbidity of collagen from calf Achilles tendons (method described in section 2.2.13.5). Thomson and Ananthanarayanan have used physicochemical techniques to demonstrate the binding of Hsp47 to native and denatured collagen. They have shown Hsp47 to prevent the *in vitro* aggregation of collagen into fibrils at neutral but not acidic pH [96], which is in agreement with the hypothesis of Hsp47 being involved in preventing the premature association of procollagen in the ER and during its transportation from the ER *in vitro*.

Much of the studies of growth of collagen fibrils *in vivo* and the precipitation of collagen *in vitro* have focused on appearance using electron microscopy. The kinetics of fibril formation from collagen solutions has been studied in detail by G.C. Wood and coworkers [190-192]. They have investigated the precipitation of collagen fibrils *in vitro* in an attempt to better understand fibrillogenesis *in vivo*. This involved systematic studies examining how the pH, ionic strength, temperature and collagen concentration effect the rate of collagen fibril formation. The source of collagen was calf dermis that comprised of non-striated filaments, without any collagen fibrils. The course of precipitation was followed by measuring the turbidity of the solution. All precipitation curves consisted of a lag phase (where no precipitation occurred) and a subsequent growth phase that had a sigmoidal shape. It has been suggested that the two phases of the curves signify two consecutive steps in the precipitation process; both of which are enhanced by a rise in temperature. Therefore collagen fibril

formation is known to comprise of two phases; the nucleation phase which involves aggregation of particles in solution to form nuclei and a growth phase that consists of accumulation of the soluble material to form a precipitate [191]. Bensusan and Hoyt [193] showed that the width of the collagen fibrils *in vitro* is determined in the course of the lag phase, prior to any notable precipitation occurring. Therefore, it can be said that the lag phase is primarily associated with the creation of microfibril nuclei [194].

The method used in this study was solely to study the collagen de-fibrillation activity of the protein; therefore measurements were taken at a constant neutral pH. The collagen used was made from fresh calf Achilles tendons and isolated using acetic acid (method described in section 2.2.13.5).

4.2 Results

The PCR and cloning of mHsp47 were carried out by Dr. Olerenshaw and Dr Taka at Imperial College. All constructs were sequenced, the results of which are shown in the Appendices (Chapter 11, Appendix C).

Purification of *Mus musculus* Hsp47

Following successful cloning, the protein was expressed in *E. coli* and purified using collagen affinity chromatography using a pH gradient from pH 9.2 to 3.0, produced by varying quantities of 0.3 M dibasic sodium phosphate (pH 9.2) with 0.3 M monobasic potassium phosphate (pH 3.0). Data of absorbance, salt concentration and pH were recorded and an elution profile plotted (Figure 4.2).

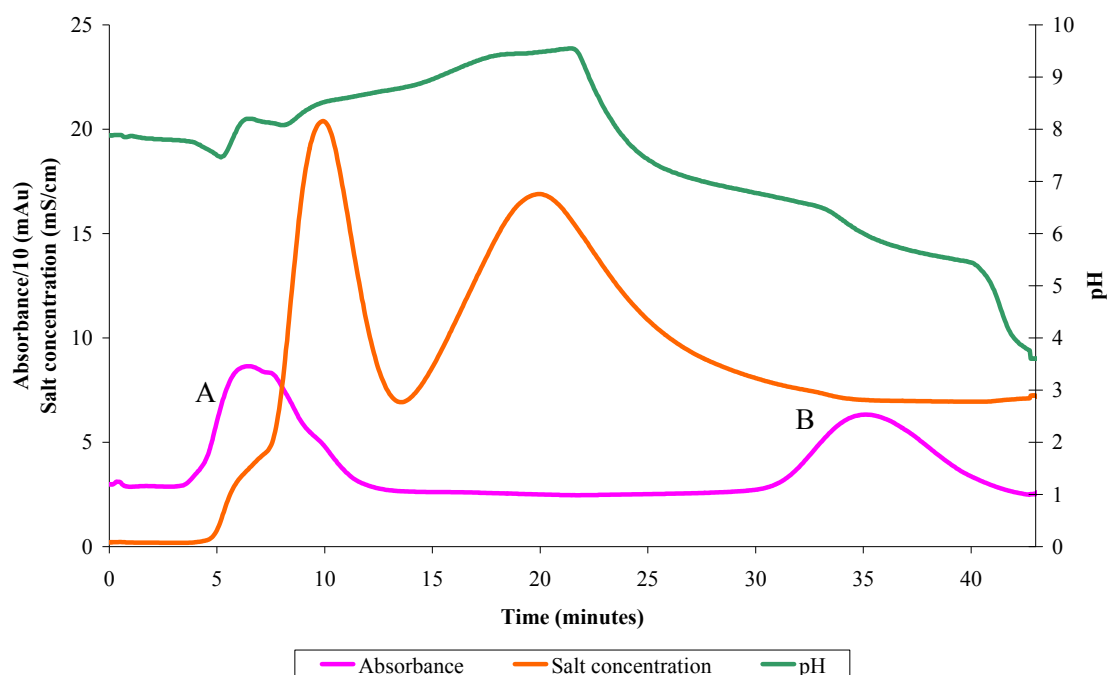


Figure 4.2. Elution profile of WT *Mus musculus* Hsp47. The pH-dependent elution of Hsp47 was probed using a gelatin-agarose column and pH elution gradient from high pH to low pH. The graph exhibits two distinct absorbance peaks (pink trace), each of which represents a point at which protein was eluted from the column. Peak A corresponds to protein that was eluted at approximately pH 8.0 whilst Peak B signifies protein that was eluted at approximately pH 6.1.

The chromatogram illustrates two distinct absorbance maxima that signify the presence of two different forms of mHsp47. Peak A (eluted between 5-10 mins) represents a form that is salt concentration dependent. This is demonstrated in the profile as protein elution correlates with an increase in salt concentration, however the pH remains above 8.0. Peak B (eluted between 32-40 mins) represents a form which is independent of salt concentration but dependent upon the pH. This is evident from the profile as the position of the peak correlates with a drop in pH whilst the salt concentration remains constant. The binding of this material to collagen was also salt-resistant. The pH at which this species was eluted was 6.1. This observation of two species is novel as this is the first instance in which peak A has been detected. Monitoring and varying the salt concentration enabled differentiation between two species of Hsp47 that require different conditions in order to be eluted from collagen.

Elution fractions were collected during the purification and analysed by SDS-PAGE.

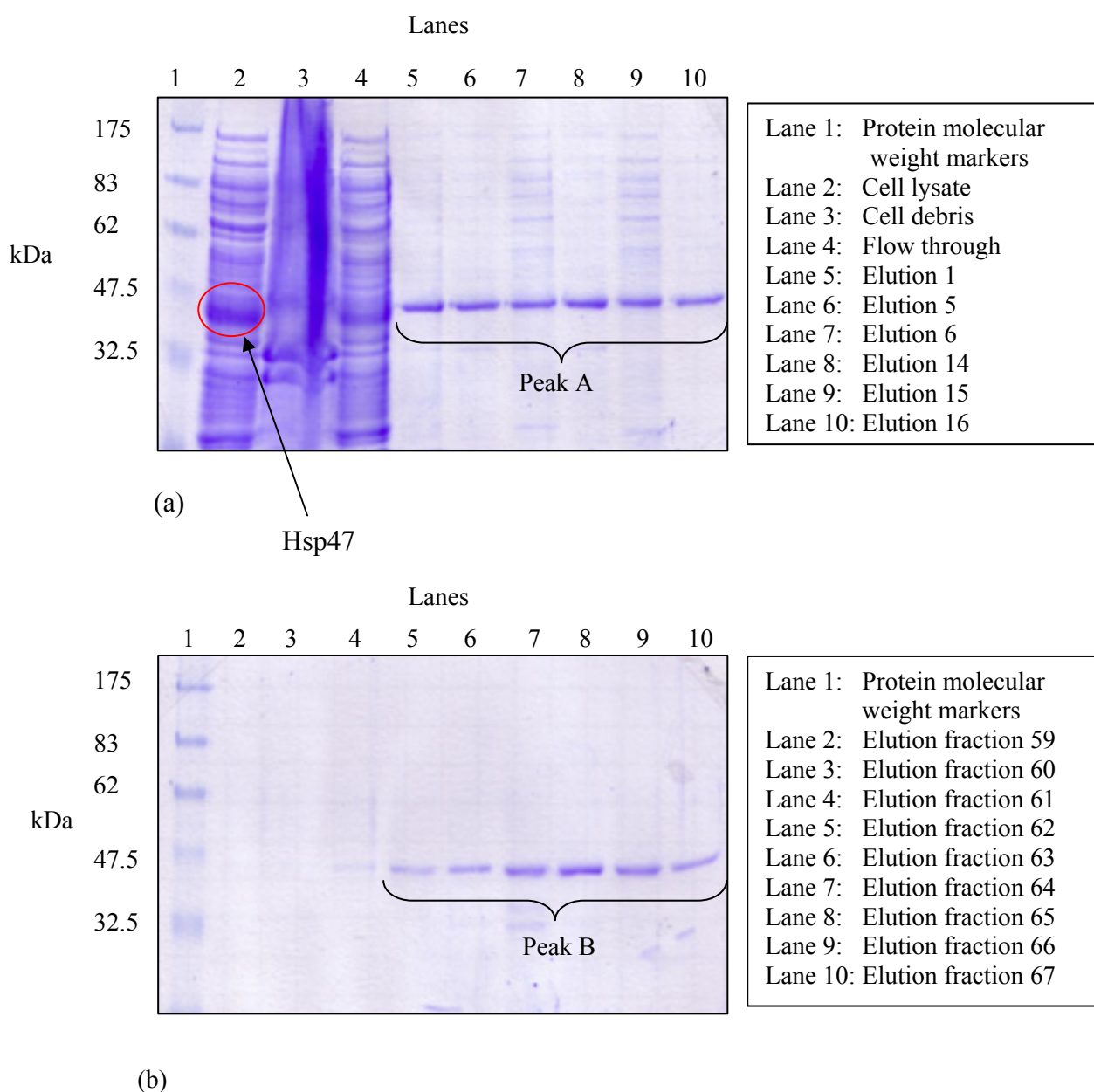


Figure 4.3. SDS-PAGE of purification of WT *Mus musculus* Hsp47. (a) Lanes 1-4 show fractions collected at each step of the purification procedure. Lanes 5-10 represent fractions corresponding to the salt dependent species (represented by the first peak in the absorbance curve). (b) Lanes 4-10 correspond to the pH dependent, salt independent protein species (represented by the second peak in the absorbance curve).

Analysis of the SDS-PAGE results show that most of the WT mHsp47 was present in the soluble form as there is a distinct band at 47 kDa in the cell lysate fraction (lane 1, Figure

4.3(a)) whereas the cell debris fraction (lane 3, Figure 4.3(a)) contained only a very small amount of insoluble Hsp47. The presence of Hsp47 in the flow-through fraction (lane 4, Figure 4.3(a)) suggested that not all of the soluble Hsp47 applied to the gelatin-agarose column bound and therefore was inactive. Lanes 5-10 (Figure 4.3(a)) illustrate protein fractions of the salt dependent, pH independent species. Figure 4.3(b) demonstrates protein fractions that were eluted upon reduction of the pH to 6.1; thereby being pH dependent. All the pH dependent fractions show a predominantly single band, indicating pure protein.

Western blotting was used to confirm that the protein seen by SDS-PAGE was Hsp47.

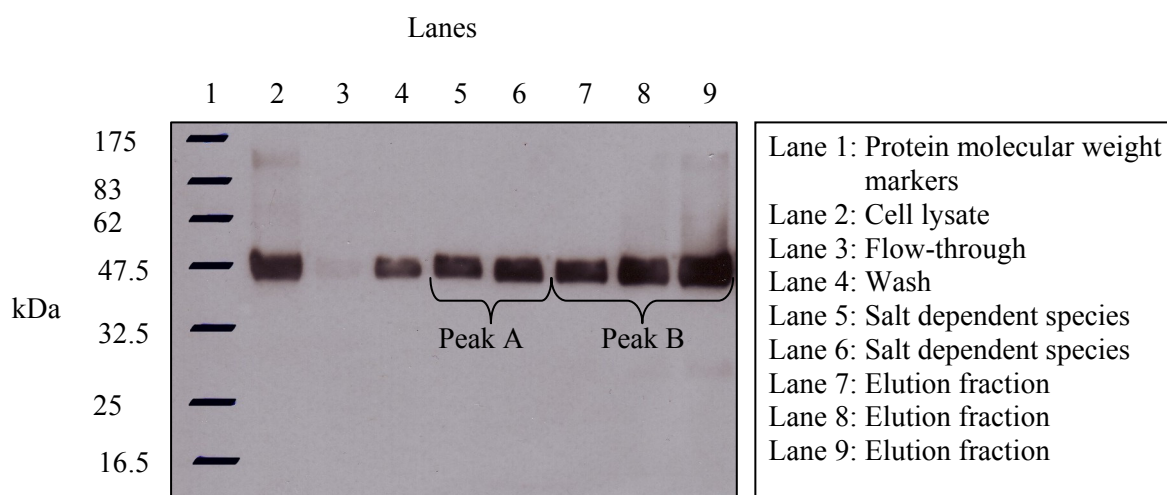


Figure 4.4. A western blot confirming the protein as Hsp47. The use of monoclonal, anti-Hsp47 antibody enabled the identification of Hsp47 in the protein samples. The presence of bands at 47 kDa confirms that the protein was Hsp47.

Structural characterisation of WT mHsp47

CD was used to determine the folded state and α -helical and β -sheet content of the protein. CD requires chiral molecules that have helical electron motion. The CD signal is due to the stereochemistry of the amide bond and other side chains. The difference in the magnitude of absorption of left- and right- handed circularly polarised light provides information on the

secondary structure of molecules. α -helical content is denoted by minima at 208 and 222 nm and β -sheet content is signified by a minimum at 216 nm. A positive signal at 200 nm represents a folded protein species.

A thermostatically controlled 0.2 cm pathlength quartz cuvette and a final protein concentration of 0.2 mg/ml were used for CD measurements.

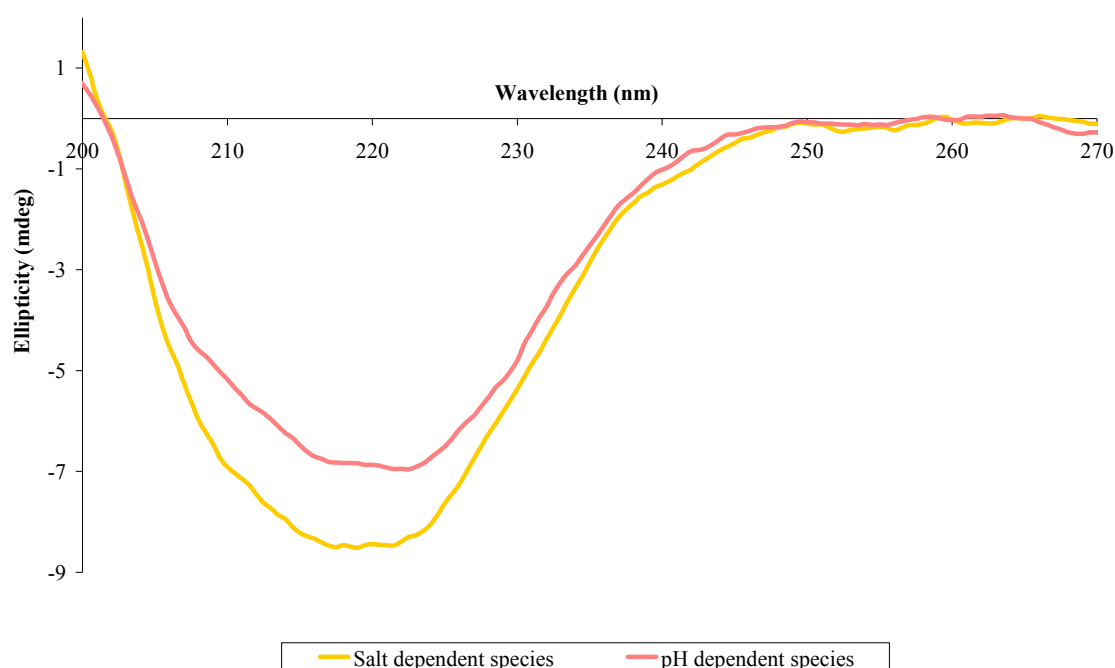


Figure 4.5. CD spectra for WT *Mus musculus* Hsp47. The data for both the salt dependent and pH dependent species show spectra that are characteristic of a folded protein. Both forms produced minima at 208, 222 and 216 nm, suggesting the presence of α -helices and β -sheets, respectively.

The stability of the protein was assessed using thermal denaturation by heating from 20 to 90 °C at a constant temperature of 1 °C/min. A CD signal at 222 nm was measured in order to detect any changes that occur in the α -helix structure. The change in the ellipticity was monitored during the heating process. Such stability experiments can distinguish between the subforms of the serpin superfamily, notably the five stranded A-sheet which has a melting point between 50-70 °C and the six stranded A-sheet that has a melting point exceeding 90 °C.

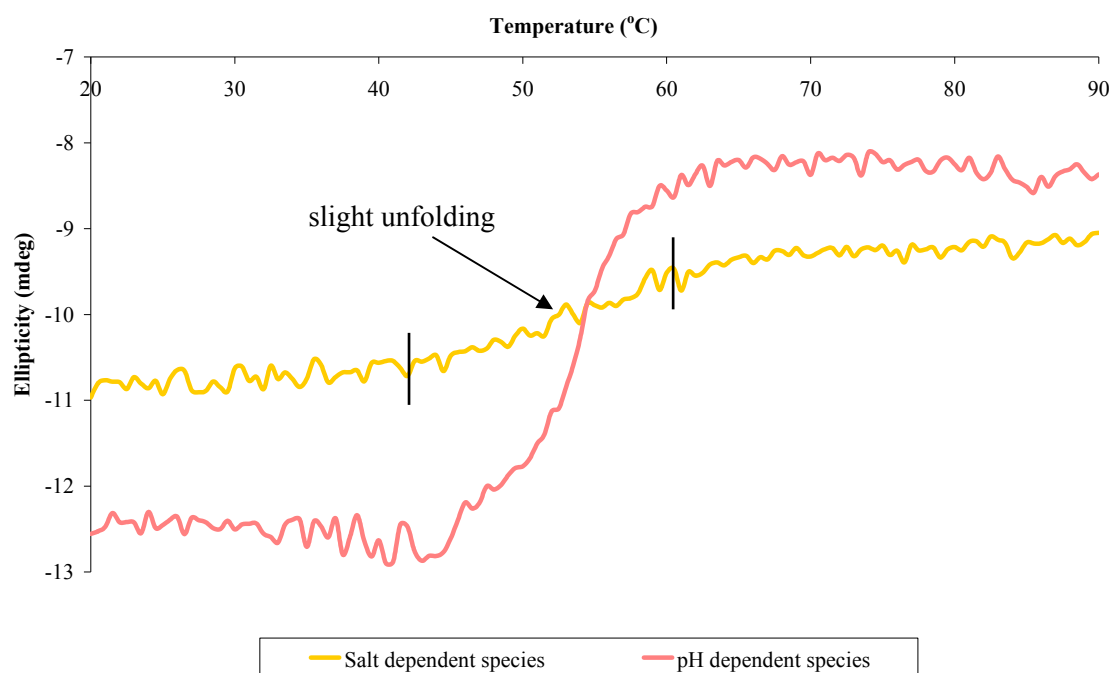


Figure 4.6. Thermal denaturation curves for WT *Mus musculus* Hsp47. The pH dependent species demonstrated a clear transition with a midpoint of 55 °C in comparison to the salt dependent species which showed a slight transition with an increase in temperature.

The CD data suggests that both the salt dependent and pH dependent forms are in the folded state and show spectra which demonstrate the presence of α -helix and β -sheet character. The thermal denaturation curves imply that the salt dependent species is of higher stability than the pH dependent species as it resists any major structural conformational rearrangement. This suggests that this species is in either the polymeric or latent serpin conformation. In the case of a polymeric conformation, the small transition observed between 50-60 °C may be due to destabilisation of the Hsp47 molecules at either end of a polymer. On the hand, if this species was to be in the latent serpin conformation, the destabilisation could be due to only partial insertion of the sixth β -strand into the A-sheet. Subsequently, the A-sheet would still lack the full complement of β -strands, making it less stable than an A-sheet containing six β -strands, and consequently reducing its resistance to destabilisation. The transition for the pH dependent species has a midpoint of 55 °C, denoting a five stranded A-sheet and thereby the

native serpin conformation. The metastable native fold is the biologically active form of a serpin due to the A-sheet lacking the full complement of β -strands.

Comparison of the CD data for *Xenopus laevis* (Figures 3.27 and 3.28) and *Mus musculus* Hsp47 (Figures 4.5 and 4.6) revealed the structural differences in the protein from each of the species. The spectrum for *Mus musculus* Hsp47 (Figure 4.5) illustrated more defined minima at 216 and 222 nm, which are characteristic of β -sheet and α -helix, respectively. In comparison to this, the spectrum for *Xenopus laevis* Hsp47 (Figure 3.27) did not show distinct minima at either 216 or 222 nm. The thermal stability curves for *Xenopus laevis* and *Mus musculus* Hsp47 (Figures 3.28 and 4.6, respectively) also exhibit differences. Whilst the pH dependent species of *Mus musculus* Hsp47 demonstrated a sharp transition with a midpoint at 55 °C and the salt dependent species revealed only a very slight unfolding, the denaturation curve of *Xenopus laevis* Hsp47 was intermediate between those of the two species of WT *Mus musculus* Hsp47.

AUC was used to determine the oligomeric state of each of the species by centrifugation of protein samples at 40,000 rpm for 8 hours at 4 °C. The protein distribution in each cell was measured by continuously monitoring protein absorbance at 280 nm. SEDFIT was used to fit the data and obtain a sedimentation distribution.

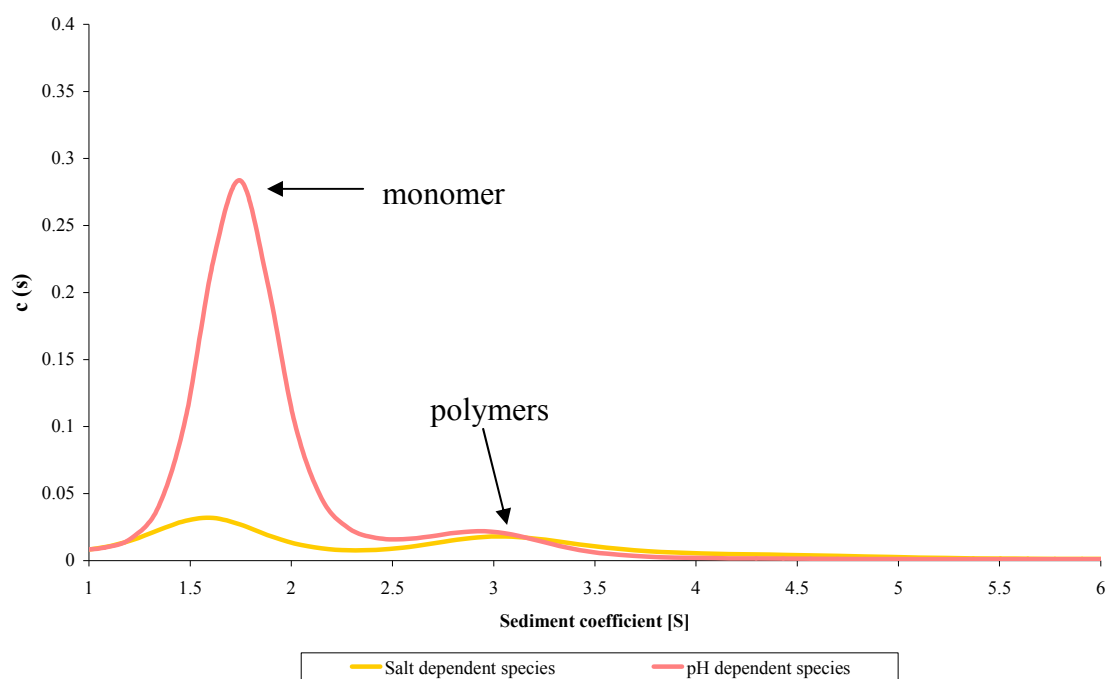


Figure 4.7. AUC for WT *Mus musculus* Hsp47. The fit of the data revealed that the salt dependent species produced two peaks of similar intensities, suggesting the presence of polymers. The pH dependent species produced a predominantly single peak at 1.7 S.

Analysis of the AUC data implied that the hyperstable, salt dependent species contains polymeric as well as monomeric species. The metastable, pH dependent species is predominantly monomeric (as illustrated by a single peak) with a sediment coefficient of 1.7 S. This is similar to that of archetypal non-inhibitory serpin ovalbumin that has a sediment coefficient of 1.8 S and a molecular mass of 45 kDa. Transformation of the data for this study allowed calculation of an approximate molecular mass of 45-50 kDa; in agreement with the mass of monomeric Hsp47.

Prior to the pH titration experiment, phosphoric acid was manually titrated into a 0.3 M phosphate solution at pH 9.3 to determine the volume and concentration of phosphoric acid required to reduce the pH of the solution to 6.0. The results were used to plot a titration curve (Figure 4.8) that allowed the identification of the points of inflection in relation to pH.

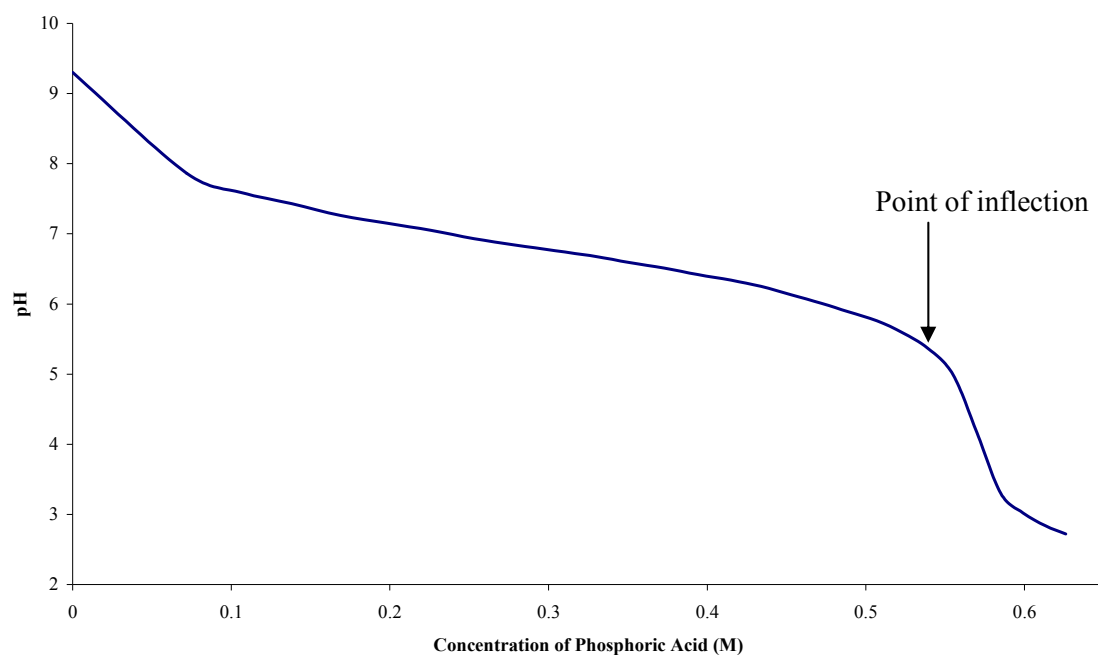


Figure 4.8. pH titration using phosphoric acid. The point of inflection is shown to be between pH 6.0 and pH 5.0.

The acid induced conformational change was probed by carrying out a pH titration experiment. This involved decreasing the pH of protein in solution from pH 8.0 to pH 3.0 by the addition of 2.5 μ l aliquots of 1.52 M phosphoric acid using an autotitrator. The change in structure of the protein was monitored by measuring the CD signal at 222 nm which is sensitive to a change in α -helical character.

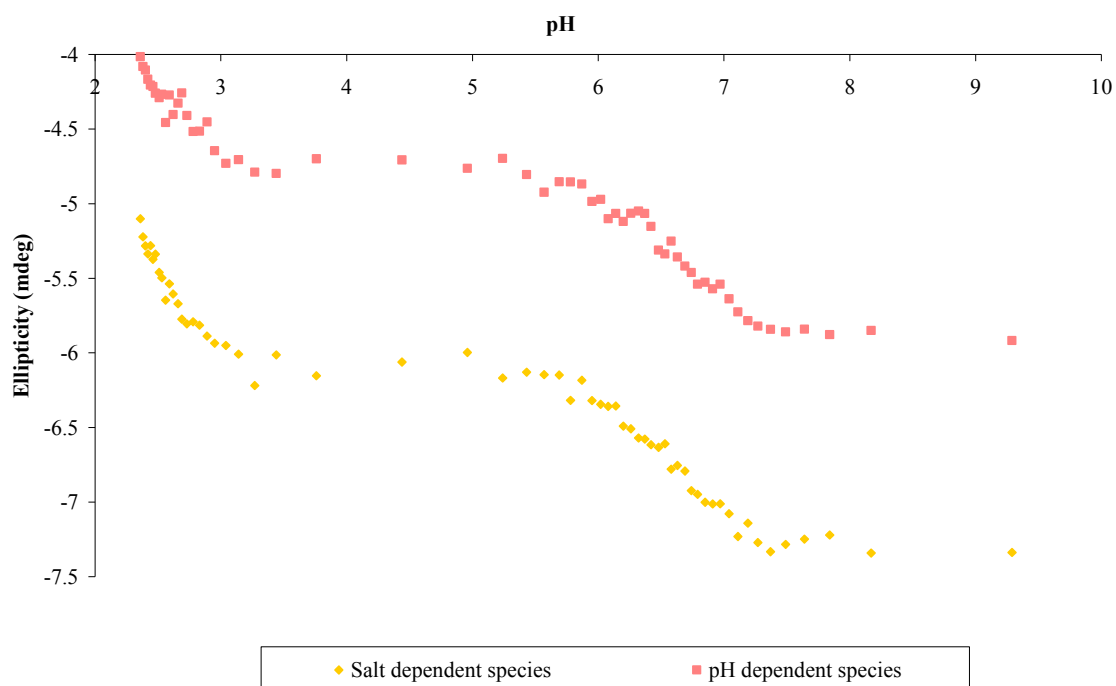


Figure 4.9. pH titrations of WT *Mus musculus* Hsp47. Both the salt dependent and pH dependent species gave a transition with a midpoint of pH 6.6, close to the pK_a of histidine residues. The transition between pH 2.0 and 3.0 is due to glutamate and aspartate residues.

The titrations illustrate that both species were able to undergo a structural transition upon a reduction in pH.

Biological activity determination

Collagen fibrillisation assays (outlined in section 2.2.13.5) were performed to establish the activity of the two forms of mHsp47 discovered. This involved investigating collagen fibril formation in the absence and presence of each of the species. The turbidity of the sample was monitored at 313 nm at 34 °C and any change would be indicative of a change in the fibrillisation of collagen. This was repeated at least three times for each species to check the reproducibility. Results for each of the species were extremely similar and repeats for the same mutant illustrated identically shaped curves.

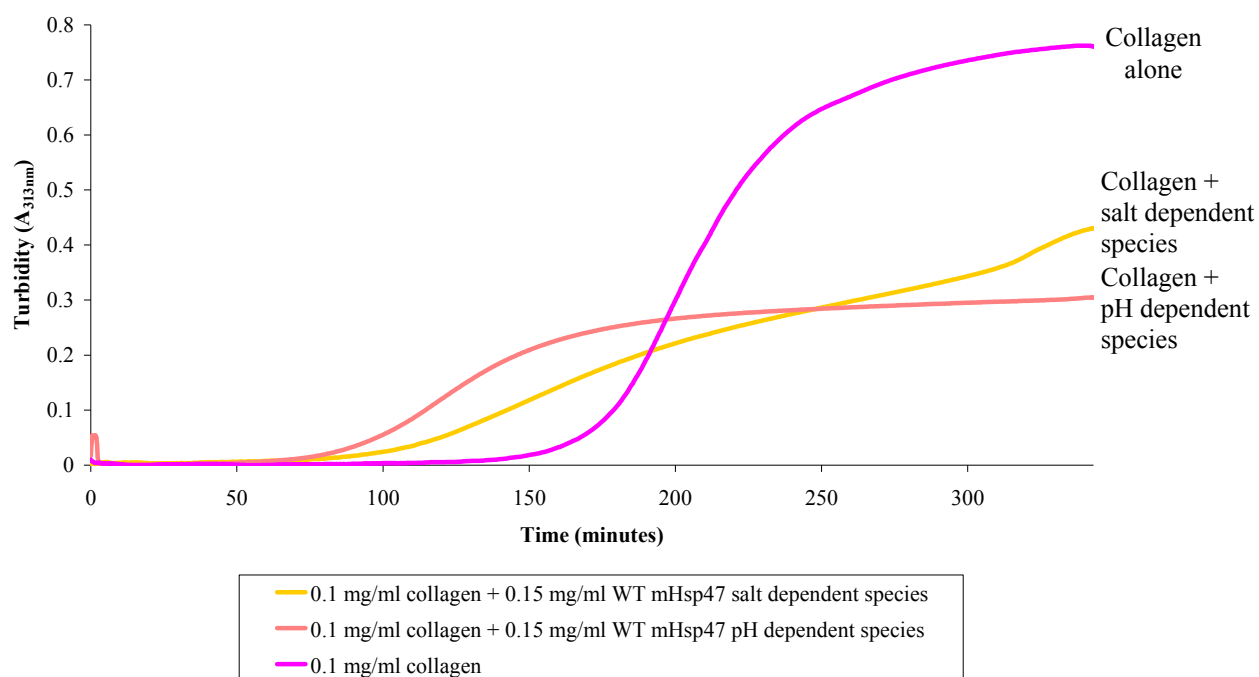


Figure 4.10. Effect of WT *Mus musculus* Hsp47 on collagen fibrillisation. The fibrillisation of 0.1 mg/ml collagen alone showed a large increase in scattering with time; reaching a maximum of 0.76 absorbance units. There was a decrease in the fibrillisation of collagen in the presence of the WT salt dependent species. A distinct decrease in collagen fibrillisation was also observed following the addition of the pH dependent species of WT mHsp47.

The fibrillisation experiments revealed some interesting results. As expected, the fibrillisation of collagen alone followed the typical collagen fibril formation pattern. Initially there was a nucleation (lag) phase (0-150 minutes) which was followed by a rapid growth phase (150-250 minutes) that preceded a deceleration phase (150-360 minutes) towards a maximum turbidity. In contrast to this, collagen fibrillisation in the presence of the polymeric WT salt dependent species had a reduced lag phase (0-100 minutes), a longer but less rapid growth phase (100-340 minutes) and a much shorter deceleration phase (340-360 minutes). The effect of the WT monomeric pH dependent species was different still; the lag phase was reduced further (0-70 minutes), the reduction in the growth of the fibrils was more pronounced (100-150 minutes) as the turbidity only increased from 0.05 to 0.24 in comparison to the 0.05 to 0.73 increase for collagen alone.

The above findings imply that both the salt dependent and pH dependent species of WT mHsp47 have effective collagen anti-fibrillisation activities, although the pH dependent species is more potent at doing so. More intriguingly, the lag phase is greatly reduced in the presence of Hsp47, suggesting that nucleation of collagen chains is facilitated by Hsp47. This suggests that Hsp47 is somehow involved in the bringing together of collagen chains but hindering the uncontrolled fibrillisation of these chains.

Electron Microscopy

Transmission electron microscopy was used to examine the topography and morphology of collagen fibrils in the absence and presence of Hsp47. Furthermore, the results from the collagen fibrillisation assays (Figure 4.10) could be compared to the electron microscopy observations to determine whether the findings from two different experiments correlate, providing strength to the proposed hypotheses. Prior to fibrillisation and in the absence of Hsp47 small, disordered collagen fibres were visualised (Figure 4.11(a)). Following fibrillisation of collagen alone, there was evidence of fibrils, however the majority of the sample had an aggregate-like morphology (Figure 4.11(b)). The fibrillisation of collagen in the presence of Hsp47 resulted in the formation of more ordered and longer fibrils (Figure 4.11(c)).

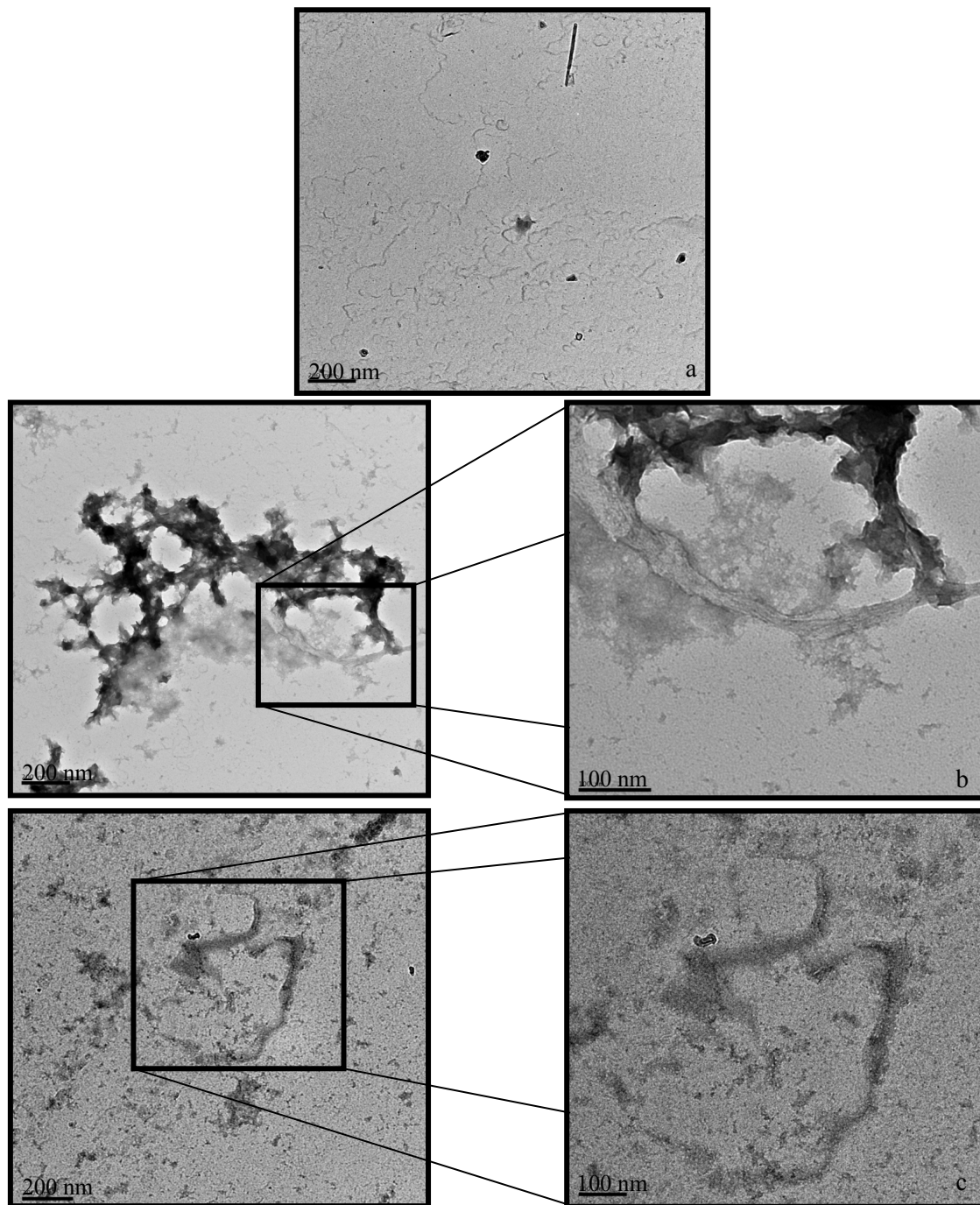


Figure 4.11. Electron micrographs of collagen fibrillisation. (a) Collagen pre-fibrillisation. The grid illustrated very few collagen fibrils alongside many small, randomly distributed fibres. (b) Collagen post-fibrillisation. Many bundles of collagen fibrils were observed as well as large aggregate-like structures. (c) Fibrillisation of collagen in the presence of WT mHsp47. WT pH dependent species and collagen were combined prior to the initiation of the fibrillisation assay. A reduced number of aggregate-like features and shorter, more ordered fibrils were visualised.

The electron microscopy results demonstrated that initially collagen comprised of small, unorganised structures which were converted to larger collagen-like fibrils after fibrillisation. This is in agreement with the results from the fibrillisation assays (Figure 4.10) that illustrated that the turbidity of collagen before fibrillisation was zero whilst the turbidity post-fibrillisation was 0.76 absorbance units; signifying the appearance of fibrils with time. Furthermore, comparison of Figures 4.11(b) and 4.11(c) showed that there was a reduction in fibril formation in the presence of Hsp47. This correlates with the fibrillisation results (Figure 4.10) that illustrated a decrease in turbidity in the presence of Hsp47 in comparison to collagen alone. This correlation between two different experiments strengthens the validity of the results obtained.

Although Figure 4.11 illustrates only single electron micrographs for each of the different fibrillisation time points, at least five fields were examined for each of the samples and similar features were visualised throughout.

4.3 Discussion

The protein purification procedure in this study involved a salt concentration gradient as well as a pH gradient. Monitoring of these conditions allowed the identification of two distinct mHsp47 species, a finding that has not been reported previously. Comparison of previous purification methods indicates that a difference in protein purification is likely to be the reason for this novel observation. For example, Nagata and coworkers [195] first isolated and purified Hsp47 from chick embryo fibroblasts using gelatin-affinity chromatography. Impurities were removed with 0.4 M salt prior to elution of gelatin-binding proteins by boiling in Laemmli's SDS gel electrophoresis sample buffer. This single-step elution produced a single elution peak that consisted of Hsp47 and other minor contaminants, notably one at 83 kDa. The high ionic strength wash buffer may have resulted in the release of the

weakly binding, salt dependent form of Hsp47. Therefore, the binding would have been reported as non-specific rather than weak and dependent upon the salt concentration. The use of a single, low pH elution buffer did not allow the exact pH of protein elution to be identified. The use of a pH gradient for protein elution, as used in this study is advantageous as it can be applied to any species of Hsp47 to determine the elution pH. It is particularly useful for the comparison of protein from different species or studying the effects (if any) of sequence mutations on the pH-dependency of Hsp47.

The purification of mrmHsp47 by El-Thaher and coworkers [97] involved affinity chromatography using collagen-sepharose followed by gel filtration. Although stringent ionic conditions (50 mM Tris-HCl, pH 8.0, containing 400 mM NaCl and 5 mM EDTA) were used for the washing of the column following the application of Hsp47, proteins present in the wash fraction were not reported. Again, a low pH, single elution buffer (50 mM Tris-HCl, pH 5.8, containing 150 mM NaCl and 5 mM EDTA) was used to elute Hsp47 and a single monomeric species was identified [97].

More recently, the pH dependent interaction of Hsp47 and collagen has been examined by Dafforn and coworkers [87] and Olerenshaw and coworkers (unpublished) [181]. As previously, Dafforn and coworkers used affinity chromatography for protein purification using collagen-sepharose. Following washing away of impurities under high ionic strength, once more a single low pH buffer was used for protein elution. Subsequently, gel filtration was used to separate pure protein from contaminant peptides. Consequently two oligomeric species were identified; a monomer at ~45 kDa and a trimer at ~147 kDa. This identification of two molecular species is similar to the finding of two pH dependent oligomeric forms of Hsp47 in this study. However, unlike in the study by Dafforn and coworkers, in this study the two species could be separated in one step using the method detailed below.

Previous work on the interaction of collagen and Hsp47 has revealed that the association/dissociation of the chaperone for its substrate is pH-dependent. Thus, a pH gradient was used in order to probe this pH-dependency and determine the point at which the Hsp47-collagen complex dissociates. Collagen affinity chromatography enabled pure, active Hsp47 to be obtained. The elution profile for WT mHsp47 illustrated a novel finding in that two species were present. The first was a salt dependent species that was eluted early in the purification and was dependent upon salt concentration. The second species was eluted late in the purification, at pH 6.1 and a constant salt concentration of approximately 7 mS/cm. This implied that the elution of this species was pH dependent and salt independent. The presence of two distinct species of this chaperone has not been reported prior to this study.

The study by El-Thaher and coworkers [97] of Hsp47 and collagen revealed that pH-dependent binding and release of procollagen/ collagen coincides with the pH-driven trans-conformational changes in Hsp47 indicating that these conformational changes alter the binding affinity of the chaperone for its substrate. CD and fluorescence allowed the identification of three structural states. The Alkali state (above pH 6.4) consisted of 25% and 36% α -helix and β -sheet, respectively. The Intermediate state (pH 6.2) comprised of 15% α -helix and 51% β -sheet and the transition from the Alkali to the Intermediate state showed marked changes in secondary structure. In the Acid state (below pH 5.7), the protein contained 11% and 57% α -helix and β -sheet character, respectively. These results suggest that a pronounced difference in secondary structure composition was indicative of the transition from the Alkali to Intermediate to Acid states. The results of collagen elution studies by El-Thaher revealed an elution maximum at pH 6.2. In relation to the results obtained in our study, this corresponds to the second peak which represents the pH dependent, salt independent species. The consistency of the elution pH of the pH dependent, salt independent species with other studies suggests that the method of a protein gradient is

reliable for monitoring the pH at which Hsp47 and collagen dissociation occurs. This all implies that the N-terminal His tag is not altering this aspect of Hsp47 behaviour.

The finding of the salt dependent, pH independent species is novel as a salt concentration gradient has not been used previously when attempting to separate Hsp47 from its substrate. The presence of two species that are released upon differing conditions makes for an interesting theory, in that there are two types of collagen binding Hsp47 forms. The first being dependent on salt concentration and the second upon the surrounding pH. Disruption of the interaction between the salt dependent species and collagen by an increase in conductivity to 7 mS/cm and a constant pH of 8.0 implies that this form is weakly binding. Once this species was eluted, even an increase in salt concentration to 20 mS/cm did not disrupt the interaction of the pH dependent species with collagen. Thereby, it can be said that this form binds collagen more strongly and the interaction involves bonds other than electrostatic/ionic interactions which are perturbed by salt. Covalent, hydrophobic and hydrogen bonds are those that are not affected by the presence of salt. Therefore, it is likely that the pH dependent species-collagen complex has additional or different bonds to the ones found in the salt dependent species-collagen complex. These stronger links between Hsp47 and its substrate collagen provide increased stability to the complex. Another possibility is that residues that are exposed in the pH dependent species, therefore able to interact with collagen, are buried in the polymer. Thus, they can no longer form bonds with the collagen α chains, resulting in weaker Hsp47-collagen association. The difference in exposed/buried residues between the monomeric and polymeric chaperone conformations could affect the binding position of Hsp47 on collagen. Furthermore, this inaccessibility to certain residues could result in reduced strength of Hsp47-collagen complex.

The results of the CD experiments suggested that the salt dependent species was folded and highly stable as it resisted denaturation when heated to 90 °C. This implied that it could not be in the native serpin conformation and was either the latent, cleaved or polymeric form. Multiple peaks from the AUC scan and a single band on SDS-PAGE implied that it was in the polymeric conformation. The sediment coefficient of the polymer is double that of the monomer, implying the presence of a dimer or trimer. The slight melting noted on the thermal CD titration (Figure 4.5) may be due to only partial insertion of the RCL of one molecule into the A-sheet of the second molecule. Such a structure was reported by Carrell and coworkers for the antithrombin dimer [158]. They reported the dimer as consisting of two molecules; one active and one inactive. The RCL of the active molecule was predicted as having a conformation compatible with the entry of two residues into the A-sheet of the inactive molecule whilst the inactive molecule was proposed to have a fully incorporated RCL, as in latent PAI-1 [158]. Results for the effect of the salt dependent species on collagen fibrillisation showed that it reduced fibril formation as the turbidity at $A_{313\text{nm}}$ at $t=360$ minutes diminished from 0.73 to 0.43. This reduction in collagen fibril formation is in agreement with earlier hypotheses that Hsp47 is involved in preventing the aggregation of collagen. The observation that the lag phase is reduced in the presence of the salt dependent species of Hsp47 implied that it not only has anti-fibrillisation activity but is also involved in the acceleration of the nucleation of collagen chains.

The electron microscopy results were in agreement with the findings of the collagen fibrillisation assays. Grids of collagen prior to the fibrillisation assay showed very few fibres. Fibrillisation of collagen produced disordered protein fibres that were distinct from those seen pre-fibrillisation. The presence of Hsp47 reduced the number of fibrils formed and these fibrils were more ordered than those formed in the absence of the chaperone. This is consistent with the finding of the collagen fibrillisation assays whereby Hsp47 reduced

collagen fibril formation (Figure 4.9) and is also in agreement with the hypothesis that the chaperone is involved in ordered association of collagen chains.

The CD results for pH dependent species revealed a folded protein that was metastable as it denatured when heated from 30 to 90 °C, with a midpoint inflection of 55 °C. This denotes the presence of the native serpin form. This is in agreement with the results obtained by Dafforn and coworkers for the native state of mrmHsp47. They demonstrated that the metastable form of the protein was ‘a typical α/β -protein’ that was folded and underwent a sharp thermal transition with a point of inflection at 54.7 °C [87]. A predominately single peak from the AUC scan and a single band on SDS-PAGE strengthened the finding of the pH dependent species being in the native conformation. As seen with the highly stable salt dependent species, collagen fibrillisation was distinctly reduced in the presence of the pH dependent species in comparison to collagen alone with a final A_{313nm} of 0.30. The native pH dependent form was more effective at de-fibrillisation of collagen than the polymeric salt dependent species. This may be due to the difference in the binding site on collagen and/or the type of interaction between the different Hsp47 species and collagen. In addition to this, like the hyperstable polymer, the metastable monomer reduced the nucleation phase of fibre formation.

The results show that the ability to reduce the nucleation phase is common to both forms of WT Hsp47 and each one is just as effective as the other. The anti-fibrillisation activity, on the other hand, although true for both forms, is greater for the metastable form. This may be due to the dimeric and monomeric nature of the salt dependent and pH dependent species, respectively. It could be that the monomer has more binding sites for the binding of collagen chains whereas in the polymer, some or part of these binding sites are inaccessible as they are being occupied/ blocked by the binding of one Hsp47 molecule to another during

oligomerisation. Monomeric Hsp47 would therefore have more points of contact with the collagen chains, thereby having enhanced collagen anti-fibrillisation activity in comparison to the polymeric, salt dependent species. It may also be that each of the species binds to the collagen molecules with different affinities which would in turn affect the binding of Hsp47 to its substrate. This would be consistent with the results of the elution profile (Figure 4.2) whereby disruption of the interaction between the hyperstable, polymeric species occurred upon increasing the salt concentration to approximately 4 mS/cm. In comparison to this, the metastable, pH dependent species required a reduction in the pH to 6.1.

To summarise, this study has revealed two distinct species of *Mus musculus* Hsp47; a salt dependent species and pH dependent species. The salt dependent form has been shown to be polymeric and hyperstable whilst the pH independent form was revealed to be predominantly monomeric and metastable. Both forms of the protein underwent a conformational change upon a reduction in pH. The presence of each of the species was able to reduce collagen fibril formation; however the pH dependent species was more potent at doing so than the salt dependent species.

CHAPTER 5: Probing the Importance of the Breach Region in the Structural Transition of Hsp47 by mutating H191, H197 and H198

5.1 Introduction

Unlike common globular proteins, where the native form is the most stable state, the serpin native form exists in a metastable state. It is this metastability that is thought to be critical for their biological functions. The serpin structure consists of three β -sheets surrounded by nine α -helices and an exposed reactive centre loop (RCL). For the inhibitory class of serpins binding of a target protease to the RCL results in its insertion into the A-sheet and the formation of a stable serpin-protease complex [122]. Mutational analyses of the archetype serpin, α_1 -AT, revealed unusual interactions, such as unusual side chain packing, unfavourable hydrophobic environments, buried polar groups and cavities as the structural basis of native metastability [21]. These interactions seem to have a destabilising effect on the interaction of helix F and the proceeding loop with the A-sheet. As the inhibitory activity of α_1 -AT was affected by mutations in the loop insertion region and not by mutations at the hydrophobic core, local stability rather than global stability was considered vital for the regulation of the inhibitory function of α_1 -AT [21, 22, 125, 196, 197].

Structural studies on the folding of plasminogen activator inhibitor-1 (PAI-1) have shown that it folds initially into an active conformation which slowly converts into a low-activity latent conformation [198]. This spontaneous conversion involves extensive conformational changes in both the proximal and distal hinge region of the RCL. P1-P1' residues on the RCL act as the pseudo-substrate for the corresponding proteinase [199]. At the N-terminal, the RCL is connected to s5A via the proximal hinge. At the C-terminal, the RCL is connected to s1C (the distal hinge region) [200]. Therefore the hinge domain is vital for the structural arrangement and functional transitions of serpins. The breach region lies adjacent to the proximal hinge

region and is also in close proximity to the distal hinge region (Figure 1.14). It is therefore likely that breach region is important in the structural rearrangement of serpins.

The denaturation and folding of active and latent PAI-1 were investigated using the EP8 mutant (T339E) [115]. This mutation in the RCL was designed to slow the kinetics of conversion from the active to latent state, based on the model that latency involved the formation of s4A [201]. The results demonstrated that folding into the active conformation is favoured over the latent conformation. Furthermore, latent PAI-1 retained its secondary structure upon denaturation whereas active PAI-1 underwent significant losses in secondary structure and was more susceptible to proteolysis near the A-sheet. It was therefore thought that the enhanced stability of the latent conformation may be due to a difference in secondary structure, specifically, the formation of s4A, loss of s1C and a surface helix [115].

Proteolysis or partial denaturation leads to conversion of the A-sheet from a 5-stranded, metastable state to a 6-stranded hyperstable form [87]. In addition to the native and latent forms, serpins can self-assemble into homopolymers via β -strand/ β -strand interactions [126, 158, 165, 178]. The biophysical properties of the polymers are determined by the type of linkage of the polymers. Polymers of α_1 -AT that are formed by linkages that involve the packing of the A-sheet are highly stable [119, 202].

Although mobility of the reactive loop is necessary for inhibitory activity, it is also involved in forming aberrant conformations that are associated with disease [123]. The Z, S_{iiyama} and M_{malton} variants of α_1 -AT are caused by the insertion of the RCL of one molecule into the β -sheet of a second molecule [126, 203, 204]. The resulting ordered polymers accumulate in the ER of hepatocytes subsequently leading to cirrhosis, juvenile hepatitis and hepatocellular

carcinoma [175, 205]. Angio-oedema, emphysema and thrombosis are associated with dysfunctional mutants of C1-inhibitor, α_1 -antichymotrypsin and antithrombin, respectively.

The expression of Hsp47 has been seen to correlate with increased collagen deposition, which subsequently leads to diseases such as myocardial infarction [206], arteriosclerosis [207, 208] and fibrosis [209-211]. Upregulation of Hsp47 expression has been seen in various animal models of fibrosis, including murine bleomycin-induced pulmonary fibrosis [212, 213], carbon-tetrachloride induced rat liver fibrosis [209] and peritoneal sclerosis in rats [214]. Furthermore, increased expression of human Hsp47 has been seen lesions of idiopathic pulmonary fibrosis [215], peritoneal sclerosis [216] and fibrotic transplanted kidney [217]. These findings along with the co-expression of Hsp47 with collagen suggest a role for Hsp47 in fibrotic diseases. Dafforn and coworkers discovered a trimeric form of mrmHsp47 and structural analysis demonstrated results that were consistent with those obtained for polymers of α_1 -AT and α_1 -antichymotrypsin that were formed by insertion of the RCL of one monomer as a fourth strand in the A-sheet of another monomer [87]. As this oligomerised form of Hsp47 maintained the ability to bind collagen, Dafforn and coworkers proposed that the binding site was not on the A-sheet but in a cleft between helices hA and hG/hH [87, 182].

The association and dissociation of Hsp47 and collagen is initiated in the lumen of the ER at neutral pH [218]. It is here where the Hsp47-collagen complex is formed. Collagen trimer formation leads to transportation of collagen chains with bound Hsp47 to the *cis*-Golgi compartment. The reduction in pH from the ER to the Golgi [219] triggers pH induced conformational changes in the chaperone and subsequent dissociation of Hsp47 from collagen. It has been previously proposed that histidine residues may be responsible for the pH dependent conformational mechanism of mrmHsp47 [87, 97]. The altered affinity of Hsp47 for collagen between pH 7.0 and 5.8 correlated with the protonation and deprotonation

of histidine which has a pK_a of 6.2. El-Thaher and coworkers reported a three phase model for the changes in structure, whereby Hsp47 underwent a 'two-step conversion' from an alkali to acid state via an intermediate state [97]. This pH dependent binding was also seen by Dafforn and coworkers during the investigation of the binding of Hsp47 to $(PPG)_{10}$. They noted a reduction in the binding following a drop in pH from 7.0 to 6.0. The change in pH from neutral to acidic is thought to promote a conformational change and consequently a diminution in the binding affinity of Hsp47 for collagen. This results in dissociation of Hsp47 from collagen and diminished biological activity [87]. El-Thaher and coworkers constructed single (H335L) and double (H197Q: H198N) site-specific mutants, both of which had irregular biological activities. The single mutant remained collagen bound even at pH 6.1 whereas the double mutant lacked any collagen binding properties at all [97].

Examination of the sequences for Hsp47 from species that have been sequenced identified fourteen histidines residues that are conserved in 90% of Hsp47 sequences. Of these fourteen histidines, five are conserved within the serpin superfamily, making them less likely to be important specifically for Hsp47. The remaining nine histidines are found in three clusters around the Hsp47 molecule. The cluster located near the top of the A-sheet (Figure 5.1 below) is of particular significance as this is the region of serpins that is involved in conformational changes involving insertion of the RCL into the A-sheet. The three histidines in this cluster are H191, H197 and H198. Histidine to alanine amino acid substitutions were designed in an attempt to better understand the structural basis and functional role of the non-inhibitory serpin, Hsp47.

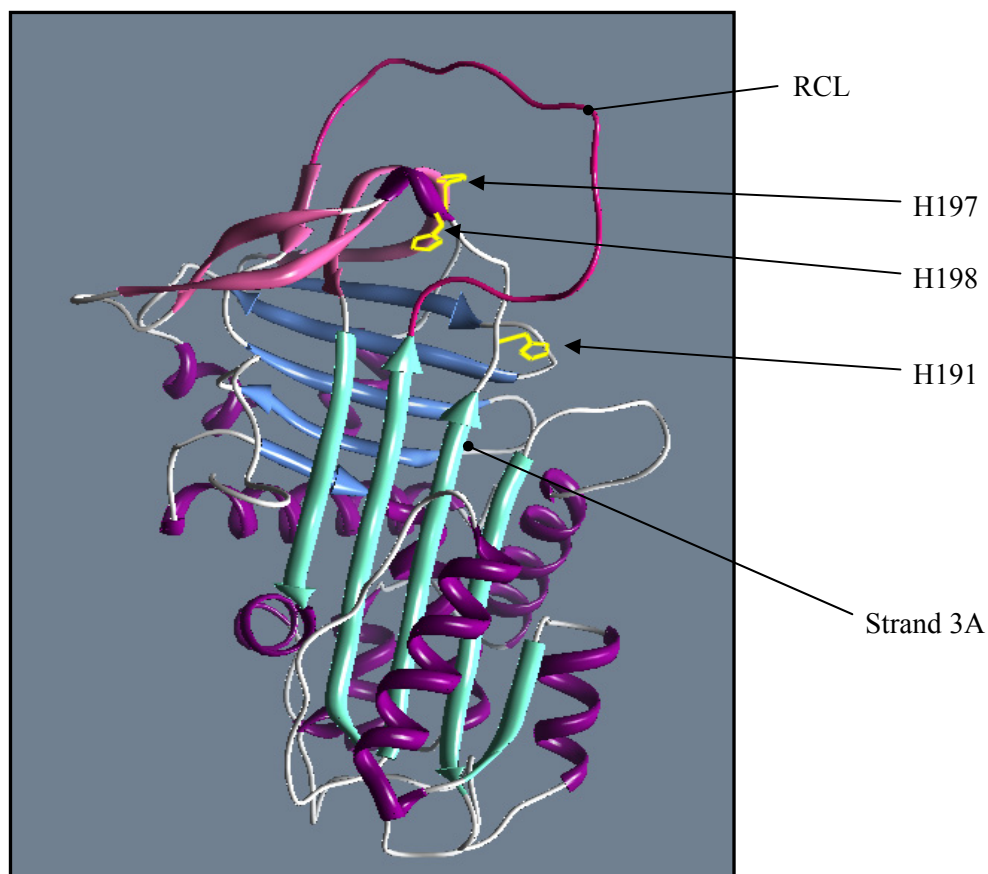


Figure 5.1. Model of native Hsp47 illustrating mutated histidine clusters.

This chapter explains three mutations (H191, H197 and H198) that were constructed to further study the importance of the histidine cluster in the breach region on Hsp47 structure and function. These mutations were made on the *Mus musculus* Hsp47 6xHis template described in chapter 4.

5.2 Results

H191A

Protein purification

As with WT Hsp47, each of the mutants was expressed in *E. coli* and purified using collagen-affinity chromatography using a phosphate buffer pH gradient from pH 9.2 to 3.0. The

absorbance, salt concentration and pH were monitored constantly throughout the purification procedure and an elution profile plotted.

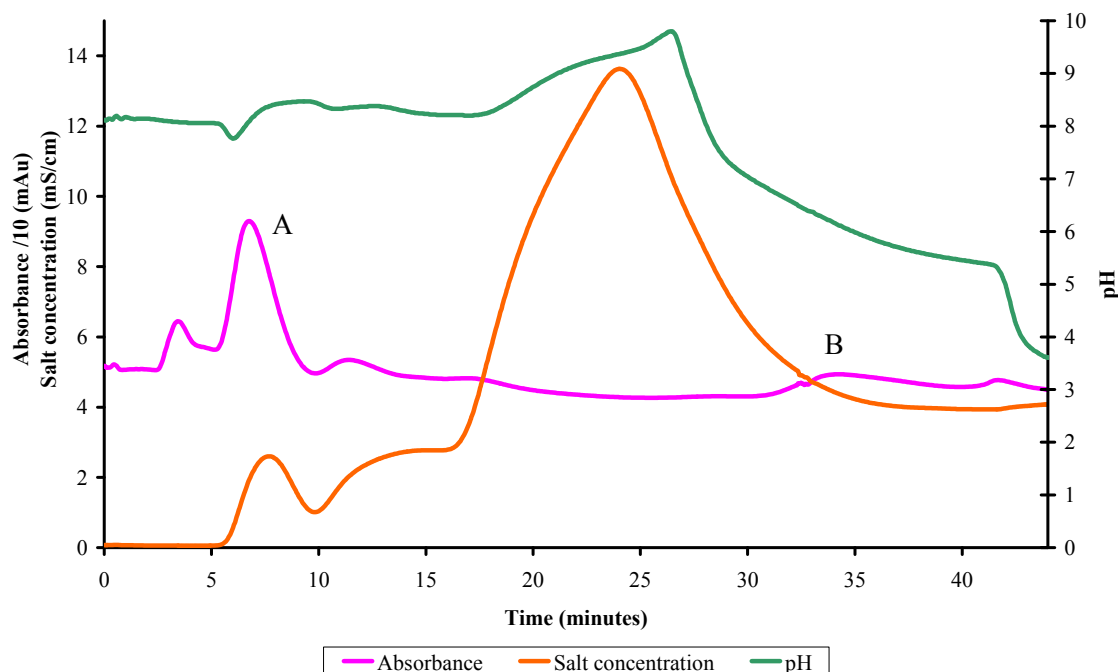
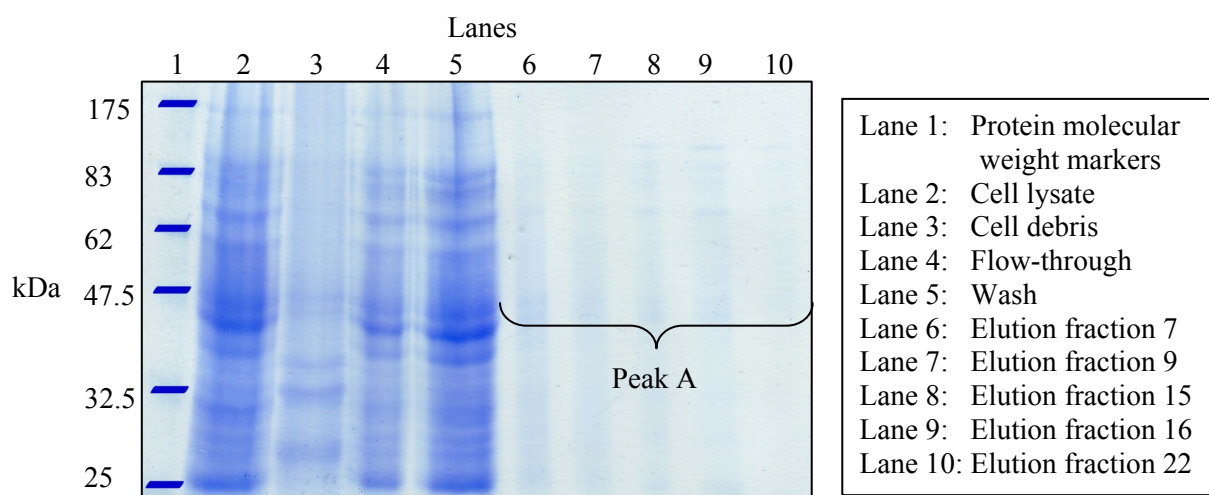


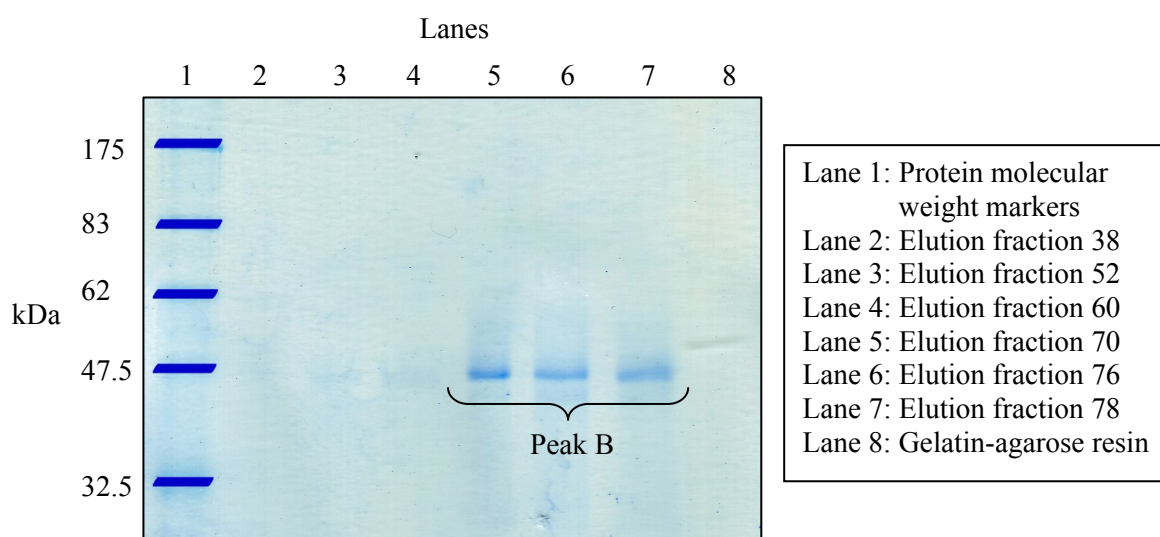
Figure 5.2. Elution profile of H191A mHsp47. The absorbance curve shows two different sets of peaks, each of which represent different forms of Hsp47. Peak A symbolises protein that is elutes by an increase in salt concentration (orange curve), thus a salt dependent species whereas the peak B illustrates protein that elutes upon a reduction in the pH (green curve), namely a pH dependent species.

The presence of two sets of absorbance peaks signifies the existence of multiple species of Hsp47. Peak A denotes protein that elutes at a relatively constant pH and an increase in salt concentration. This therefore designates a salt dependent, pH independent species. The peak of the elution of this salt dependent species correlates with the peak for the salt concentration. Peak B corresponds to a second species of Hsp47, elution of which occurred upon a reduction in the pH, during which the salt concentration was being reduced from its maximum. These findings suggest that it is a pH dependent, salt independent species.

Samples were collected at steps during the purification of the protein in order to monitor the path taken by the protein and to determine the point(s) of elution of Hsp47.



(a)



(b)

Figure 5.3. SDS-PAGE for H191A mutant. (a) Lane 1 illustrates that some Hsp47 was produced as there is a band at 47 kDa. Occurrence of bands at 47 kDa in lanes 4 and 5 demonstrate that there was some protein that did not bind (hence was eluted in the flow-through) and some that bound weakly (hence being eluted in the wash fraction). Lanes 8 and 9 represent the species that eluted in the first peak of the purification. (b) The protein bands present in lanes 5, 6 and 7 designate protein that was eluted in the second peak. Lane 8 shows protein that remained on the gelatin agarose beads following completion of the purification.

Results of SDS-PAGE showed that most; if not all the Hsp47 being produced was in the soluble fraction due to the absence of a 47 kDa band in the cell debris lane (lane 3, Figure 5.3 (a)). The bands in lanes 8 and 9 of Figure 5.3 (a), albeit very faint, correspond to the salt dependent, pH independent species that was eluted between 5-10 minutes. The intensity of the bands implies that only a small amount of protein being produced by the H191A Hsp47 mutant was the salt dependent form. The bands in lanes 5-7 of Figure 5.3 (b) correspond to the pH dependent species that was eluted at pH 6.1.

Structural characterisation

CD and AUC were used to determine secondary structure composition, stability and oligomerisation state, respectively. pH titration experiments were performed to detect the ability of the protein to undergo a conformational transition upon a reduction in pH. The biological activity was determined by collagen fibril formation assays in the presence of variant Hsp47.

As with WT, a 0.2 cm quartz cuvette was used for all CD measurements. The protein concentration was 0.2 mg/ml.

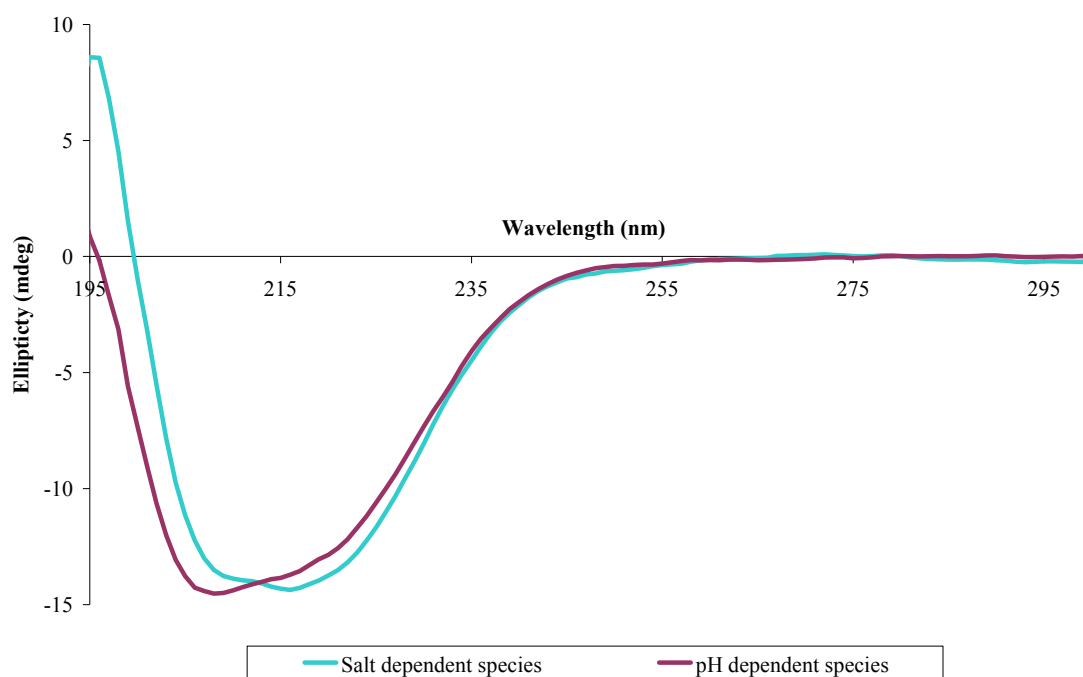


Figure 5.4. CD spectra for H191A mutant. Spectra for both species are typical of that for a folded protein. Minima at 208 and 222 nm are typical for α -helical structures and a minimum at 216 nm denotes the presence of β -sheet character.

Thermal denaturation was used to determine protein stability. Changes in the CD signal were monitored whilst the samples were heated from 20 to 90 °C.

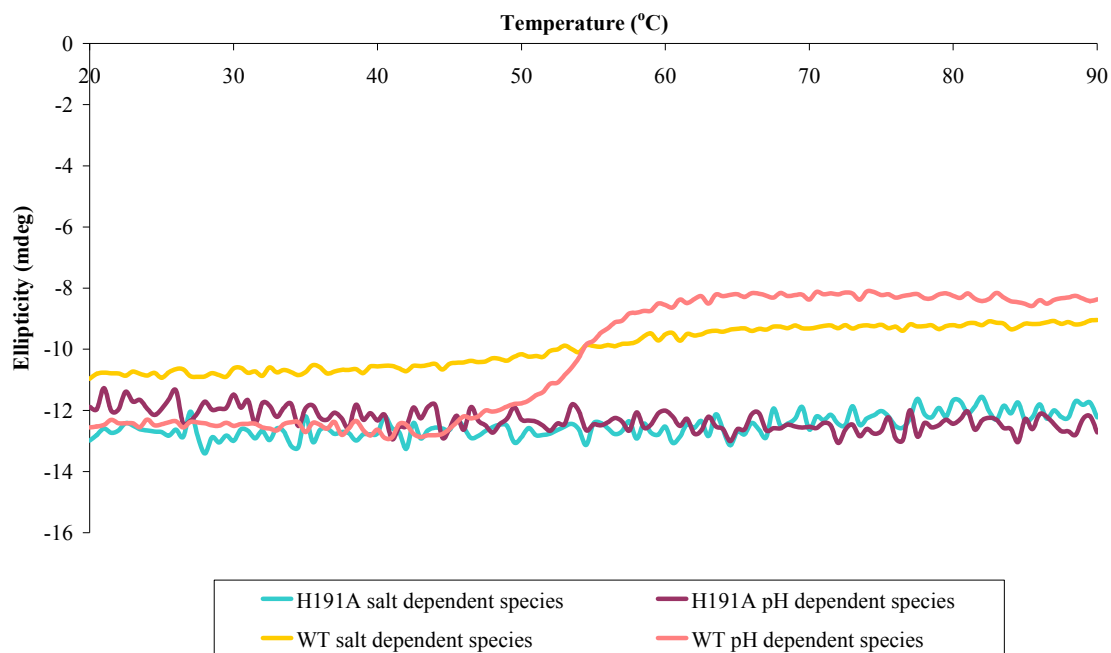


Figure 5.5. Thermal stability curves for H191A mutant. Both the salt dependent, pH independent and the pH dependent, salt independent species show that there is no change in structure upon heating from 20 to 90 °C.

The CD results demonstrated that the salt dependent and pH dependent species were folded with α -helix and β -sheet character. Figure 5.5 illustrates that the two forms of Hsp47 produced by the H191A mutant were highly stable as they were resistant to a conformational transition during the thermal denaturation experiment. This indicated that these species are either in the polymeric or latent conformations.

AUC was used to clarify which of the hyperstable forms was present for each of the species. The samples were centrifuged at 40,000 rpm for 8 hours at 4 °C and the protein absorbance at 280 nm was monitored.

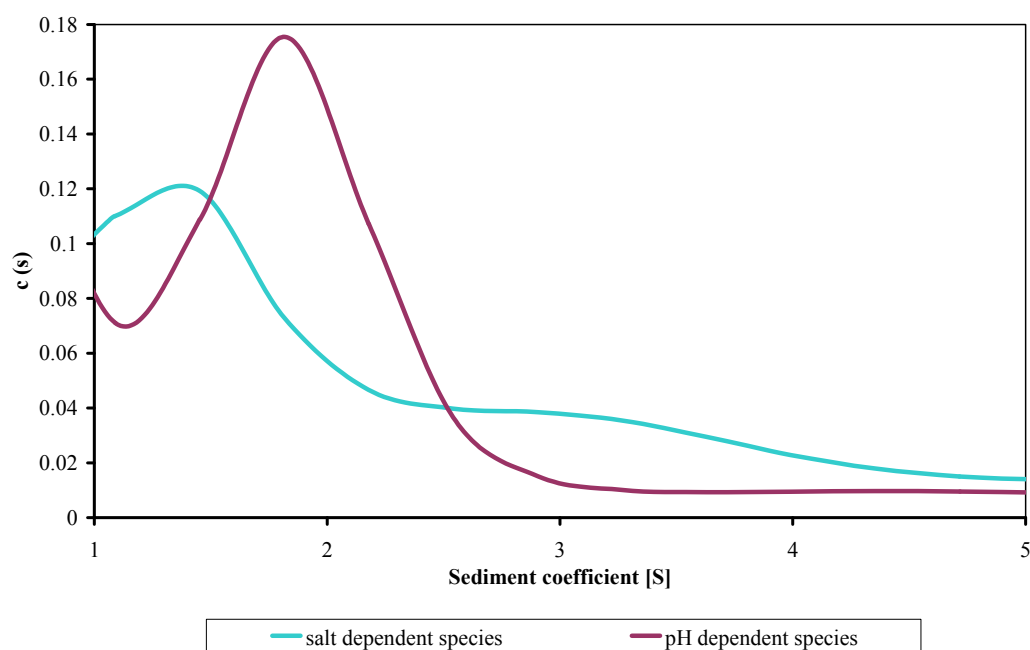


Figure 5.6. AUC results for H191A mutant. The data fit illustrates that the salt dependent species produced one distinct peak and a smaller broad peak. This is indicative of the presence of protein with more than one molecular mass. The pH dependent species produced a single peak, indicative of a monomeric species.

The salt dependent species illustrated two peaks, one at approximately 1.6 S and a second at 3.2 S. It is therefore possible that the first peak represents a monomer whilst the second peak signifies a polymeric species. These data are in agreement with the AUC data for the salt dependent species of WT mHsp47, which revealed two peaks at similar positions. The pH dependent species illustrated a single peak at approximately 1.7 S, consistent with the pH dependent species of WT mHsp47 (Figure 4.7).

Understanding the pH dependence of each of the Hsp47 mutants will provide an insight into Hsp47 function. The pH sensitivity of Hsp47 is thought to be mediated by the ionisation of histidine residues thus a shift in the pH dependency of conformational change in the mutants would highlight any histidines that are important in controlling the conformational change. The six-stranded latent conformation adopted by the pH dependent species stabilises the

central β -sheet. Involvement of the A β -sheet in the pH dependent conformational change would perturb the extent of the conformational change and/or its pH dependency.

The pH titration experiment was thus performed to investigate this pH dependency of conformational change. The CD signal was monitored, by measuring change in α -helical content (at 222 nm), whilst the pH of the protein solution was decreased from pH 8.0 to 3.0.

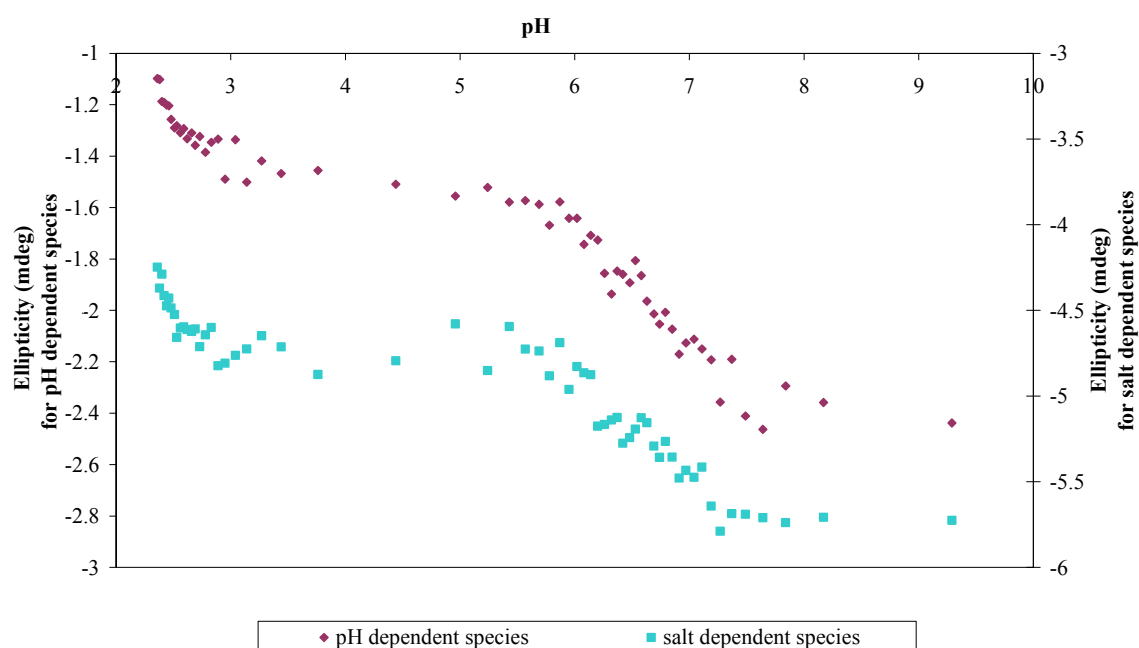


Figure 5.7. pH titrations of H191A mutant. The pH was reduced from 9.0 to 3.0 by the addition of 2.5 μ l aliquots of 1.52 M phosphoric acid. Both the salt dependent and pH dependent species showed a characteristic conformational change with a midpoint of 6.6.

Each of the species for the H191A mutant was able to undergo a conformational change, suggesting that mutation of the histidine at position 191 does not affect the pH dependency of the protein. This is the first instance in which the conformational change has been observed for the stabilised (latent) conformation. WT polymers were also shown to undergo this pH dependent conformational change.

The collagen fibrillisation assay was performed to test the ability of the two species of the H191A mutant to arrest collagen fibrillisation.

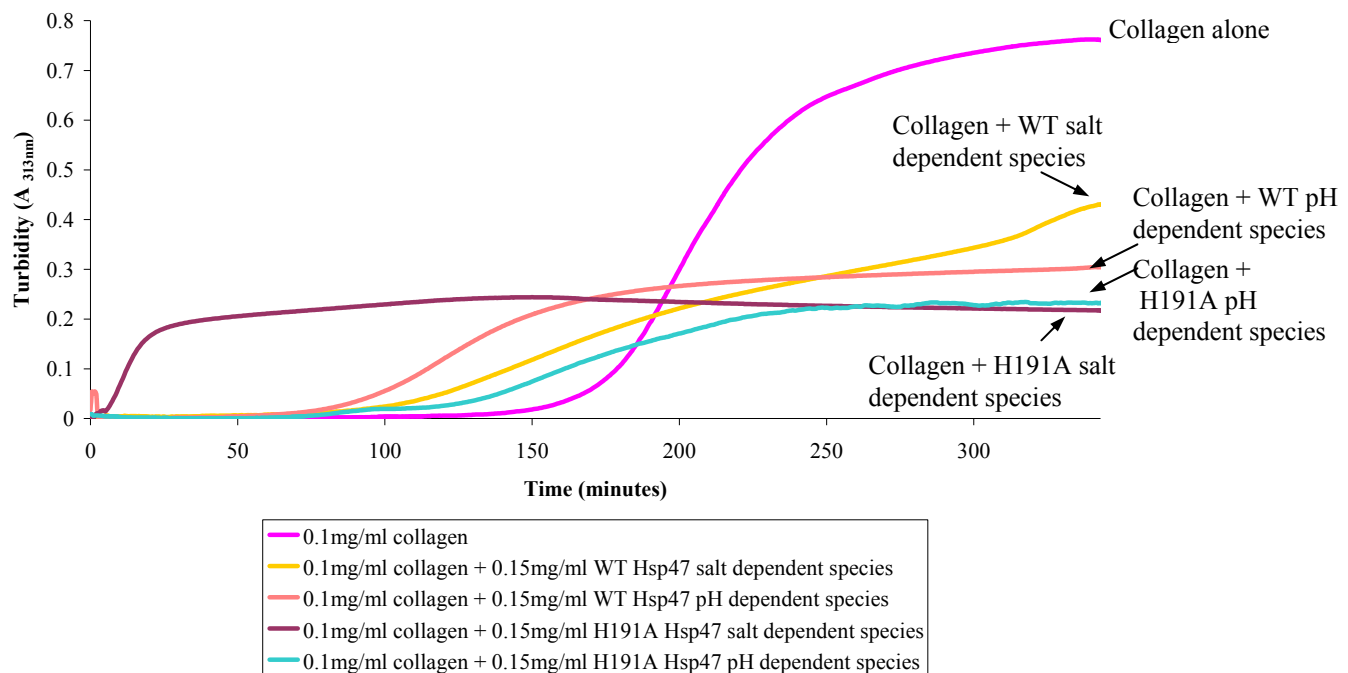


Figure 5.8. Effect of H191A mutant on collagen fibrillisation. The salt dependent species of the H191A mutant effected the total collagen fibre formation greatly as the final turbidity was similar to that of collagen fibrillisation in the presence of WT pH dependent species. The pH dependent species of the mutant was equally as potent as the salt dependent species at preventing collagen fibrillisation as it reduced the turbidity as much as the former.

The salt dependent species of the H191A mutant was more active than the WT salt dependent species as the turbidity of the final sample (at $t=360$) was reduced further in the presence of the variant than the WT equivalent. More interestingly, however, mutation of the histidine at position 191 appeared to affect the lag phase of collagen fibrillisation, reducing it considerably in comparison to collagen alone. Consistent with WT results, the mutant H191A pH dependent species possessed anti-fibrillisation activity of collagen but unlike WT Hsp47, the salt dependent species was just as effective at reducing collagen fibril formation as the pH dependent species. This implied that the mutation of H191 not only enhances the biological activity of the pH dependent species but also has a positive effect on the activity of the salt

dependent species. Moreover, collagen fibril formation in the presence of the latter species comprised a very small lag phase, suggesting that this variant somehow decreased the nucleation time before fibre formation was initiated.

H197A:H198A

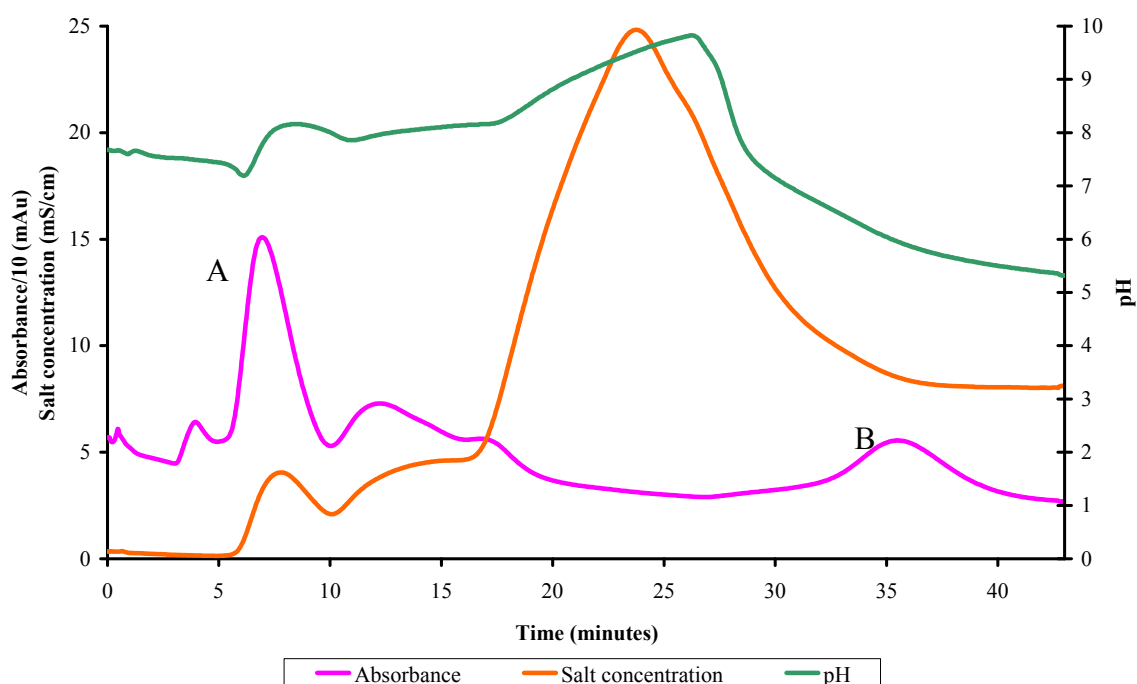


Figure 5.9. Elution profile for H197A:H198A mutant. Purification using a pH and salt concentration gradient revealed two distinct peaks of protein elution. The first peak represents protein that eluted upon an increase in salt concentration whilst the second peak denotes protein that required a reduction in the pH for dissociation from collagen.

As with the WT and H191A mutant, the H197A:H198A double mutant of mHsp47 yielded two species of the protein. A salt dependent species; elution of which was parallel to an increase in salt concentration and a pH dependent species; elution of which correlated with a reduction in pH. Once all the salt dependent protein was eluted, a further increase in the salt concentration had no effect on the binding of the pH dependent protein to collagen. This suggests that the pH dependent species was independent of salt concentration. Likewise, the salt dependent species does not depend on the pH as application of the pH gradient prior to the salt concentration gradient did not result in protein elution.

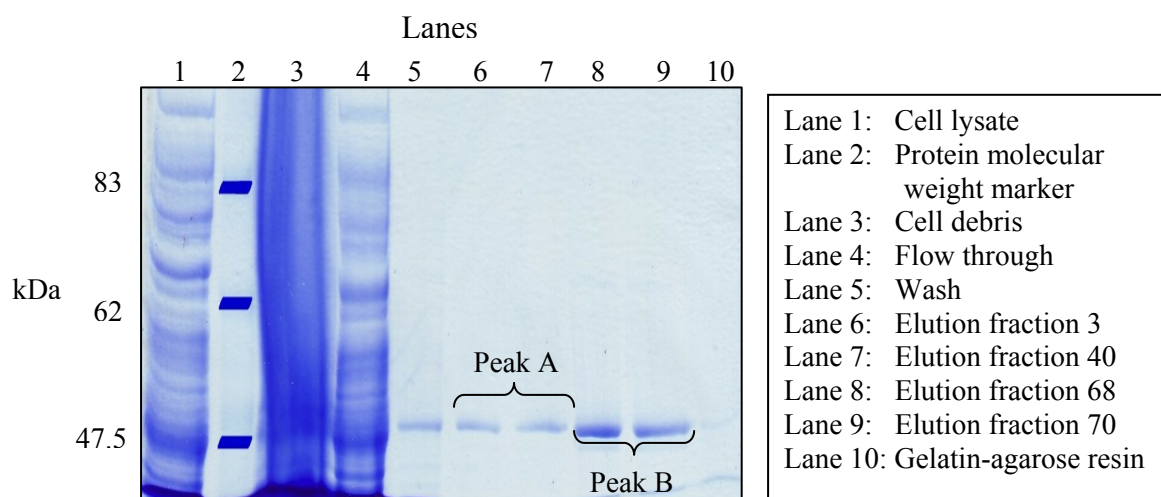


Figure 5.10. SDS-PAGE for H197A:H198A mutant. Lane 1 shows a band at 47 kDa, demonstrating the production of Hsp47. Small amounts of protein were present in the flow through and wash fractions (lanes 4 and 5, respectively). Lanes 6 and 7 represent protein that was eluted in the first absorbance peak of the elution profile while lanes 8 and 9 represent protein that eluted in the second absorbance peak.

The results of SDS-PAGE illustrate that the H197A:H198A mutant yielded a significant amount of soluble Hsp47 (Figure 5.10, lane 1). Absence of a band at 47 kDa in the cell debris fraction (Figure 5.10, lane 3) indicates the lack of insoluble material. Hsp47 in the flow-through and wash fractions (Figure 5.10, lanes 4 and 5, respectively) shows that some of the protein did not bind and some was weakly binding as it eluted when washing the gelatin column with water. Salt dependent Hsp47 is present in lanes 6 and 7 and pH dependent Hsp47 is present in lanes 8 and 9. Lane 10 shows that there was no protein retained on the gelatin-agarose beads following completion of the purification procedure.

Structural characterisation

Determination of the structural information on each of the species for the H197A:H198A mutant used CD and AUC.

CD was used to establish the secondary structure composition by measuring the CD signal between 105-300 nm and analysing the resulting spectra. The stability of each of the species was investigated by monitoring the change in CD signal whilst heating from 20 to 90 °C. A 0.2 cm quartz cuvette and a protein concentration of 0.2 mg/ml were used for all CD measurements.

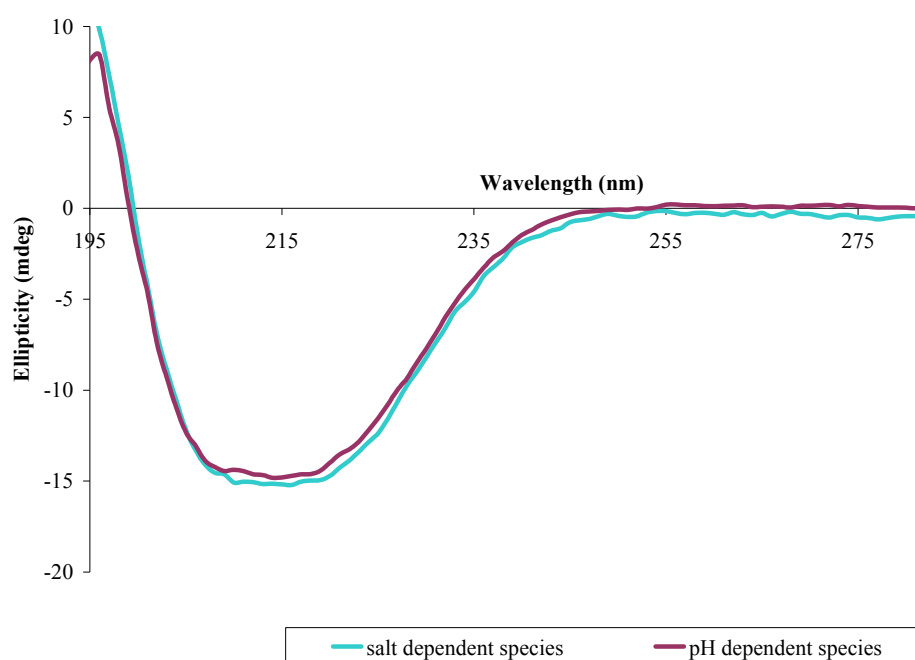


Figure 5.11. CD spectra for H197A:H198A mutant. Both the salt dependent and pH dependent species illustrate spectra with a positive signal at 200 nm and minima at 206, 218 and 222 nm typical for a folded protein containing α -helix and β -sheet character.

The spectra revealed that both the salt dependent and pH dependent forms of the H197A:H198A variant were folded α/β proteins.

Thermal denaturation was used to determine protein stability. This involved heating of the protein and constantly measuring the CD signal. Transformation of the data into graph form allowed any conformational transitions to be distinguished as this would be accompanied by a large change in CD signal.

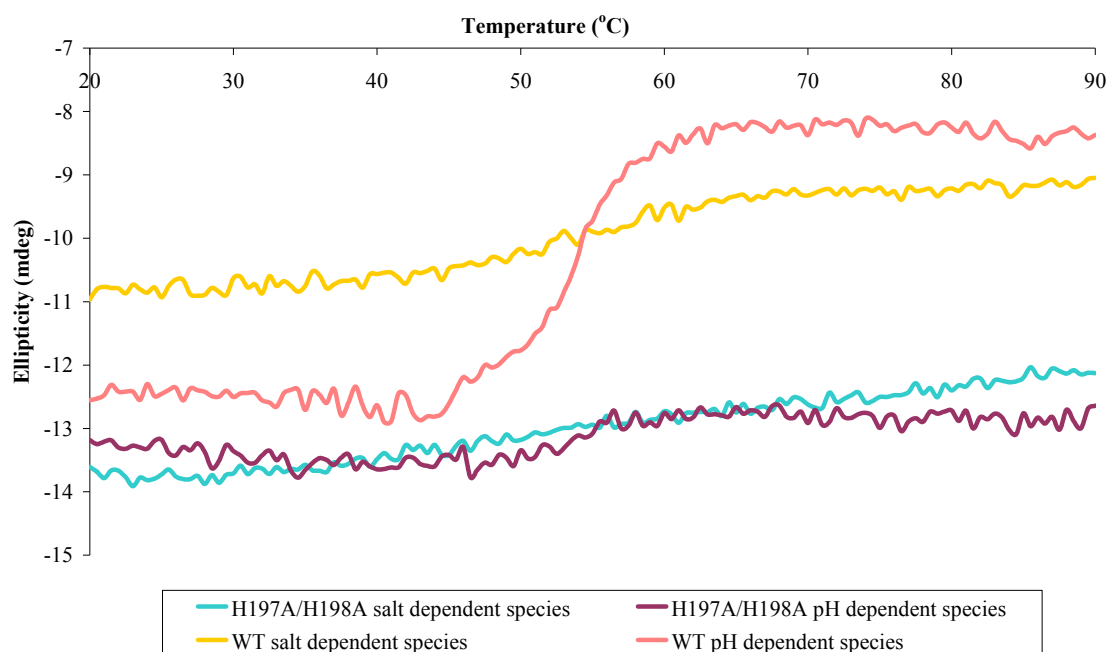


Figure 5.12. Thermal stability curves for H197A:H198A mutant. The CD signal for both species remained relatively constant when heated from 20 to 90 °C. No distinct transition can be seen from the results above.

The CD results (Figures 5.11 and 5.12) indicate that both the salt dependent and pH dependent species were folded with minima at 208 and 222 nm for α -helical composition and a minimum at 216 nm for the presence of β -sheets. Both forms were hyperstable as they resisted a distinct conformational change upon being heated to 90 °C. These findings suggest the presence of the polymeric or latent form of Hsp47. It cannot be the cleaved form as SDS-PAGE revealed a single band, signifying a single peptide.

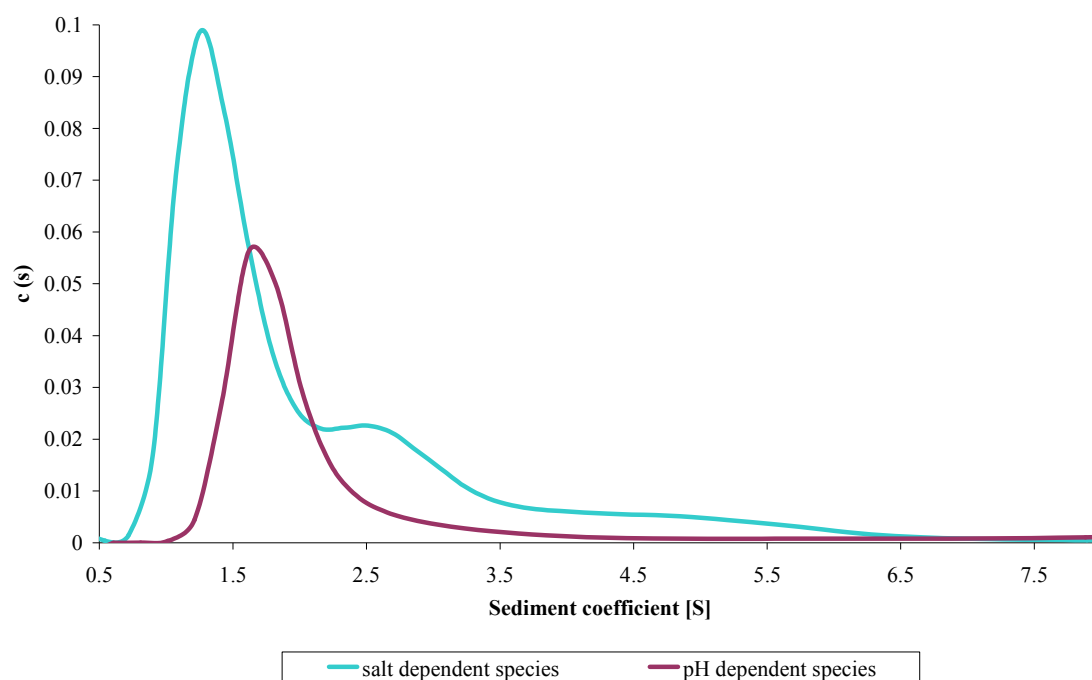


Figure 5.13. AUC results for H197A:H198A mutant. The salt dependent species revealed a curve with a predominant peak at approximately 1.5 S and a smaller peak at 2.6 S. The pH dependent species produced a single peak at 1.7 S.

The AUC results suggest that the salt dependent species comprised a mixture of monomeric and oligomeric protein whilst the pH dependent species was monomeric.

Taken together, the CD and AUC data imply that the salt dependent species was folded, hyperstable and oligomeric; all characteristics that are typical for a polymer. The pH dependent species was shown to be folded, hyperstable and monomeric; distinctive of the latent serpin conformation.

The pH dependent conformational change of the latent and polymer forms of H197A:H198A mHsp47 was probed using the pH titration method.

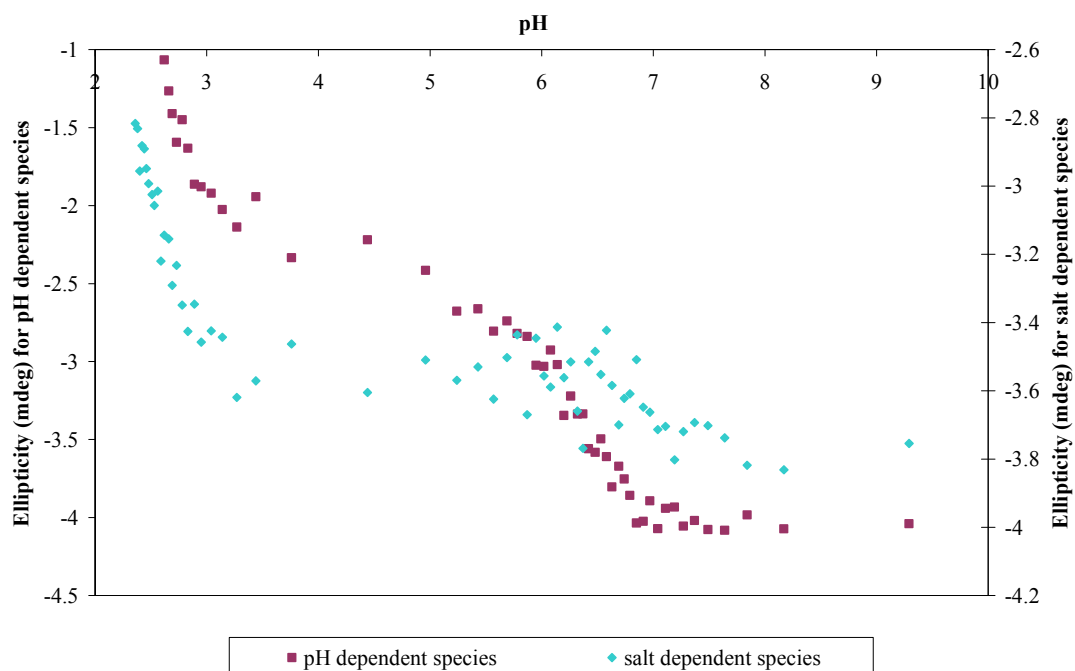


Figure 5.14. pH titrations of H197A:H198A mutant. Both species showed a conformational transition with a midpoint of pH 6.6.

Mutation of the histidines 197 and 198 did not affect the ability of the protein to undergo a conformational change as both species underwent a transition between pH 6.0 and 7.0.

The effect of the mutations of histidines at positions 197 and 198 on the biological activity of Hsp47 was investigated by the collagen fibrillisation assay. Fibril formation was monitored in the absence and presence of the mutant.

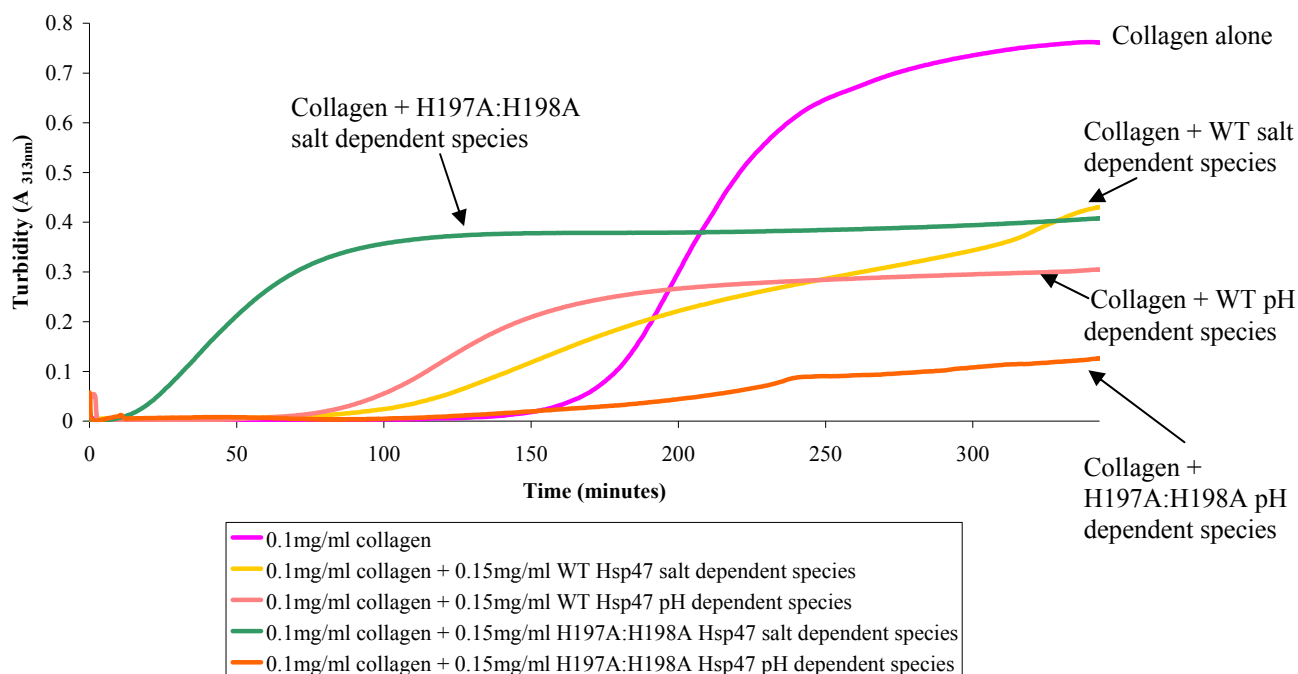


Figure 5.15. Effect of H197A:H198A mutant on collagen fibrillisation. The absorbance of collagen in the presence of each of the mutant species was greatly reduced in comparison to collagen alone.

The salt dependent form of the H197A:H198A mutant was able to reduce collagen fibrillisation as the final turbidity of collagen with Hsp47 was 0.4 in contrast to the 0.76 of collagen alone. The pH dependent form of H197A:H198A mutant mHsp47 significantly reduced collagen fibril formation, giving a final turbidity of 0.13. Therefore, both the salt dependent and pH dependent species were able to reduce collagen fibril formation, albeit to different extents. Comparison of the effect of WT and H197A:H198A mutant Hsp47 salt dependent species showed that the mutant form was less effective at the anti-fibrillisation of collagen chains than its WT counterpart. On the other hand, the mutant form of the pH dependent species was a more potent collagen anti-fibrillisation agent than the WT equivalent. This suggested that mutation of the histidine residues enhances the biological activity of the pH dependent protein but reduced the activity of the salt dependent species.

Furthermore, each of the mutant species seemed to affect the lag phase of the fibrillation process. The lag phase for the fibrillation of collagen in the presence of the WT salt dependent species was reduced in comparison to that for collagen alone. Moreover, the mutant salt dependent species further decreased the lag phase to 10 minutes. The effect of the pH dependent species on lag phase duration, however, was less obvious. The WT pH dependent species reduced the lag phase in comparison to collagen alone, but less so than the H197A:H198A mutant mHsp47 pH dependent species which seemed to prolong the lag time. However, this species also greatly reduced the acceleration phase thus it was difficult to distinguish whether the reduced turbidity was due to lag phase elongation or acceleration phase suppression.

Electron microscopy was used to visualise the effect of the H197A:H198A variant of Hsp47 on collagen fibre production.

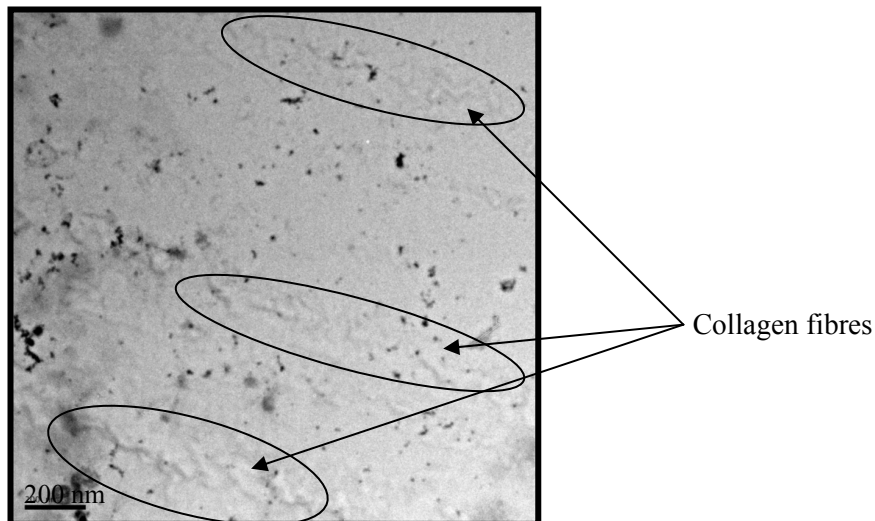


Figure 5.16. Electron micrograph of collagen post-fibrillation, in the presence of **H197A:H198A Hsp47**. The grid illustrates that some fibre formation was occurring; however the short length of the fibres imply that fibril formation was reduced.

The electron micrograph suggested that some fibre formation was occurring, as illustrated by the thin, rod-like structures. Comparison of Figures 4.10(b) and 5.16 revealed that the aggregate-like structures formed by collagen fibrillisation alone were not evident post-fibrillisation in the presence of the H197A:H198A mutant, implying that this mutant prevented fibril formation/aggregation. The difference between Figures 4.10(a) and 5.16 demonstrate that some fibrillisation of collagen was taking place as the fibres visualised post-fibrillisation in the presence of H197A:H198A (Figure 5.16) were larger than those observed pre-fibrillisation (Figure 4.10(a)). However, the size of these fibres was reduced in comparison to those formed in the presence of WT Hsp47, indicating that the H197A:H198A mutant was a more potent anti-fibrillisation protein than its WT counterpart. Furthermore, the scarcity of the fibrils in the presence of the variant strengthened the hypothesis that mutation of histidines 197 and 198 to alanine enhanced the anti-fibrillisation potential of Hsp47. In addition to this the fibres formed in the presence of the variant were more ordered, taking the appearance of long rods, than those formed by collagen alone. Moreover, the observation of reduced fibre production correlated with the results of the collagen fibrillisation assay (Figure 5.15), whereby the turbidity of the solution after completion of the fibrillisation experiment was lower in the presence of the H197A:H198A variant than WT Hsp47.

5.3 Discussion

The aim of these mutations was to probe the acid induced conformational change and the effect on collagen binding biological activity of Hsp47 by substitution of histidine residues near the apex of the A-sheet. This hinge region, near the A-sheet is known to be important in the inhibitory mechanism of serpins and it has been suggested that a similar conformational change may be involved in the structural transition of Hsp47. The pH of the Hsp47

conformational transition is thought to be close to pH 6.0, which is close to the pK_a of histidine, therefore it was decided that each of the histidines would be mutated to alanine.

Sequence alignment and homology modelling identified a histidine cluster within this region. A single mutant (H191A) and a double mutant (H197A:H198A) were constructed and both produced pure, soluble protein that was active. Structural analysis of the protein produced by each of the mutant constructs revealed a salt dependent species that was in the native serpin conformation and a pH dependent species that was in the latent serpin conformation. The discovery of the two different species as well as the latent serpin conformation was novel for Hsp47. However, the latent form of mrmHsp47 was modelled by Davids and coworkers and was shown to comprise of a six stranded A-sheet, comparable to latent, non-inhibitory PAI-1 [139, 182]. Formation of the latent serpin state is accompanied by insertion of the RCL into the A-sheet (the central β -sheet). In inhibitory serpins, the latent conformation is associated with reduced or no activity [139, 220] whereas in non-inhibitory serpins it frequently modifies serpin activity. This conformation provides the serpin with enhanced structural stability as the A-sheet now contains a full complement of β -strands, allowing the formation of numerous backbone hydrogen bonds.

The elution profiles for the single H191A mutant and the double H197A:H198A mutant were similar to that of the WT Hsp47 in that a salt dependent and pH dependent species were produced. Closer examination of the chromatograms revealed that the mutants produced a smaller peak for elution of the pH dependent species than WT mHsp47, suggesting that substitution of histidine at positions 191, 197 and 198 to alanine reduced the production of the more tightly binding pH dependent species. This implies that the interaction of the pH dependent species and collagen is weakened/ disrupted by the removal of the histidine residues. Therefore, it is feasible to propose that these histidine residues may play an

important role in the binding of collagen to its chaperone, Hsp47. Production of the salt dependent species, on the other hand, seemed to be unaffected by the mutations as a similar amount was produced by each of the mutants and the WT. This indicates that the ionic interaction between collagen and Hsp47 is unperturbed by the mutation of H191, H197 and H198.

Structural analysis studies of the salt dependent species of the double and single mutants demonstrated similar results. CD spectra and thermal titration data implied an α -helix and β -sheet containing protein that was hyperstable. The AUC results revealed multiple peaks, suggesting the presence of oligomeric species. The H191A mutant produced two peaks; the first at 1.7 S and a second at 3.2 S, possibly representing monomeric and polymeric Hsp47. The H197A:H198A mutant also showed two peaks; one at approximately 1.3 S followed by a second, smaller peak at 2.6 S. In both cases, the first peak was dominant; therefore the protein being produced was predominantly monomeric but oligomeric protein was also formed. This high molecular weight, non-denaturing species must therefore be in a polymer form. Although serpin oligomerisation is common, it is interesting that only two distinct forms (for example, monomer and dimer) are present in this case. Antithrombin is the only other serpin that is known to form discrete oligomers. In circulation, insertion of the RCL of one molecule into the β -sheet C of a second latent antithrombin molecule causes dimer formation [158, 183]. The oligomers produced by the H191A and H197A:H198A mutants are different from those observed in the study of mrmHsp47 by Dafforn and coworkers [87] although the results obtained do have some similarities. As with their trimeric mrmHsp47, the polymer form in this study, resisted thermal denaturation when heated to 90 °C. Furthermore, comparison of the thermal titration results suggest that the polymer was more resistant to thermal unfolding as the trimer underwent a slow linear reduction in secondary structure [87] whereas polymeric mHsp47 did not show as much of a transition (Figures 5.5 and 5.12 for H191A and

H197A:H198A, respectively). The trimer was able to bind to the (PPG)₁₀ collagen mimic peptide, therefore possessing biological activity [87]. Likewise, the dimer was able to reduce collagen fibrillisation, retaining its biological activity. Furthermore, the anti-fibrillisation activity of each of the mutant polymer species was enhanced in comparison to WT mHsp47.

The CD spectra for the pH dependent species of the H191A and H197A:H198A mutants were characteristic of the folded protein state, thus it can be said that mutation of these residues does not prevent the folding of Hsp47. However, although the mutations did not interfere with the protein adopting a folded state, the actual folded conformation of the pH dependent species of H191A and H197A:H198A mHsp47 was different to that of WT mHsp47. This is evident from the results of the experiments performed for structure determination. The stability curve and AUC data (Figures 4.5 and 4.6, respectively) for the WT pH dependent species signify a low stability, monomeric serpin state, namely the native conformation. By contrast, the thermal stability curves for the pH dependent species of the H191A and H197:H198A mutants (Figures 5.5 and 5.12, respectively) revealed curves characteristic of highly stable proteins that retained a folded conformation even when heated to 90 °C. The AUC data (Figures 5.6 and 5.13) illustrated a single peak for the pH dependent species produced by each of the mutants, denoting the presence of a monomeric species. Taken together, these results propose a highly stable, monomeric serpin state, namely the latent conformation. Therefore, substitution of histidines 191, 197 and 198 to alanine promotes the formation of the latent serpin conformation of mHsp47. This high stability, 6 stranded A-sheet, monomer has not been seen for Hsp47 prior to this study.

The conformation of Hsp47 is known to be very pH-sensitive and is thought to provide a highly pH-dependent mechanism for the binding and release of Hsp47 and collagen. Previous studies and results for WT mHsp47 in this study have revealed extreme changes in Hsp47

structure below pH 7.0. These conformation changes occur in a two stage process, involving two structural transitions. The first transition takes place at a pK_a of 6.5 and results in reduced α -helix and increased β -sheet character. This is thought to involve the ionisation of one or more histidine residues in the protein. The second transition occurs at lower pH, with a midpoint at pH 4.0 and is a result of altered ionisation of glutamate and aspartate residues. Discovery of an active, latent form of Hsp47 allowed the possibility of further investigating this highly pH dependent mechanism. As the latent state was active, it was thought that this conformation could either disrupt the acid induced conformational change or leave it unperturbed. If the conformational change was disturbed, it could be proposed that the substituted residues were involved in the structural transition or that the strand insertion did not alter the ability of the protein to undergo a pH induced conformational change. On the other hand, unaffected conformational changes would imply that structures involved in latent formation were distinct from those involved in the acid induced conformational transition. Data analysis for the conformational change and affinity of the native WT and latent mutant species suggested that the structural transition has not been altered by the formation of the latent serpin conformation. Previous other studies have shown that formation of the latent state, resulting from strand insertion into the A-sheet, greatly enhances the stability of the serpin as a consequence of a more stable A-sheet. This increased stability is signified by a dramatic enhancement in thermal stability of the serpin. The latent forms of both mutants revealed a stable A-sheet that resisted thermal denaturation. The fact that structural change occurred despite the insertion of the RCL into the A-sheet, implies that neither the A-sheet or the RCL are not involved in the acid induced conformational change. If either of these did play a role in the structural transition, perturbation of structural changes would have been seen.

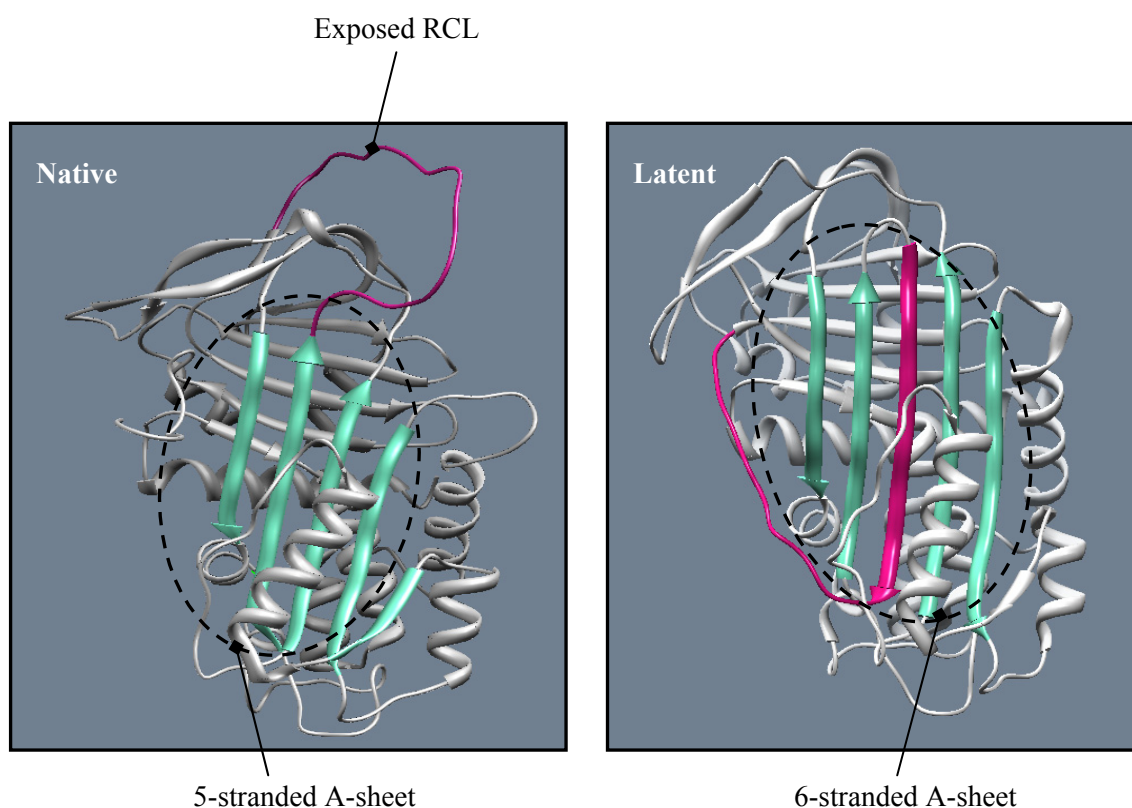


Figure 5.17. Model structures of native and latent *Mus musculus* Hsp47.

The transition from the native, 5-stranded A-sheet to the latent, 6-stranded A-sheet causes numerous changes in the surface of Hsp47 that may subsequently affect the binding of the chaperone to collagen. The biological activity was investigated by measuring collagen fibrillisation in the absence and presence of Hsp47 (Figures 4.8, 5.8 and 5.15 show results for WT, H191A mutant and H197:H198A mutant, respectively). These data show that the latent Hsp47 conformation was able to bind collagen and reduce collagen fibrillisation. This suggests that the collagen binding site is not close to the regions that are altered by the native to latent transition. There are three regions in particular that are affected by this conversion. The RCL is removed following its insertion into the A-sheet, resulting in exposed residues at the top of β -sheets B and C. Insertion of the RCL into the A-sheet leads to expansion of the central β -sheet and as a consequence, movement of surface residues in strands 1, 2, 3, 5 and 6. Finally, the C-terminal end of the RCL is stretched from the C-sheet down the entire edge of

s6A. Our findings therefore imply that these regions are not vitally important for the binding of Hsp47 to collagen. The data obtained from this study is consistent with the results obtained by Davids and coworkers [182] and Dafforn and coworkers [87], leaving their hypothesis of the collagen binding site being in a cleft between helices hA and hG/hH feasible.

As the acid induced conformational change occurs at a pH that is close to the pK_a of histidine, it is likely that the structural rearrangements are instigated by the ionisation of histidines. Sequence studies have highlighted histidine residues that are conserved across Hsp47 expressing organisms but not amongst the serpin superfamily as a whole. It is therefore proposed that one or more of these histidines is involved in the conformational transition of Hsp47 in acidic conditions. The location of histidines 191 197 and 198, around the hinge region, makes them probable candidates for such a role. The H191A mutant of mHsp47 produced a conformational transition that was similar to that of WT mHsp47, with a midpoint of pH 6.6 (± 0.04). However, the pH dependency of the latent conformation was perturbed by the mutation of histidines 197 and 198, implying the involvement of one or both of these residues in the pH dependent conformational change of Hsp47. Although a conformational rearrangement did occur upon reduction of the pH, the midpoint was shifted to pH 6.2 (± 0.06). This is a further 0.3 pH units from that of WT and the H191A mutant of mHsp47, implying that removal of these histidines makes mHsp47 less sensitive to reducing pH. This requirement of more acidic conditions and alteration in the pK_a indicates the involvement of this cluster in the pH dependent conformational mechanism.

CHAPTER 6: Probing the Importance of the Breach Region in the Conformational Transition of Hsp47 by mutating H220

6.1 Introduction

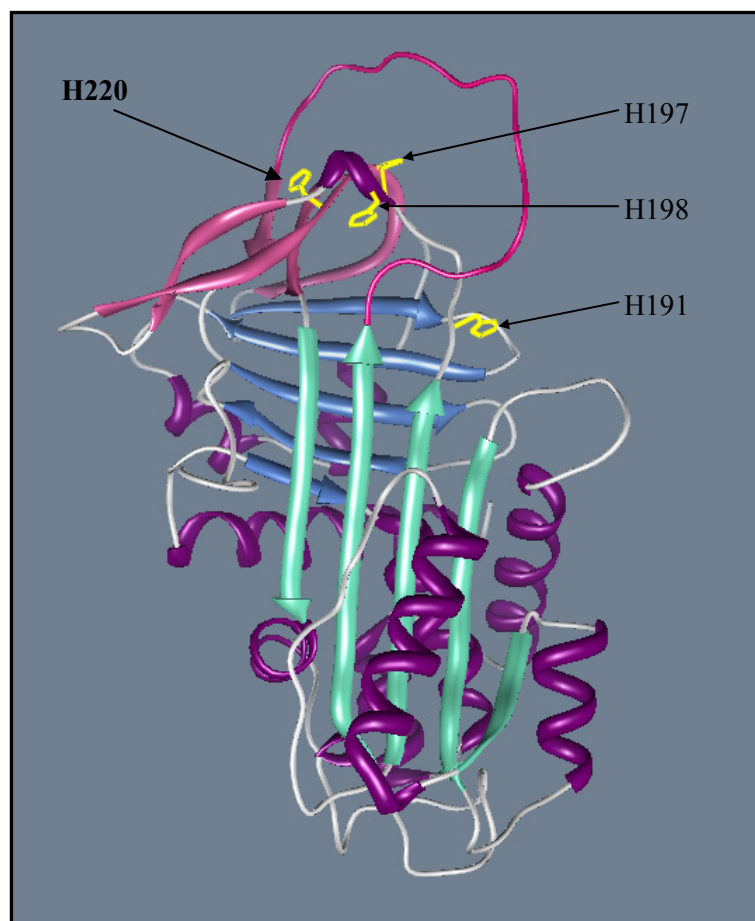


Figure 6.1. Location of histidine 220 on model Hsp47 structure.

The effect of the H197A:H198A variant on Hsp47-collagen binding led to the opinion that other histidines within this region may also play a role in the interaction between mHsp47 and collagen. Histidine 220 is another of the conserved histidines and is located in the breach region (above the A-sheet), close to histidines 197 and 198. Consequently, a H220A mHsp47 mutant was designed to investigate the importance (if any) of this histidine in the binding/release of Hsp47 from its substrate.

As mentioned in Chapter 5, the region above the A-sheet has been shown to be important in the conformational transition from the native to the latent state. Furthermore, mutation of H197 and H198 in this region resulted in the formation of the latent state of mHsp47, suggesting that the mutation of H220 may also effect the conformation or biological activity of the protein.

As with the other mutants, the H220A construct encoded a His tag that would allow non-gelatin binding proteins to be purified via metal affinity chromatography.

6.2 Results

Protein purification

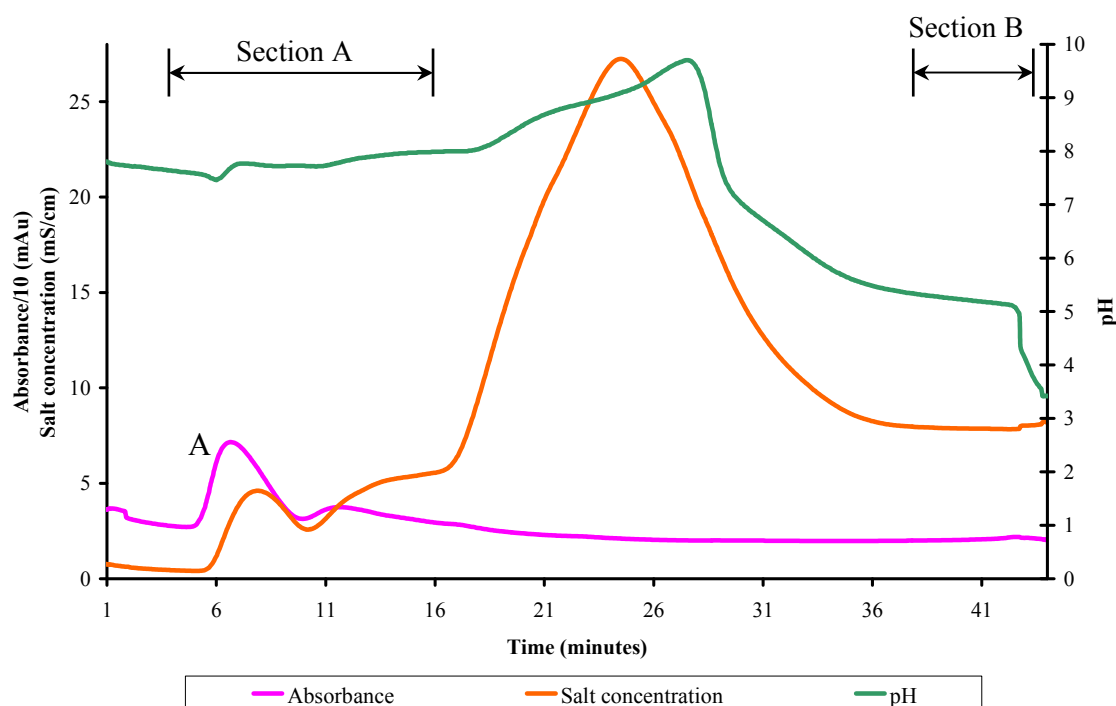


Figure 6.1. Elution profile of H220A *Mus musculus* Hsp47. The absorbance trace shows a dominant single peak between 5-10 minutes (Peak A). This peak coincides with an increase in the salt concentration (orange curve) but the pH remains relatively constant. A reduction in the pH below 6.0 did not result in further release of Hsp47.

The chromatogram produced during gelatin agarose purification showed an absorbance peak between 5-10 minutes which correlated with a peak in the ionic strength (for example, salt concentration) in the early washing steps of the column where the pH remained above 6.0. This peak represented a salt dependent binding species of mHsp47. A further increase in the salt concentration did not release further protein, implying that all the salt dependent form of Hsp47 was released in the first peak. This is consistent with the results for WT, H191A and H197A:H198A mHsp47. However, there is a key difference with this mutant in comparison to each of those studied thus far, that is an absence of a peak in the latter part of the purification following reduction of the pH from alkali to acid which signifies a lack of the pH dependent species. This implied that mutation of H220 abolished any pH dependent binding of Hsp47 to collagen, suggesting that this residue may be important in the interaction.

Each of the fractions collected during the purification was analysed by SDS-PAGE.

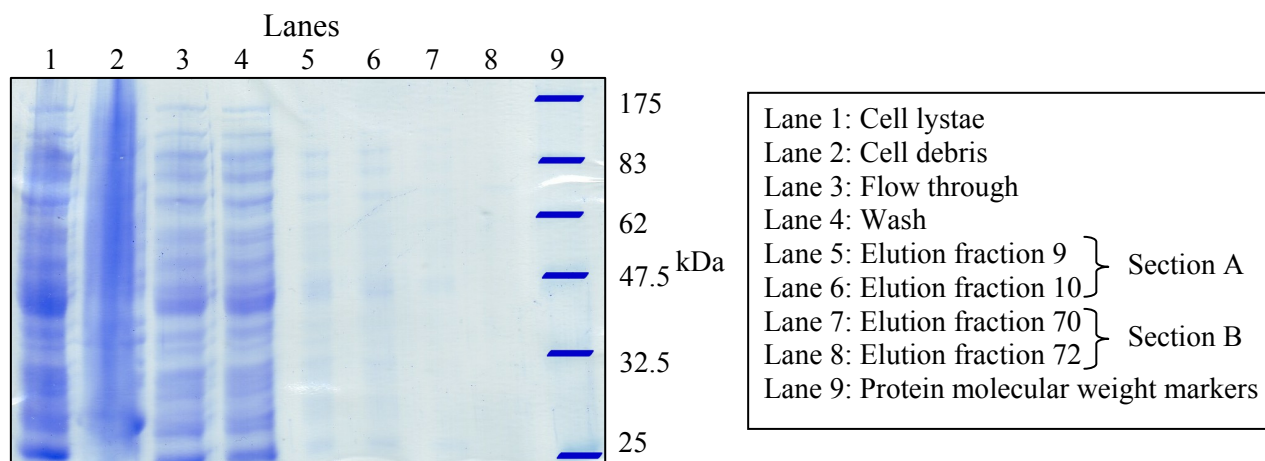


Figure 6.3. SDS-PAGE for H220A mutant. Lanes 3 and 4 show that much of the protein produced was present in the flow-through and wash fractions, respectively. Lanes 5 and 6 represent the salt dependent species. Lanes 7 and 8 correspond to protein that was eluted upon reduction of the pH and confirm the absence of any pH dependent species.

Analysis of the SDS-PAGE results show that most, if not all, of the Hsp47 produced was soluble as it was present in the cell lysate fraction (Figure 6.3, lane 1) and not in the cell debris fraction (Figure 6.3, lane 2). As most of the protein was released in the flow-through and wash fractions; it implied that it was not binding to the gelatin-agarose, thereby suggesting that mutation of H220 interfered with the binding of Hsp47 to collagen. Lanes 5 and 6 (Figure 6.3) signify the salt dependent species and the low intensity confirmed that only a minute amount of this weakly binding species was produced. Absence of a peak for the pH dependent species in the elution profile was confirmed by the absence of a protein band at 47 kDa in lanes 7 and 8; consistent with the proposal that the H220A mutation abolished the pH dependent species of mHsp47. The low intensity of the bands for the salt dependent species inferred that the mutation also affected the binding of the salt dependent species.

The concentration of the salt dependent species was too low for analysis of secondary structure by CD and AUC. The flow-through and wash fractions were therefore subjected to purification on Nickel-Nitrilotriacetic acid (Ni-NTA) resin.

Ni-NTA affinity purification procedure

Purification of 6xHis tagged mHsp47 was performed under native conditions using a batch procedure. The batch process involved binding of the protein to Ni-NTA resin in solution prior to packing the protein-resin complex in a column for washing and elution steps. This method is thought to promote efficient protein binding, particularly when the protein concentration is low.

Cells were lysed by washing with Tris-HCl, pH 8.0 buffer containing 300 mM NaCl and 20 mM imidazole. Low concentrations of imidazole reduce the potential for binding background contaminants. The imidazole group is part of the structure of histidine and can therefore bind

to the nickel ions immobilised by the NTA groups on the matrix. Binding of imidazole to the nickel ions interferes with the binding of dispersed histidines of non-tagged background proteins. Low concentrations of imidazole can therefore reduce low-affinity, non-specific binding of contaminant proteins whilst still allowing high-affinity binding of 6xHis tagged proteins. The matrix was then washed with 20 mM imidazole, pH 8.0. This was repeated three times to ensure removal of background contaminants. Protein elution involved washing the matrix with 250 mM imidazole. The increased imidazole concentration resulted in dissociation of the 6xHis tagged protein as they could no longer compete for the binding sites on the Ni-NTA resin.

Samples of the purification using Ni-NTA were analysed by SDS-PAGE.

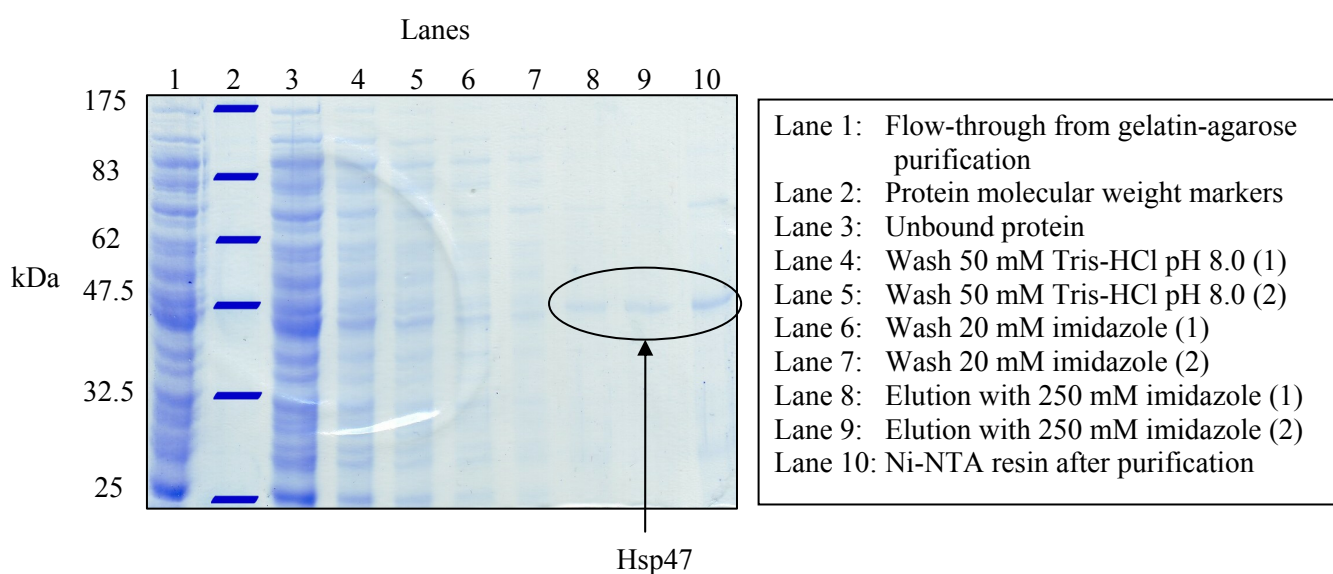


Figure 6.4. SDS-PAGE of purification of H220A *Mus musculus* Hsp47 on Ni-NTA. Lane 3 represents protein that did not bind the Ni-NTA. Lanes 4 and 5 denote non-specifically bound protein. Lanes 6 and 7 show weakly binding protein that was eluted with 20 mM imidazole. Lanes 8 and 9 show tightly binding protein that required an increased imidazole concentration to be released from the Ni-NTA. Protein in Lane 10 is Hsp47 that remained bound to Ni-NTA after elution with 250 mM imidazole.

Purification on Ni-NTA revealed that only a small amount of the flow-through from the gelatin-agarose column interacted with the Ni-NTA. Reapplication of the unbound protein onto Ni-NTA resulted in further binding and elution as shown in Figure 6.4. Of the protein that bound to the Ni-NTA beads, some bound non-specifically and eluted with 50 mM Tris-HCl (pH 8.0) (Figure 6.4, lanes 4 and 5) whilst some was weakly binding and eluted with 20 mM imidazole (Figure 6.4, lanes 6 and 7). 250 mM imidazole was used to elute any strongly binding Hsp47 (Figure 6.4, lanes 8 and 9) whilst some protein remained bound to the Ni-NTA, even after the stringent wash and elution conditions (Figure 6.4, lane 10).

Purification on Ni-NTA produced pure Hsp47 (single band on SDS-PAGE) that was used for secondary structure analysis.

Structural characterisation of H220A mHsp47

As with WT and previous mutants, H220A Hsp47 was subjected to circular dichroism studies to determine secondary structure composition and protein stability whilst AUC was used to determine its oligomerisation state (results shown below). Subsequently, pH titration experiments were performed to investigate the ability of the mutant to undergo the pH-dependent conformational transition and the collagen fibrillisation assay was carried out to determine the biological activity of the chaperone.

All CD experiments were performed using a 0.2 cm quartz cuvette and a protein concentration of 0.2 mg/ml.

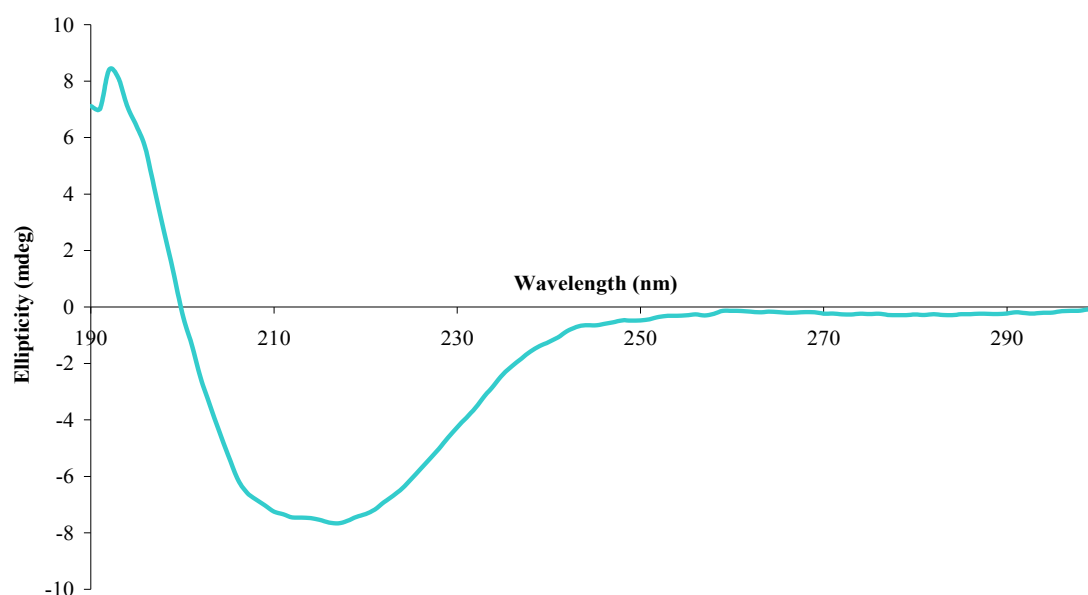


Figure 6.5. CD spectrum of H220A mHsp47. Minima at 208, 216 and 222 nm are typical of an α -helix and β -sheet containing protein, whilst a positive signal at 195 nm signifies protein that is in a folded state.

The thermal stability of the protein was investigated using a thermal titration experiment that involved an increase in temperature from 20 to 90 °C by 1 °C increments.

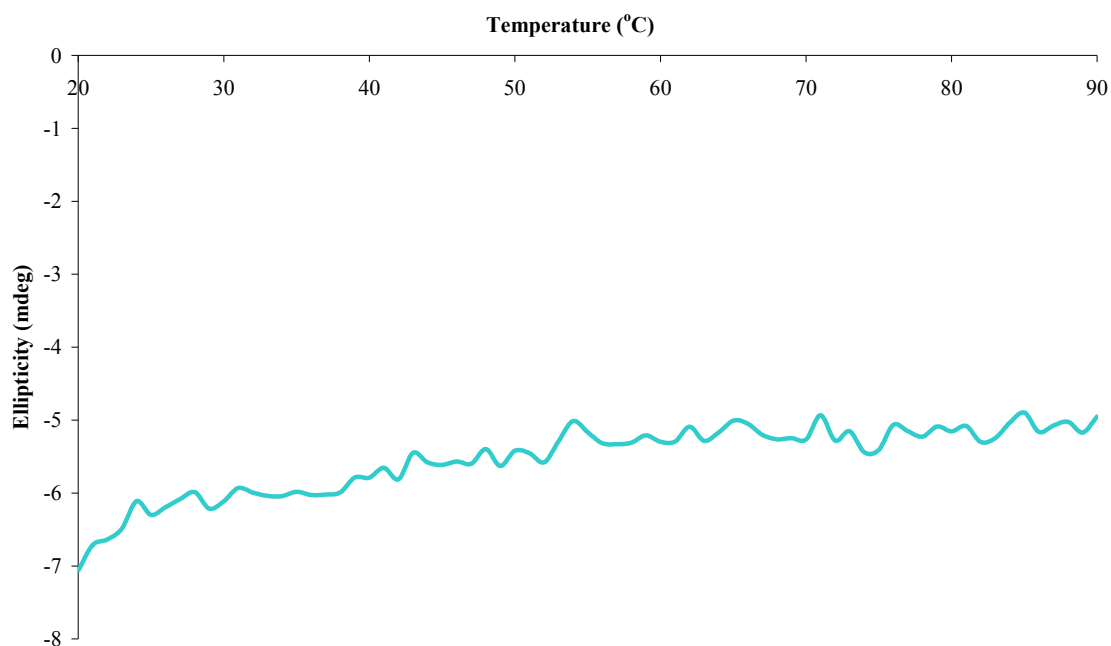


Figure 6.6. Thermal stability curve for H220A mHsp47. Heating the protein from 20 to 90 °C shows a slight change in the CD signal. This implies that there may have been some gradual loss of secondary structure; however the protein was not denatured as there was no sharp transition and no distinct change in CD signal.

The CD results revealed an α/β -protein that was folded and resisted thermal denaturation upon heating to 90 °C, therefore being highly thermostable.

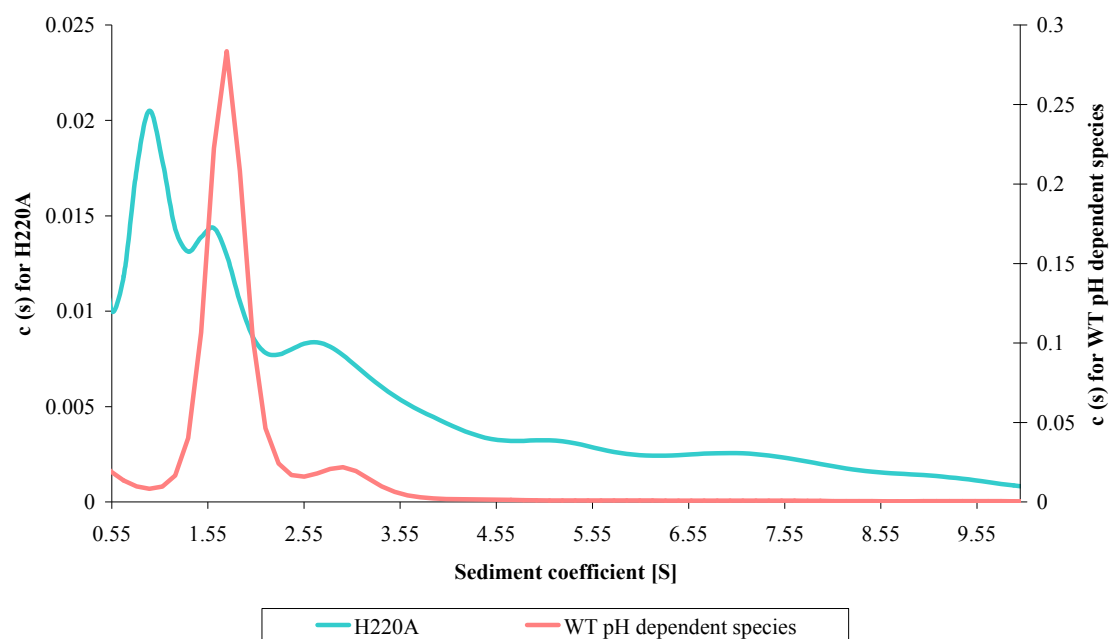


Figure 6.7. AUC for H220A mHsp47. Multiple peaks signify the presence of oligomers.

Numerous peaks from the AUC results illustrate that the protein was a mixture of monomeric and polymeric.

The dependency of the protein fold on pH was monitored by reducing the pH from alkali to acid using phosphoric acid and monitoring the change in CD at 222 nm.

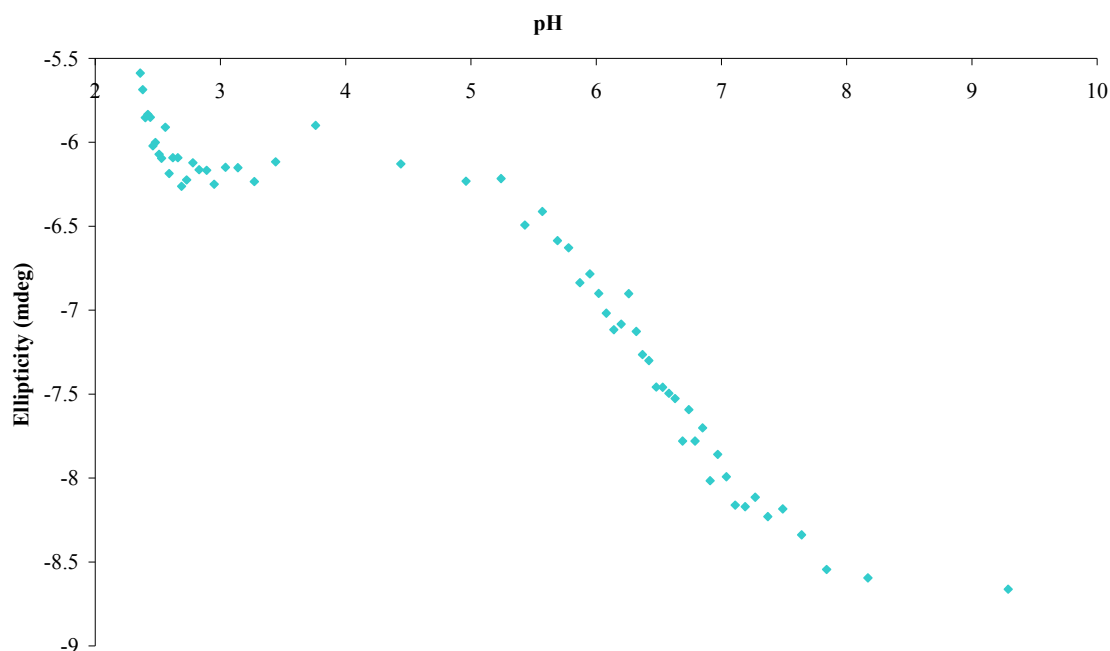


Figure 6.8. pH titration of H220A *Mus musculus* Hsp47. The pH transition was characteristic of Hsp47; whereby the acid induced conformational change had a midpoint inflection at pH 6.6.

The ability of the mutant to undergo a pH induced conformational change suggests that mutation of H220 did not affect the pH sensitivity of the protein structure. As the pH dependent conformational transition is thought to be mediated by the ionisation state of histidine residues, it is expected that the mutation of histidine residues which are involved in this conformational change would be accompanied by a shift in the pH dependency of the protein. Thus, it can be said that H220 is not important for controlling the conformational transition of mHsp47.

Biological activity determination

As with the WT and other mHsp47 variants, the effect of the H220A mutant on collagen fibril formation was investigated by monitoring the turbidity of collagen in the absence and presence of the H220A Hsp47 variant (Figure 6.9).

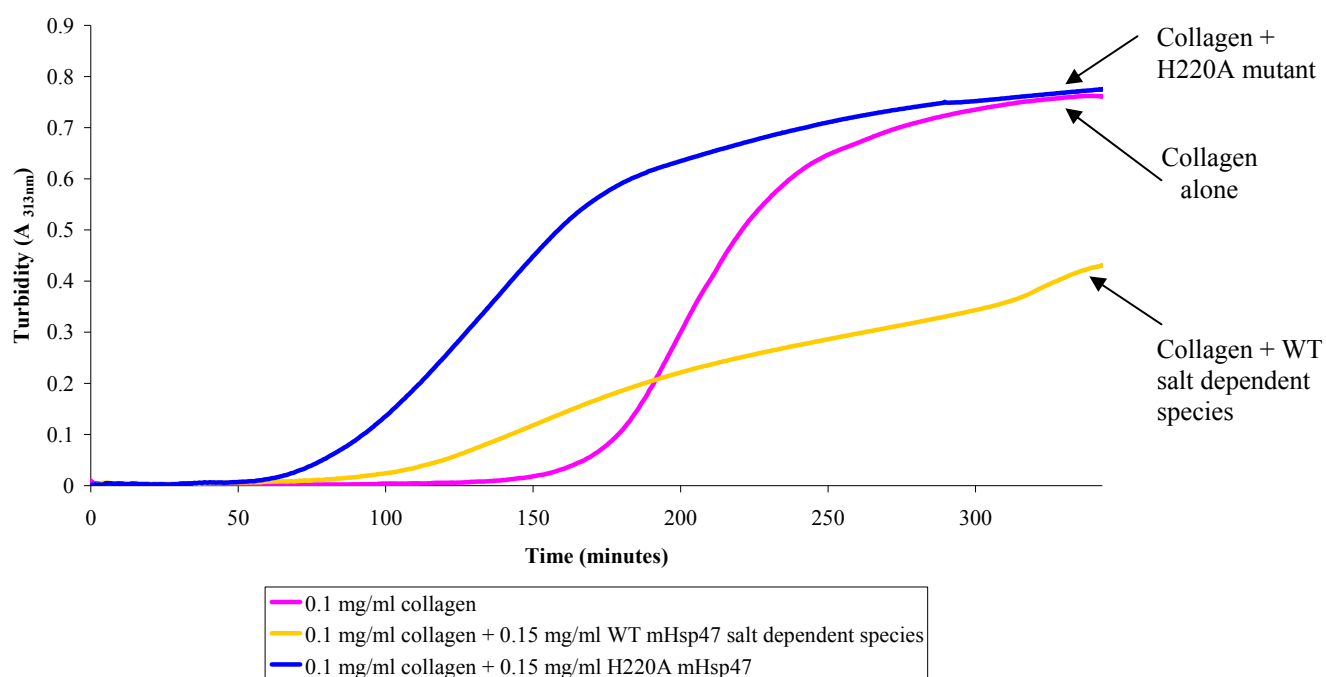


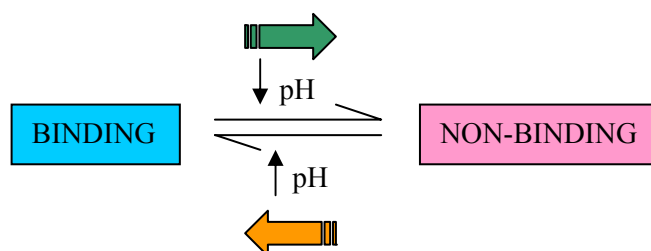
Figure 6.9. Effect of H220A mutant on collagen fibrillisation. The blue curve illustrates that the mutant was unable to reduce the fibrillisation of collagen as the turbidity at the end of the experiment was very similar to that of collagen alone.

Upon comparison of collagen fibril formation results for WT and variant it was evident that the H220A mutant of mHsp47 did not possess anti-fibrillisation activity despite mutation of H220. The final turbidity of the solution comprising of collagen and mutant protein was 0.78 compared with 0.73 of collagen alone. Interestingly, this non-gelatin binding (purified by Ni-NTA chromatography) protein was the only mutant which lost its biological activity upon substitution of histidine 220 for alanine. In addition to this, the collagen fibrillisation curve in the presence of H220A mHsp47 showed a much reduced nucleation phase compared with that of collagen alone and collagen with WT mHsp47.

6.3 Discussion

The elution profile of H220A Hsp47 from gelatin-agarose demonstrated a single peak that represented the salt dependent species of mHsp47 as it coincided with a peak for the salt concentration. This is consistent with the result for WT and H197A:H198A, both of which produced a salt dependent species that correlated with an increase in the salt concentration at an alkali pH. However, when the corresponding fractions were analysed by SDS-PAGE, the protein was hardly visible suggesting that there was very little salt dependent collagen binding species. The absence of a pH independent species for this mutant implied that mutating histidine 220 somehow increased the likelihood of the protein to elute from collagen with salt, thereby suggesting that H220 is important in the pH dependent binding/release of Hsp47 to collagen. The low concentration of the H220A variant salt dependent species in comparison to WT, H191A and H197A:H198A mutants suggested that the interaction of this species with gelatin was also affected by the H220A substitution. The finding that most of the Hsp47 produced by the H220A mutant did not bind gelatin and was present in the flow through fraction of the purification, indicated that the binding of the chaperone to its substrate was perturbed. These results allow for two feasible proposals:

- (i) The binding site of Hsp47 for collagen is likely to be at or near the region of H220. This could be investigated by the ability of the mutant to bind its substrate.
- (ii) H220 is in close proximity to the important site for the conformational change trigger. This could be investigated by a pH gradient as the conformational transition is as follows:



Each of these proposals is discussed below.

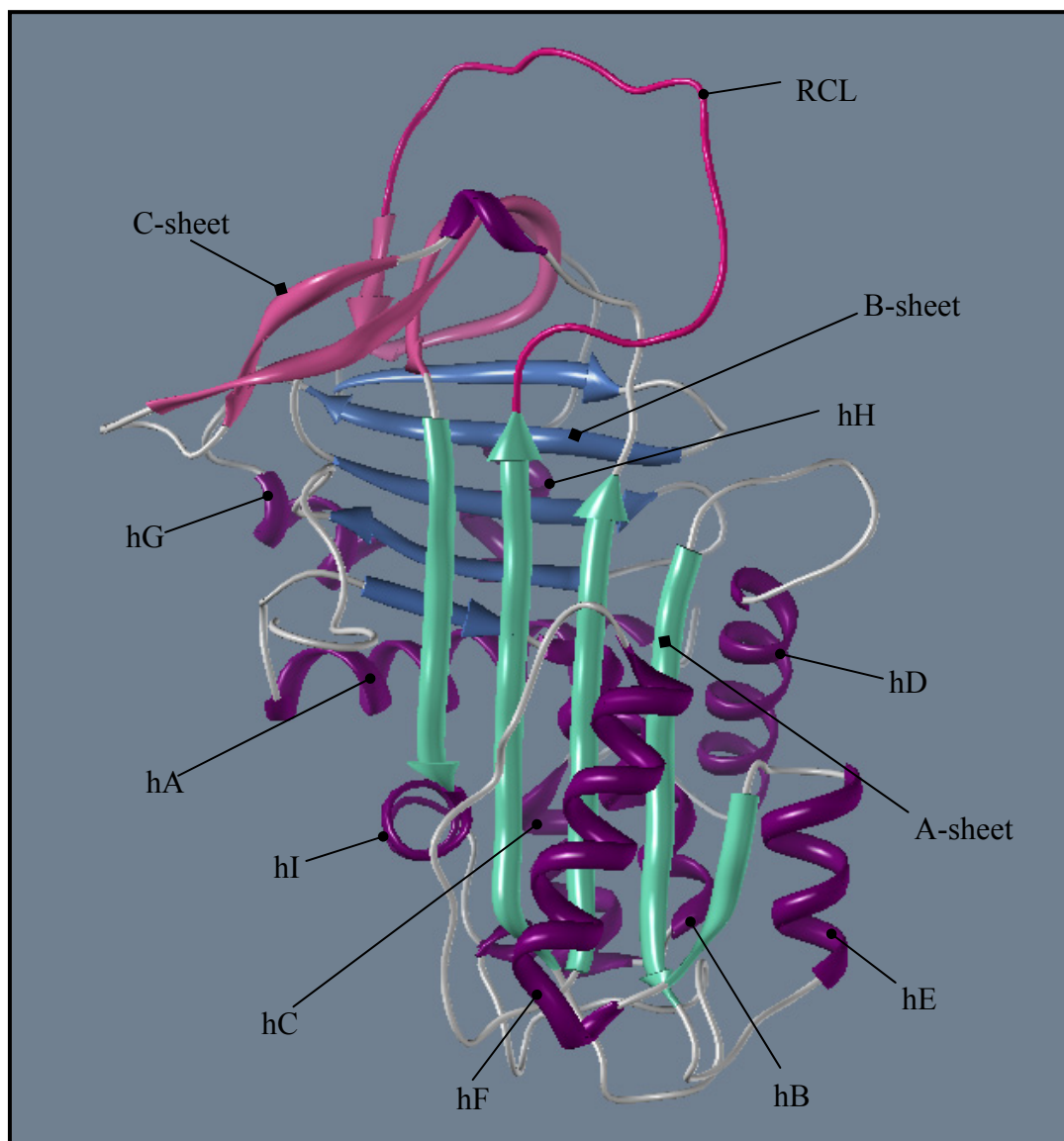


Figure 6.10. Model structure of native *Mus musculus* Hsp47.

Collagen binding studies by El-Thaher and coworkers in which part of helix A of Hsp47 was deleted resulted in diminished collagen binding. They therefore concluded that this region comprised of all or part of the collagen binding site. Since this study, Davids and coworkers have modelled the three-dimensional structure of inhibitory and latent mrmHsp47 using cleaved hPCI as template [182]. The specificity and high affinity binding of Hsp47 to procollagen/collagen peptides at physiological pH [180, 221, 222] can be explained by the long, deep cleft of mrmHsp47. Helices hA and hG/hH form the sides of the cleft whilst the base is formed by β -sheet B. The B β -sheet projects hydrophobic side chains toward the cleft

whilst the helices project hydrophilic amino acid side chains at the sides of the cleft. Deletion mutants of mrmHsp47 were designed to test the hypothesis of the proposed binding cleft. Two deletion mutants, NΔ1 and CΔ3, were unable to bind collagen (Nagata, unpublished). The NΔ1 mutant lacked the first 32 amino acids of the N-terminal whilst CΔ3 was missing the last 34 amino acids of the C-terminal. Both deleted regions comprised of extensive regions of the putative procollagen/collagen binding groove [182]. These findings together with the findings of C-terminal His tagged *Xenopus laevis* and *Danio rerio* Hsp47 in this study (Chapter 3) denote the importance of an unperturbed Hsp47 C-terminus for the function of the protein. As helix A has been discovered as being central to the mechanism of conformational change for serpins, it was suggested, it is likely that this region is also structurally important Hsp47.

The H220A mutation had a similar effect to that of the L266A mutation by Miller and coworkers whereby the substitution resulted in a complete loss in the collagen binding ability of Hsp47. As the L266A mutation is relatively conservative it was thought unlikely that it would have an adverse affect on the conformational dynamics of the protein (Olerenshaw and coworkers unpublished). The proposed hydrophobic cleft was further investigated by a double mutant; L32A:F34S that was located in helix A. The placement of the side chain of L32 implied that it would affect the hydrophobicity of the cleft, thereby affecting collagen binding. And indeed, the double mutant (L32A:F34S) was unable to bind and release collagen.

SDS-PAGE confirmed the result of the elution chromatogram by illustrating the presence of only a salt dependent and not a pH dependent species. The intensity of the salt dependent species was also reduced in comparison to that produced by WT mHsp47. Furthermore, it was evident that most of the Hsp47 produced by the H220A mutant either did not bind (pH dependent species) or bound very weakly (salt dependent species) to gelatin and was present

in the flow-through and wash fractions, respectively. This implied that mutation of the H220 resulted in a change in Hsp47 that disrupted the ability of the chaperone to bind its substrate. Thus, both the proposals above are still possible.

Following purification using Ni-NTA, unbound protein from the flow-through fraction was used for structural analysis in an attempt to investigate the reasoning behind the non-binding property of the H220A mHsp47 mutant. The CD spectrum showed the protein to be folded, demonstrating that the mutation did not affect the folding ability of the protein. It remained unknown however, whether it was true at a side chain level. Thus, it remained indefinite whether the substitution of histidine with alanine resulted in a disruption in Hsp47-collagen binding due to a difference in the folded conformation of the protein or by disabling the protein from forming a bond with the gelatin. The thermal stability experiment demonstrated that the protein was highly stable, resisting denaturation upon heating to 90 °C. AUC results exhibited multiple peaks, characteristic of a polymeric protein and a peak at approximately 1.7 S, corresponding to monomeric material. The number of peaks was more than was observed for the WT salt dependent species (Figure 6.7), thus it is possible that the mutation led to an increase in the number of oligomeric species of the protein. Taken together, these results suggest the presence of a stable, oligomeric form of Hsp47 that was non-collagen binding.

The acid induced conformational change was probed by the pH titration experiment. It was evident from the results that the mutant underwent a conformational change similar to that of the salt dependent species of WT mHsp47, denoting that the H220A substitution did not interfere with the folding of the protein into a pH labile conformation or the ability of the serpin to undergo a conformational transition upon a reduction of pH. As mentioned previously, histidine residues are thought to be central to the pH sensitivity of Hsp47 and so

modification of histidine residues that play a role in instigating the conformational change of the protein is likely to bring about a variation in the pH induced transition. Therefore, as the transition of the H220A mutant is largely similar to that of WT mHsp47, it can be said that histidine 220 is not involved in the mechanism of conformational change of the protein and that polymeric Hsp47 can still undergo the transition.

According to the collagen fibrillisation data, the H220A mutant was unable to reduce fibril formation. This is not surprising as the inability of the protein to bind tightly to gelatin-agarose during protein purification led to the opinion that it would have no or very little biological activity. The agreement and consistency of results provide strength to the data collected. Figure 6.9 illustrated that there was no reduction in collagen fibril formation in the presence of H220A mHsp47 in comparison to collagen alone. This implied that the region of the protein responsible for the sensitivity to pH may be distinct from the protein region involved in collagen binding as the H220A variant lost its ability to bind to gelatin/collagen but retained its capability to undergo conformational change upon a reduction in pH (Figure 6.8).

Hence, when analysed together, the secondary structure data from this study and other Hsp47 mutation studies are not in complete accordance with either of the proposals put forward earlier. The finding that the H220A variant was unable to bind gelatin suggests that histidine at position 220 is important for the binding of mHsp47 to its substrate. It could be that the histidine itself is involved in the binding or that the substitution of histidine for alanine abolished any interaction(s) holding the Hsp47 in the non-binding conformation; thereby preventing an Hsp47-collagen complex. His 220 however, is located on s3C; at the top β -sheet A and evidence from previous work (discussed above) strongly suggests that the binding site of Hsp47 for collagen on helix A. It is therefore unlikely that His 220 is part of

this binding site. H220 is in close proximity to H197 and H198 and as seen in Chapter 5.0, mutation of histidines 197 and 198 perturbed the pH sensitivity of the protein, highlighting their importance in the Hsp47-collagen binding. The inability of the H220A mutant to bind collagen further emphasises the significance of the breach region in the Hsp47-collagen interaction. Thus, it is feasible to propose that there may be a secondary collagen binding site on Hsp47, involving His 220. Requirement of an Hsp47-collagen interaction at both primary and secondary binding sites would explain the inability of the H220A variant to associate with collagen.

CHAPTER 7: Probing the Importance of the Gate Region in the Structural Transition of Hsp47 by mutating H255 and H256

7.1 Introduction

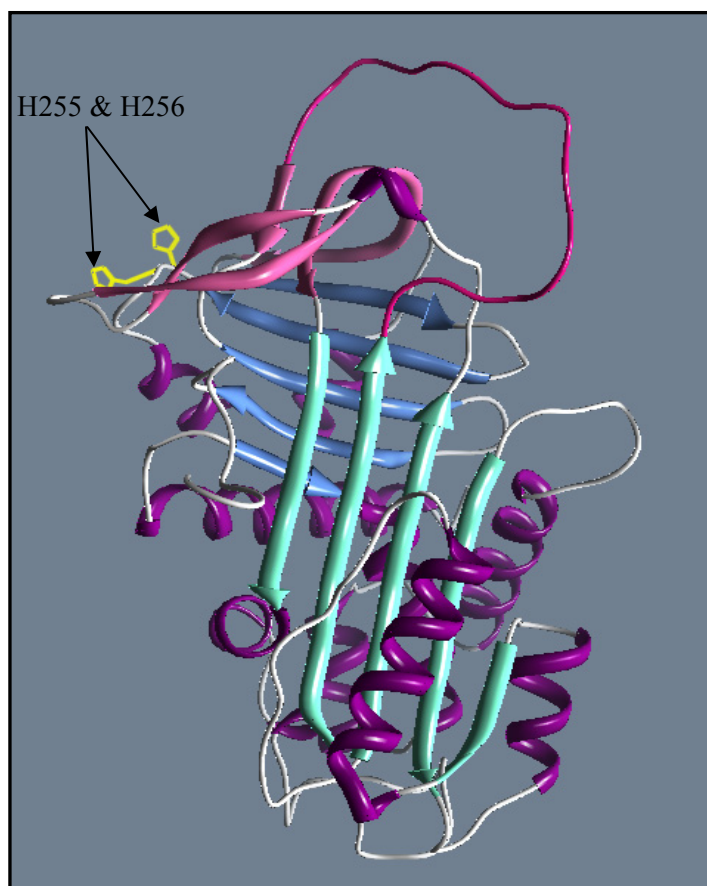


Figure 7.1. Location of histidines 255 and 256 on model Hsp47 structure.

Histidines 255 and 256 make up the gate cluster of Hsp47 which neighbours the gate region that is located near the end of β -sheet C. In the serpin superfamily, H255 is frequently substituted with tyrosine or asparagine and therefore may not play such an important role in maintaining the serpin fold. H256 on the other hand, is conserved within the Hsp47 serpin family, and thus may be involved in the conformational dynamics of the protein.

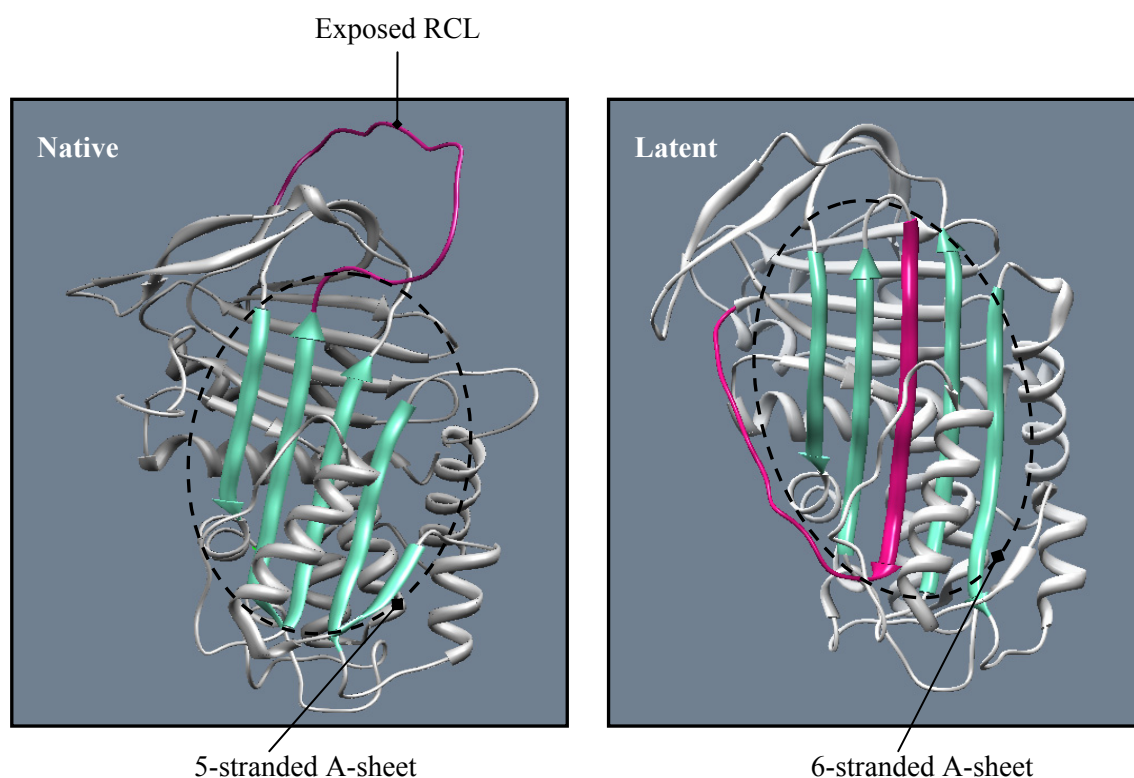


Figure 7.2. Model structures of native and latent Hsp47.

Although not very much is known about the gate cluster, the gate region is thought to undergo movement upon formation of the latent state in many serpins. Partial insertion of the RCL into the A-sheet requires repositioning of s1C away from the remaining β -sheet C, leaving the C-sheet with only three β -strands [157]. This movement of s1C is thought to be hindered by s3C, s4C and the turn connecting these strands [139]. Latent serpins are unable to interact with proteases and are not protease inhibitors. The transition to latency represents a control mechanism for the serpin PAI-1. PAI-1 is released in the inhibitory conformation, however, it spontaneously adopts the latent conformation following removal from vitronectin, which holds the serpin in an active form [223, 224]. The spontaneity of conversion of PAI-1 to the latent form is thought to be due to flexibility of this gate region (composed of s4C-s3C) [123]. This hypothesis was based on the gate region of native PAI-1 having strong hydrophobic interactions with turn connecting s1C to s4B which need to be overcome so as to permit s1C to reside as in the latent conformation [201, 220].

In α_1 -AT, this region is restrained by a disulphide bond; thereby impeding spontaneous conformational adaptation to the latent state. Mutation of one of the cysteines involved in the formation of the double bond abolishes the inhibitory activity of α_1 -AT. Destabilisation of native interactions in α_1 -AT by substitution of residues in s1C resulted in conversion from the native, metastable state to a stable, latent conformation. This signified the importance of the interactions in the B/C barrel region in the conformational transition of serpins. These findings highlight the importance of the gate region in the activity of serpins. Furthermore, mutations of residues that would perturb hydrophobic interactions in this region of α_1 -AT, namely F208A of s4C and F370A of s4B, did not result in conversion to the stable, latent state [220]. These findings imply that there may be a difference in the controlling of the conformational transition amongst serpin proteins. As the native and latent states of various serpins share the same features however, it is highly likely that destabilisation of the B/C β barrel is involved in the conformational transition of other serpins [161].

As mentioned above, the spontaneous transition to the latent state of PAI-1 is thought to be due to the disordered and mobile gate region [139]. PAI-1 variants have been engineered to either prevent or reduce the rate of s4A insertion into the A-sheet, these included mutants that were designed to stabilise the gate region. The mobility of the gate region was covalently restrained by engineering a disulphide bond in PAI-1, producing a CC variant. Furthermore, residues 203-213 of the gate of most serpins are bound in a hydrophobic pocket. In latent PAI-1, however, this position is occupied by the C-terminal, preventing the gate region from occupying the pocket, allowing its mobility [201]. According to modelling studies, leucine at position 280 and arginine at position 30 were likely to be responsible for the displacement of the C terminus of PAI-1. Mutants (L280I and R30T) that enabled the C terminus to take up its normal position and restrain mobility of the gate were engineered to test this hypothesis. Analysis of the gate mutants showed a 2.6-fold increase in the half life relative to WT PAI-1.

This increase in half life signified a decreased rate of transition to the latent state. Moreover, an R30E mutant, designed to disrupt the salt bridge with Glu 350 and destabilise the latent conformation, resulted in a two-fold increase in the $T_{1/2}$ [225]. The significant enhanced stability of the gate and disulphide variants provide evidence of the interactions in the gate region being important for the stability of active serpins [201].

A variant of histidines in position 255 and 256 of the gate cluster was engineered to investigate the importance of these residues in the serpin fold and/or conformational transition. Perturbation of the pH dependency of the mutant would indicate that one or both of these residues play a significant role in the pH dependent conformational change of Hsp47. As explained above, the gate cluster has been shown to be important in the latency transition for PAI-1 and α_1 -AT. It would therefore be interesting to investigate whether mutations in this region of Hsp47 have a similar stabilising effect on the non-inhibitory serpin.

7.2 Results

Protein Purification

Protein was purified using collagen-affinity chromatography as per previous methods by loading 50 ml of the cell lysate onto a column of gelatin-agarose beads and reducing the pH from alkali to acid.

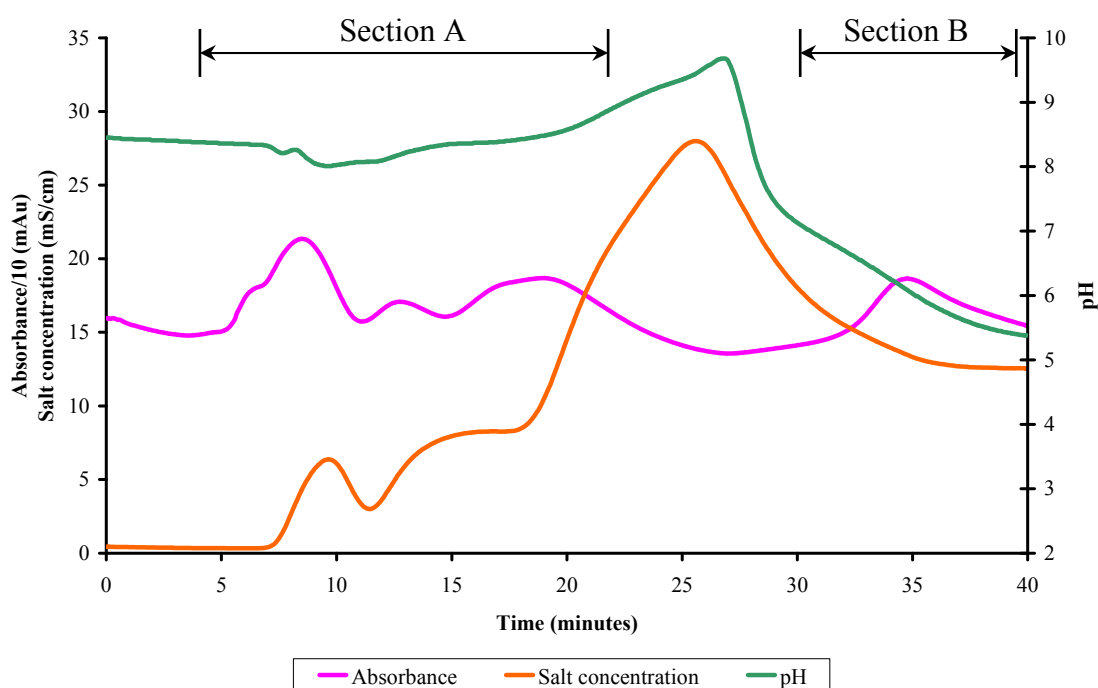


Figure 7.3. Elution profile of H255A:H256A *Mus musculus* Hsp47. The absorbance curve shows multiple peaks; the peaks between 5-20 minutes represent the salt dependent species whereas the peak at 35 minutes corresponds to elution of the pH dependent species. The peaks denoting the salt dependent species coincide with an increase in the salt concentration whilst the pH curve demonstrates that a reduction in the pH to acid correlates with the absorbance peak at 35 minutes.

The elution profile for the H255A:H256A mutant was similar to that of WT in that there were salt dependent and pH dependent species. However, the salt dependent species for the WT was eluted in three peaks rather than a singleton. Each peak however, did coincide with a peak for the salt concentration on the column. The first peak for salt concentration (between 5-10 minutes) resulted in the elution of some salt dependent protein. A further increase in the salt concentration released further Hsp47 that was salt dependent (represented by the peak between 10-15 minutes). A further increase in the salt concentration still, resulted in more salt dependent species being eluted (as illustrated by the peak between 15-20 minutes). This implied a concentration dependent release of the salt dependent species, whereby the protein affinity for gelatin varied. Thus, the more tightly bound protein required an increased salt concentration to abolish its interaction with gelatin in comparison to protein that had a lower

affinity for the substrate. The pH dependent species was eluted in a single peak (at 35 minutes) when the pH was decreased to 6.3. The elution of this species occurred when the salt concentration was constant; implying the release from gelatin was independent of salt concentration.

Samples taken during the purification procedure were analysed by SDS-PAGE.

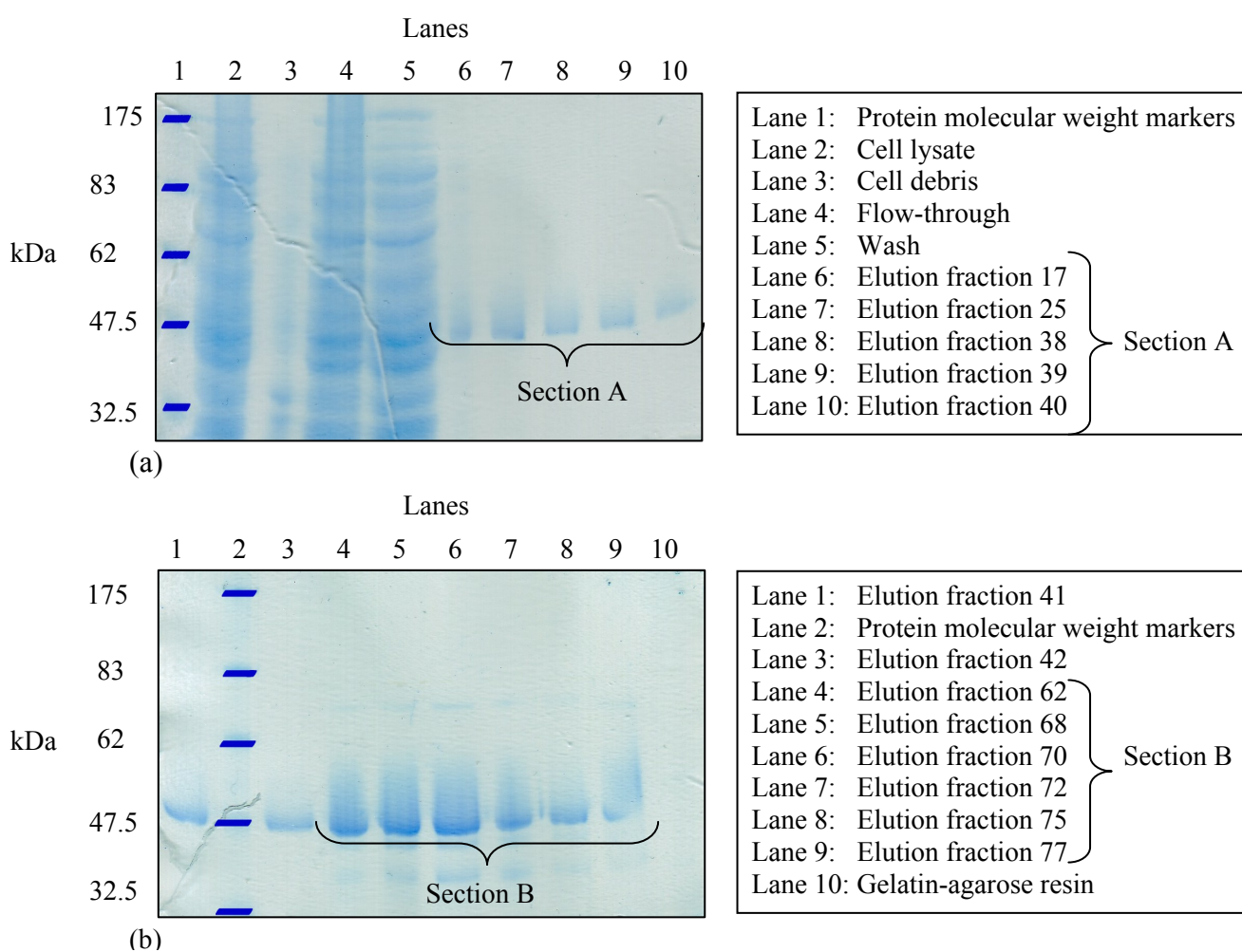


Figure 7.4. SDS-PAGE for H255A:H256A mutant. (a) Lanes 1 and 2 show that the Hsp47 produced was present in the cell lysate and not the cell debris fraction, respectively. This signified that the protein was soluble. Lanes 7 and 8 represent the salt dependent species that eluted in the first peak, whilst lanes 8-10 correspond to the salt dependent species that eluted in the second peak. (b) Lanes 1 and 2 denote the salt dependent species that was released in the third and final peak for this species. Lanes 4-9 illustrate the fractions of the pH dependent species. Lane 10 demonstrates that there was no

protein on the gelatin-agarose resin following the purification and had therefore eluted in the purification.

Structural characterisation

The secondary structure of each of the species was investigated by CD. Protein used was from the first peak in Section A of Figure 7.3. A thermostatically controlled 0.2 cm pathlength quartz cuvette and a final protein concentration of 0.2 mg/ml were used for CD measurements.

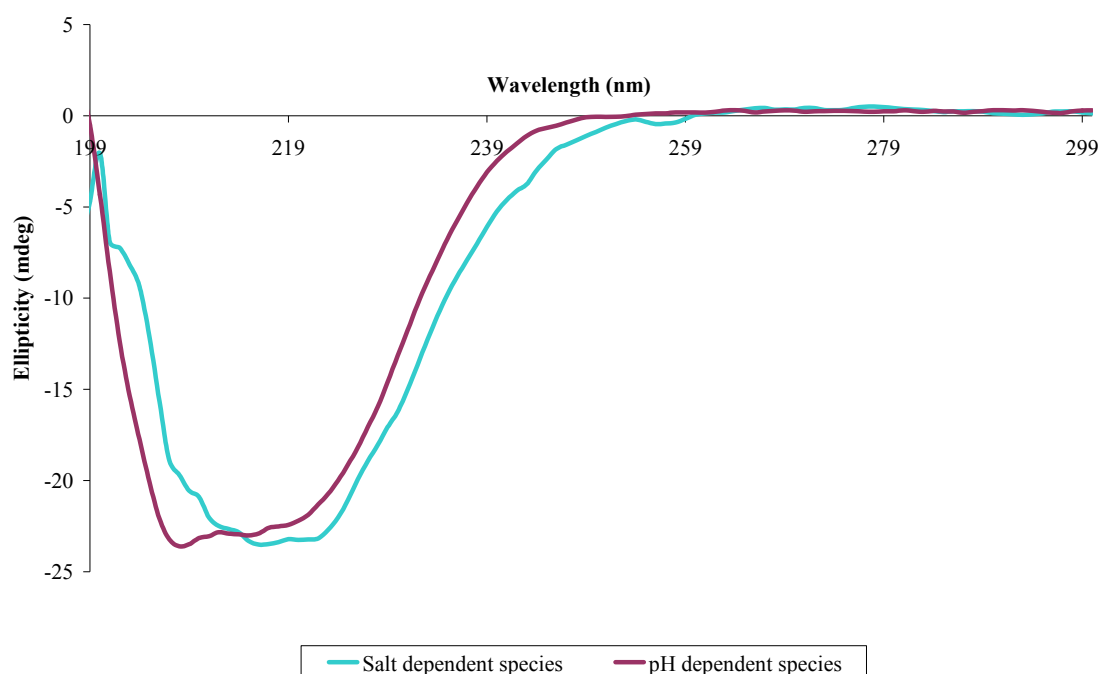


Figure 7.5. CD spectra for H255A:H256A mutant. Both the salt dependent and pH dependent species illustrated spectra characteristic of a folded protein. Minima at 208 and 222 nm signify α -helices and a minimum at 216 nm denotes β -sheet character.

The stability of the protein was determined by thermal denaturation studies. Any change in the ellipticity was monitored whilst the protein was heated from 20 to 90 °C.

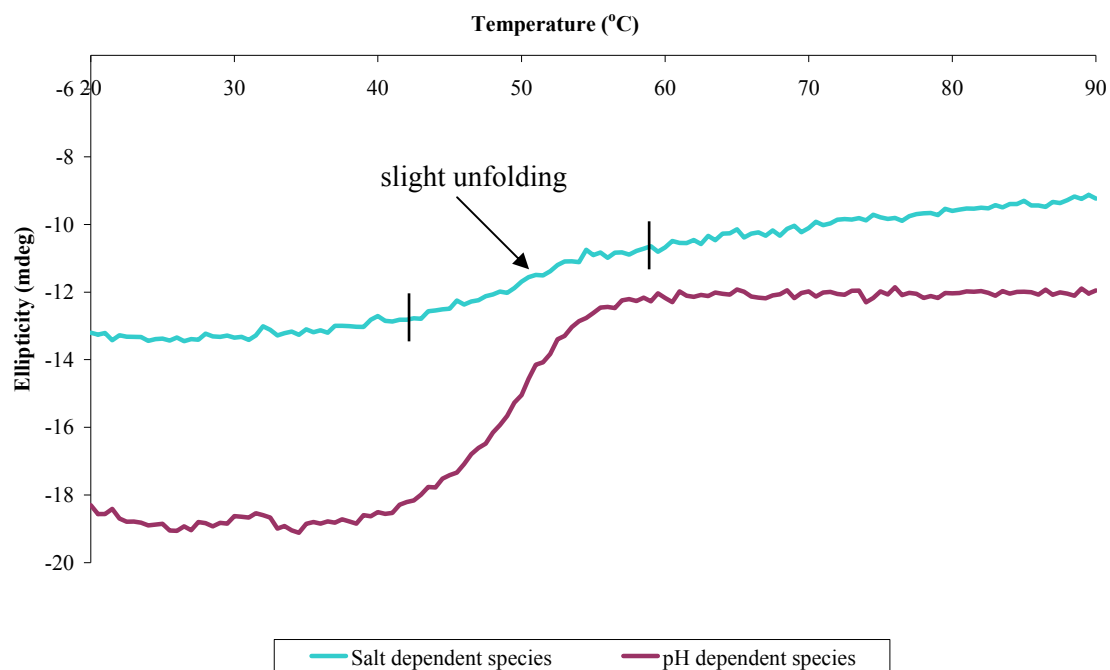


Figure 7.6. Thermal denaturation curves for H255A:H256A mutant. The salt dependent species showed a gradual loss of some secondary structure as illustrated by a slow change in the ellipticity. The pH dependent species demonstrated a clear transition with a midpoint at 45 °C, denoting a significant conformational alteration.

Results from the CD experiments imply that salt dependent species was folded, comprised of α -helix and β -sheet character (typical of serpins) and was highly stable as it resisted denaturation upon thermal titration. From this it was deduced that the salt dependent species could be in either the latent or polymeric serpin conformation. Although the salt dependent species resisted complete denaturation when heated, some unfolding can be seen on the denaturation trace (Figure 7.6). In the case of the polymeric state this could be due to the unfolding of monomers at either end of the polymer and in the case of the latent state, this unfolding could be due to incomplete insertion of the RCL into the A-sheet. The pH dependent species was illustrated as being in a folded, low stability form that denatured when heated producing a transition with a midpoint at 45 °C. Again, α -helix and β -sheet character was evident and the pH dependent species was thought to be in the native serpin conformation.

AUC was used to determine the oligomeric state of the protein and to deduce which of the serpin conformations each of the species was in.

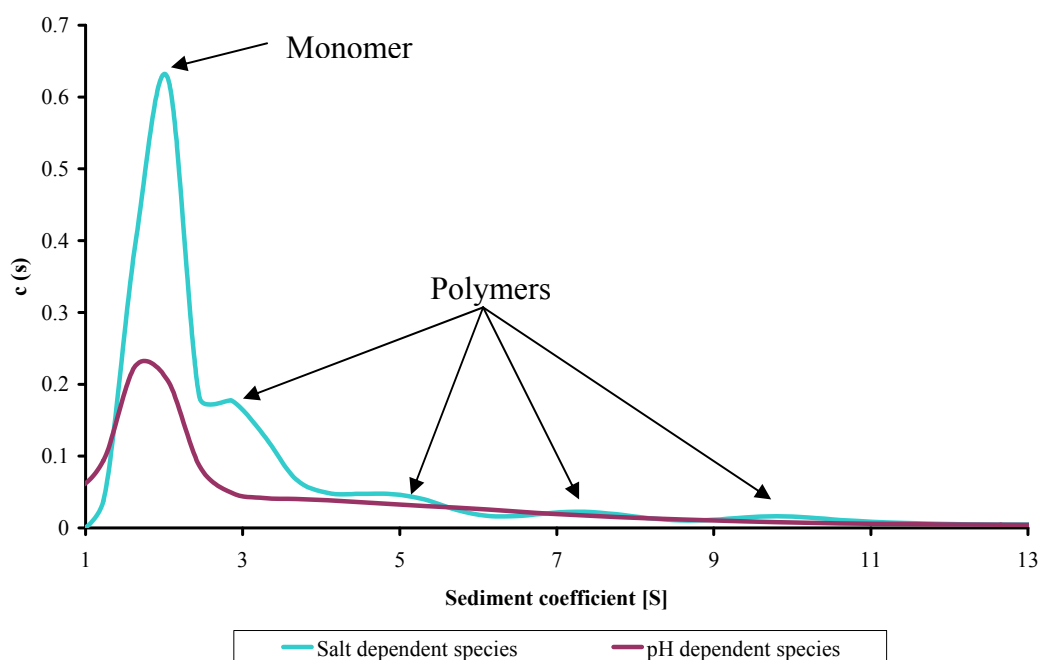


Figure 7.7. AUC for H255A:H256A mutant. The salt dependent species illustrated multiple peaks, suggesting the presence of polymers. The pH dependent species produced a single peak at 1.8 S, signifying a monomer.

Taken together, the structural analysis data suggested that the salt dependent species contained monomer as well as high molecular weight species that resisted thermal unfolding and comprised of α -helix and β -sheet content; satisfying properties of the polymeric serpin conformation. The pH dependent species was revealed to be a low molecular weight species, with a sediment coefficient of 1.8 S (approximately 45 kDa), that underwent a thermal transition with a midpoint of 45 °C and also comprised of α -helix and β -sheet content, properties of the native serpin fold.

The pH dependent conformational change of each of the H255A:H256A species was probed by the pH titration experiment.

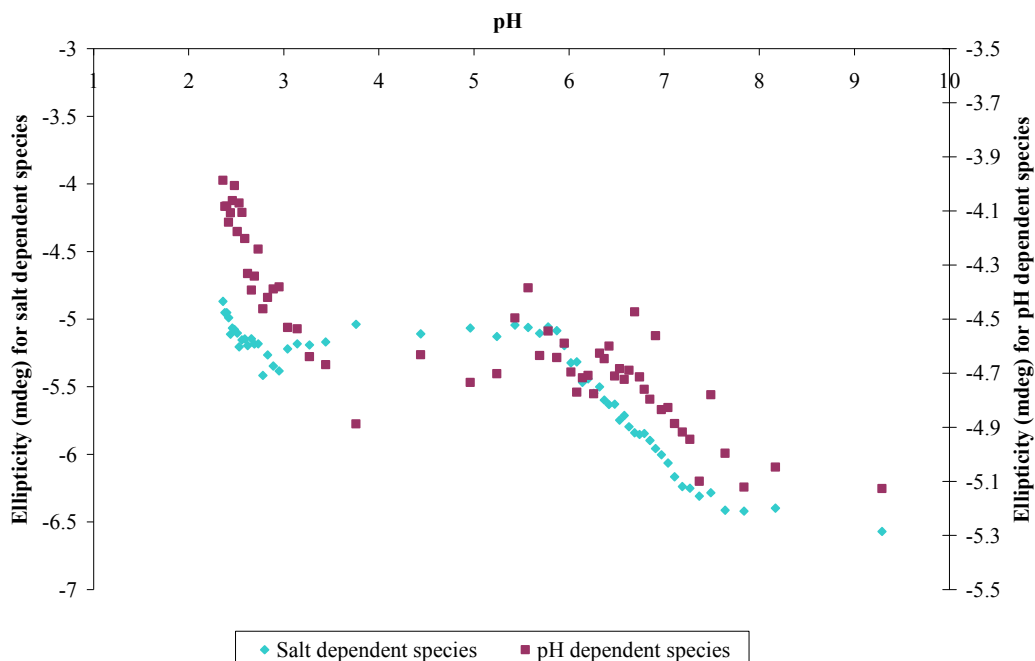


Figure 7.8. pH titration of H255A:H256A mutant. Reduction of the pH from alkali to acid resulted in a conformational transition for both the salt dependent and pH dependent species. Both species showed a transition with a midpoint of pH 6.6.

As both the salt dependent and pH dependent species demonstrated a conformational transition, it implied that mutation of histidines 255 and 256 did not alter the pH dependency of the protein. This therefore suggested that H255 and H256 are not involved in the mechanism of conformational change of mHsp47.

The activity of the species was investigated by the collagen fibrillisation assay which determined protein activity by measuring its anti-fibrillisation efficiency.

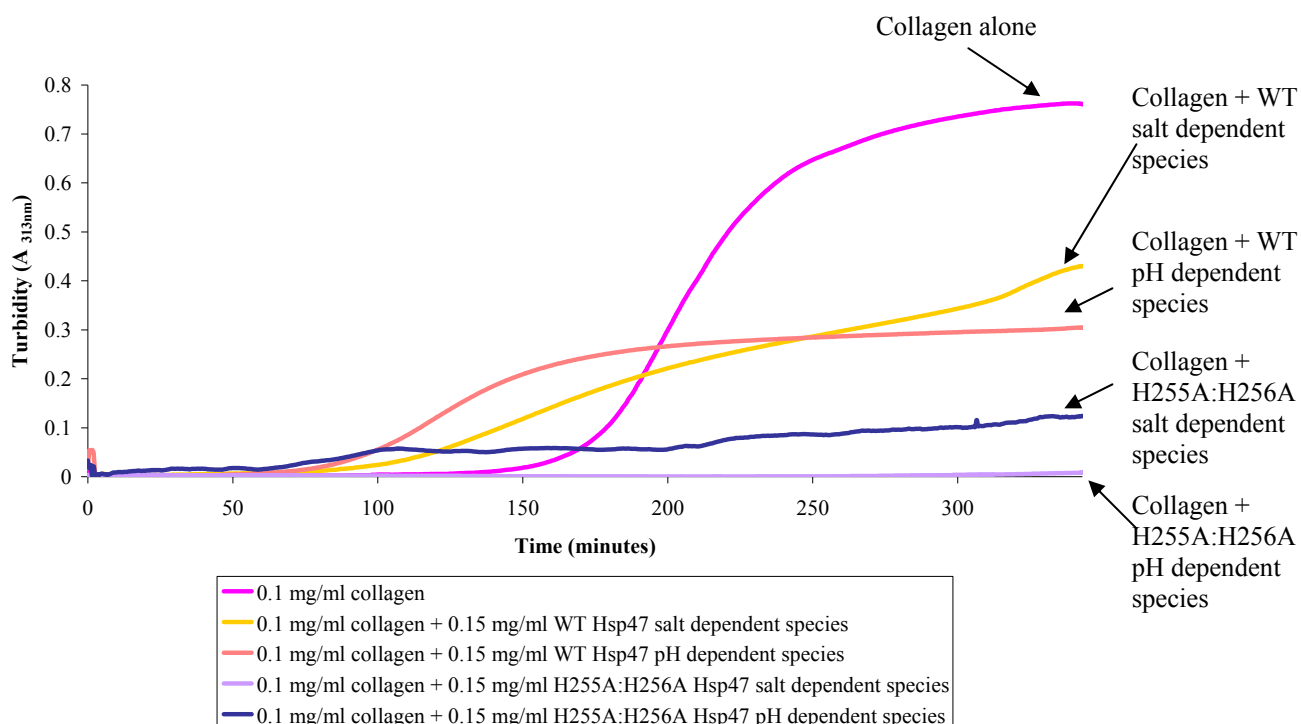


Figure 7.9. Effect of H255A:H256A mutant on collagen fibrillisation. The salt dependent species produced by the mutant suppressed fibril formation completely as the turbidity remained constant and no nucleation, growth or deceleration phase was seen. The H255A:H256A pH dependent species greatly reduced collagen fibrillisation as the final turbidity was 0.12, compared with 0.73 of collagen alone.

The results of the collagen fibrillisation assays revealed that both the mutant species were effective at reducing collagen fibrillisation. The H255A:H256A salt dependent species completely abolished fibril formation as the turbidity was kept at the basal level throughout the assay. The pH dependent species, although less effective than the salt dependent species, markedly diminished collagen fibril formation as turbidity of the solution did not exceed 0.13. Furthermore, each of the mutant species was found to be more potent at anti-fibrillisation of collagen than their WT equivalent. The presence of WT salt dependent species did decrease collagen fibrillisation compared with collagen alone, reducing the turbidity from 0.73 to 0.43. However this was less so than the H255A:H256A salt dependent species which did not show any significant increase in the turbidity, suggesting lack of fibril formation. In the same context, the WT pH dependent species reduced the turbidity to 0.30 and the H255A:H256A

pH dependent species decreased this further still to 0.13. These results therefore implied that mutation of histidines 255 and 256 enhanced the biological activity of mHsp47.

7.3 Discussion

This double mutant was produced to investigate the importance of the histidines near the Hsp47 gate region in the conformational transition of Hsp47. Histidines 255 and 256 are located in close proximity to the end of β -sheet C, forming the gate cluster (Figure 7.1). This region is important in the conformational dynamics of serpins but it has not been probed in terms of Hsp47 and its interaction with collagen. The gate region undergoes movement upon formation of the latent serpin state, and lies adjacent to the gate cluster. It was therefore feasible to propose that the histidine gate cluster might be involved in the serpin fold or conformational transition of Hsp47.

Protein elution by a pH gradient from collagen produced two species; a salt dependent and pH dependent species. This is consistent with WT, H191A and H197:H198A mutants, suggesting that mutation of histidine at positions 255 and 256 does not affect the production of two distinct species. Comparison of the chromatogram for the H255A:H256A mutant with those of the H191A and H197A:H198A mutants demonstrated a slight increase in the salt concentration at which this species (H255A:H256A) eluted from the gelatin-agarose column. The chromatogram for the H255A:H256A mutant (Figure 7.3) showed that a salt concentration of 8.278 mS/cm was required for the release of further salt dependent species. The release of all the H191A and H197A:H198A salt dependent species was at 4.644 and 4.654 mS/cm, respectively. In comparison to this, the release of all the WT salt dependent species required a salt concentration of 20.331 mS/cm. The observation that the mutants required a reduced salt concentration to be eluted from collagen compared to its WT

counterpart is another fascinating finding as it suggests that mutating histidine residues perturbed the binding of the salt dependent species. As the release of the salt dependent species from the gelatin-agarose column was instigated by a lower salt concentration, it seems that substitution of histidine for alanine reduced the binding affinity of the salt dependent species for collagen. Ionic interactions are amongst those that would be disrupted upon substitution of histidine residues for alanine. A reduction in these interactions would result in the destabilisation of the whole molecule. This suggests that the double mutation of the histidines in the gate cluster may diminish the binding affinity of Hsp47 salt dependent species to collagen as it required a reduced salt concentration, compared with WT, to be released. This mutant, however, required a higher salt concentration for the release of all the salt dependent species compared with the H191A and H197A:H198A mutants, implying that mutation of histidine at positions 191, 197 and 198 further reduced the affinity of the salt dependent species for collagen. It can therefore be said that the breach region and gate cluster are important in the binding affinity of the salt dependent species of Hsp47 to collagen and the breach region is more so than the gate cluster.

Mutation of the histidines showed a less pronounced effect on the H255A:H256A mutant pH dependent species, however, with maximal elution at pH 6.08. This was close to pH 6.1 of WT pH dependent species, implying that mutating histidines 255 and 256 did not alter the binding affinity of the pH dependent species for collagen. Therefore, it can be proposed that the binding site of collagen on Hsp47 does not comprise the gate cluster.

The pH titration experiments revealed that both the salt dependent and pH dependent species of the H255A:H256A mutant produced the characteristic two-phase transition. This suggested that histidines 255 and 256 were not involved in the acid-induced conformational transition of Hsp47.

The results of secondary structure analysis were similar to that of WT mHsp47. The CD spectra were characteristic of α -helix and β -sheet containing, folded proteins. The thermal denaturation curves were consistent with the WT form of the protein in that the salt dependent species showed resistance to temperatures as high as 90 °C with only slight unfolding whilst the pH dependent species underwent a transition with a midpoint of 45 °C. The midpoint of the transition for the pH dependent species was lower than for the WT (55 °C), implying that mutation of histidines 255 and 256 reduced the stability of this species. The AUC data was again similar to that of WT mHsp47 whereby the salt dependent species produced multiple peaks, denoting the presence of oligomeric species and the pH dependent species revealed a dominant single peak, representing monomeric protein. A difference to be noted, however, the potentially trimeric peak produced by WT pH dependent species was not seen for the H255A:H256A mutant. Therefore, it was possible that mutation of histidines 255 and 256 reduced the likelihood of trimer formation for this species.

The ability of the mutant species to bind and suppress collagen fibrillisation implied that the histidines that made up the gate cluster were not part of the binding site of mHsp47 to collagen. Furthermore, mutation of H255 and H256 enhanced the de-fibrillisation activity of both the salt dependent and pH dependent species of the protein in comparison to the WT equivalents. For the WT salt dependent species and pH dependent species, nucleation, growth and deceleration phases were evident and a final turbidity of 0.43 and 0.30 was reached, respectively. The H255A:H256A mutant salt dependent and pH dependent species, on the other hand, did not show clear nucleation, growth or deceleration phases, with the turbidity increase being very slow indeed. The salt dependent species prevented any collagen fibrillisation from taking place, maintaining the turbidity at baseline levels throughout the experiment, whilst the pH dependent species produced a final turbidity reading of 0.13. The

reduced turbidity levels corresponded to reduced fibril formation in the presence of mutant species, indicating that the histidine to alanine mutations improved the biological activity of mHsp47.

The enhanced biological activity of the H255A:H256A mutant was an interesting finding as the mutation of the histidines did not alter the binding affinity or the ability of the mutant to undergo an acid-induced conformational transition in comparison to the WT equivalent. The secondary structure analysis was also similar to the WT, indicating a folded, thermostable conformation for the salt dependent species and a folded, metastable state for the pH dependent species. Therefore such an improvement in collagen anti-fibrillisation is an interesting finding. It could be that the substitution of histidine with alanine residues removed any hydrogen bond interactions, allowing a less restrained Hsp47 structure that was able to bind collagen in a way that interfered with collagen fibril formation more readily. It could also be that this less restrained H255A:H256A mutant had more than one point of contact with its substrate, allowing further 'bending' of collagen chains, which ultimately resulted in reduced collagen fibril formation.

		WILD TYPE			H197A: H198A			H255A: H256A			H191A			H220A		
		pH6	pH9	FT	pH6	pH9	FT	pH6	pH9	FT	pH6	pH9	FT	pH6	pH9	FT
Protein purification		**	**	-	*	*	**	*	**	*	*	**	*	-	-	**
AUC	Monomer	✓			✓			✓			✓					✓
	Oligomer		✓			✓			✓			✓				✓
CD	Spectrum	F	F		F	F		F	F		F	F				F
	Melt	✓	✗		✗	✗		✓	✗		✗	✗				✗
Collagen fibrillisation	Lag phase	<	<		>	<		>	n/e		>	<				<
	Fibril formation	<	<		<	<		<	<		<	>				n/e
pH titration	Transition	✓	✓		✓	✓		✓	✓		✓	✓				✓
	Spectrum	F	F		F	F		F	F		F	F				F

F: folded

n/e: no effect

Table 7.1. Summary of results for *Mus musculus* Hsp47.

CHAPTER 8: Crystallography

8.1 Introduction

X-ray crystallography is a means by which the arrangement of atoms within a crystal can be determined. When a beam of X-rays makes contact with a crystal, it scatters into many directions. The angles and intensities of these scattered beams can be used to construct a three-dimensional model of the electron density within the crystal. This density of electrons can be used to deduce the position of atoms within the crystal, as well as their chemical bonds and disorder.

Protein crystals are most commonly grown in solution and involve a slow decrease in solubility of the protein. Crystal growth in solution is characterised by two processes; nucleation and growth (Figure 8.1). Identification of solution conditions that favour the formation of a single, large crystal is important as larger crystals provide higher resolution.

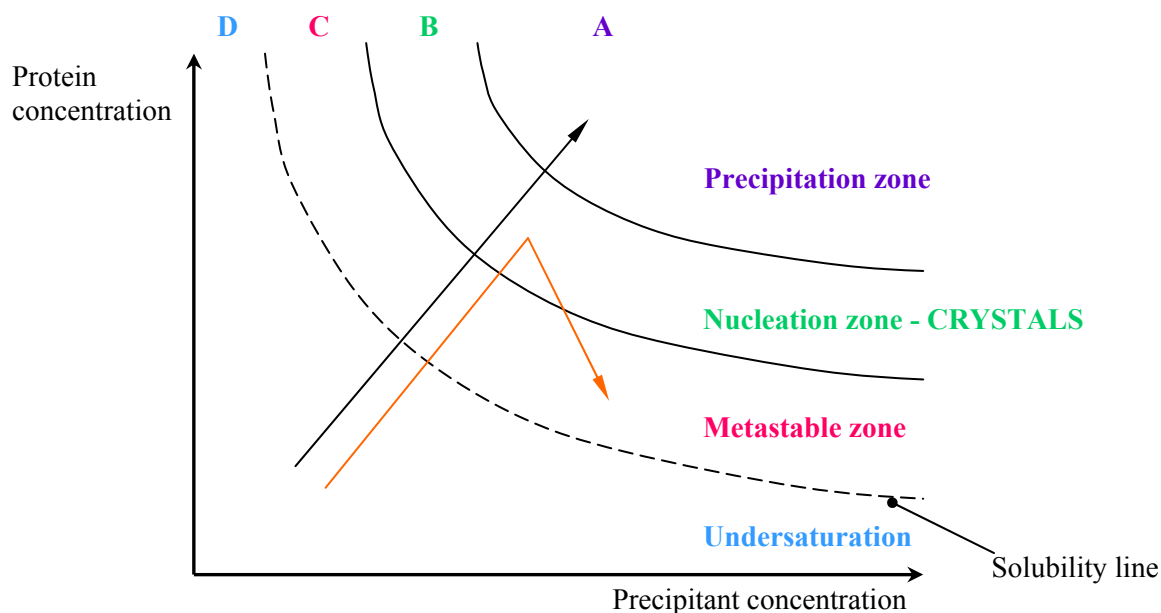


Figure 8.1. X-ray crystallography phase diagram. The space represented when plotting the concentration of crystallisable protein against concentration of precipitant can be divided into several areas.

High concentrations of both protein and precipitant result in protein precipitation and formation of amorphous material. Lowering the concentrations may result in the formation of crystal nuclei that can develop into diffracting crystals. A further reduction in concentration stops nuclei formation and therefore crystals do not appear. However, if a nucleus is placed in this solution it will form a large crystal. This area where crystal growth but not nucleation can proceed is the metastable zone. The lowest protein and precipitant concentration results in completely soluble protein and therefore undersaturation. As it is difficult to predict beneficial conditions for the nucleation or growth of well-ordered crystals, screening is used to identify favourable conditions.

Towards the end of the research period, a pilot attempt was made to produce crystals of Hsp47. The latent state of the protein that was discovered in this research was used for the crystal trays due to its enhanced stability and resistance to denaturation. Latent Hsp47 from the H197A:H198A variant was concentrated to approximately 4 mg/ml using a viva spin column. Subsequently the protein was dialysed against 50 mM Tris-HCl, pH 8.0 to remove the phosphate that was present in the elution buffer. The ‘Hanging Drop Method’ was used for growing protein crystals whereby a drop of protein solution was suspended over a reservoir of buffer and precipitant (Figure 8.2).

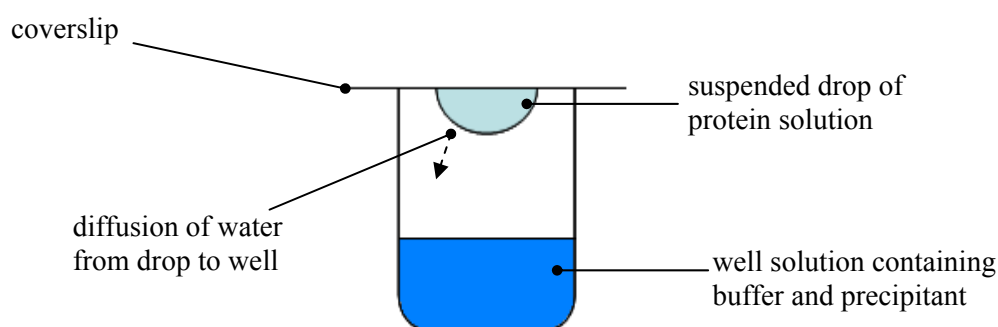


Figure 8.2. The Hanging Drop Method.

Structure and JCSG commercial screens were used to set up crystal trays to determine optimum conditions for crystal growth. Out of the many conditions used a few produced interesting results that could be Hsp47 crystals. All potential Hsp47 crystals shown below were grown for just over a year.

8.2 Results

Results from JCSG Screen

Buffer conditions: 0.02 M calcium chloride, 0.1 M sodium acetate, pH 4.6 and 30 % v/v MPD.

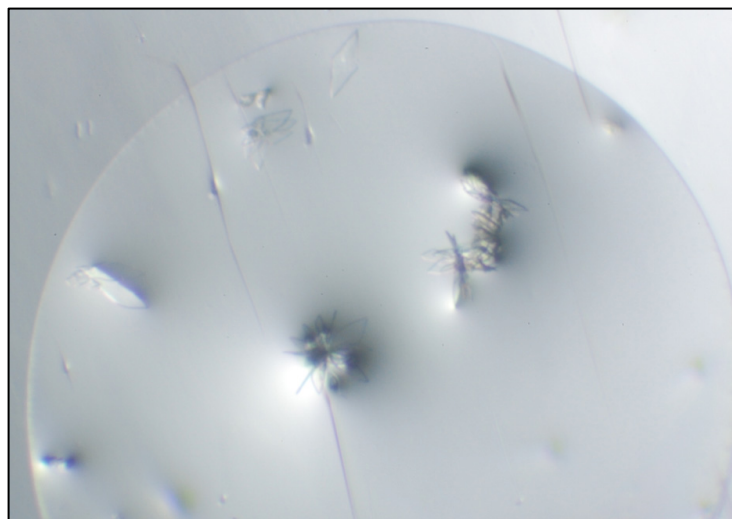


Figure 8.3. Crystal growth in the presence of 0.02 M calcium chloride, 0.1 M acetate and 30 % v/v MPD. The larger structures in the centre of the photograph illustrate crystals.

Buffer conditions: 0.2 M zinc acetate, 0.1 M sodium acetate, pH 4.5 and 10% w/v PEG 3000.

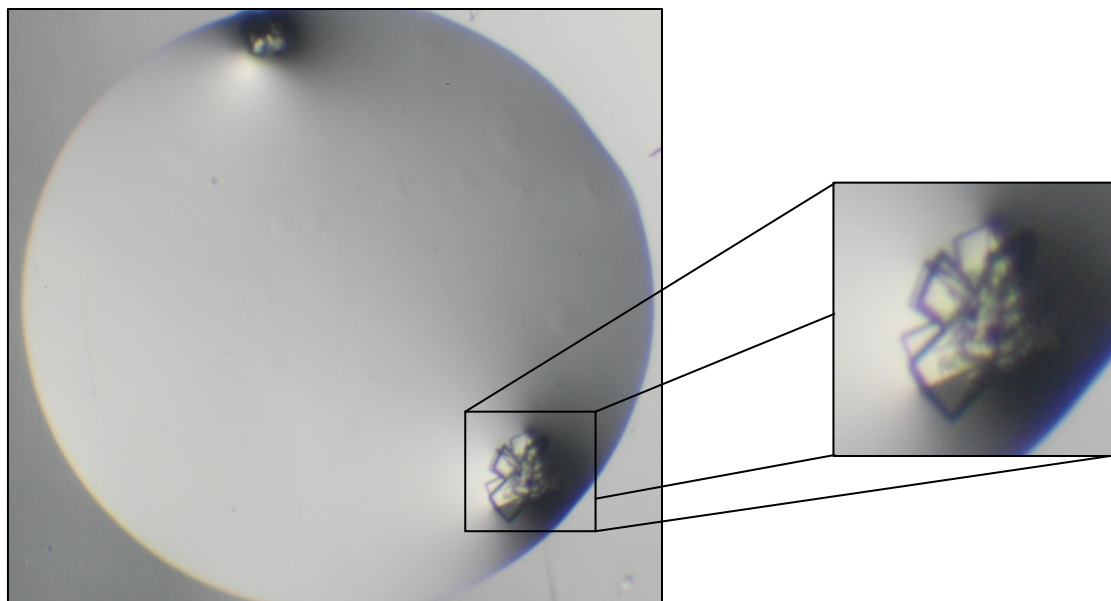


Figure 8.4. Crystals growth in the presence of 0.2 M zinc acetate, 0.1 M sodium acetate and 10% w/v PEG 3000.

Results from Structure Screen

Buffer conditions: 0.1 M Na Hepes, pH 7.5 and 1.5 M lithium sulphate monohydrate.

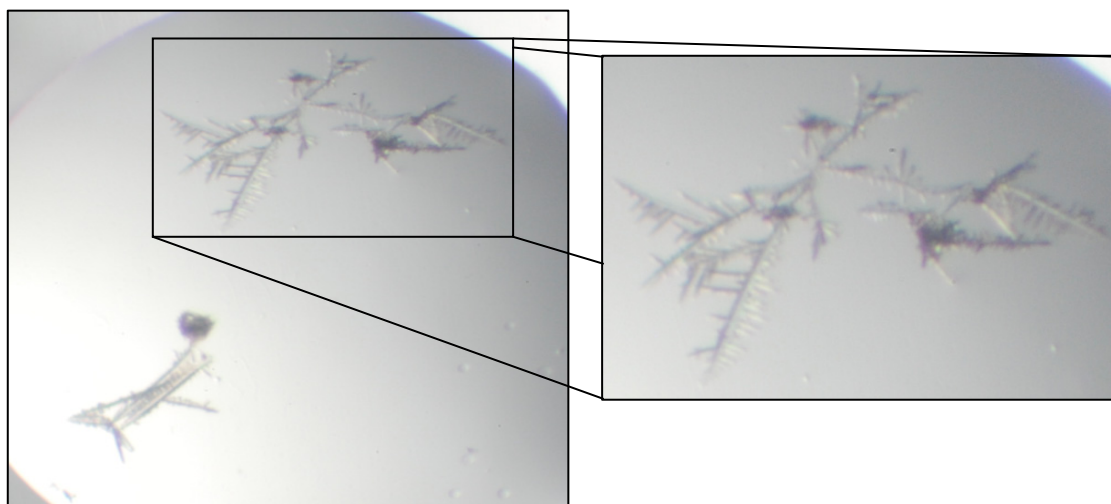


Figure 8.7. Crystal growth in the presence of 0.1 M Na Hepes and 1.5 M lithium sulphate monohydrate.

Buffer conditions: 0.01 M nickel chloride hexahydrate, 0.1 M Tris, pH 8.5 and lithium sulphate.

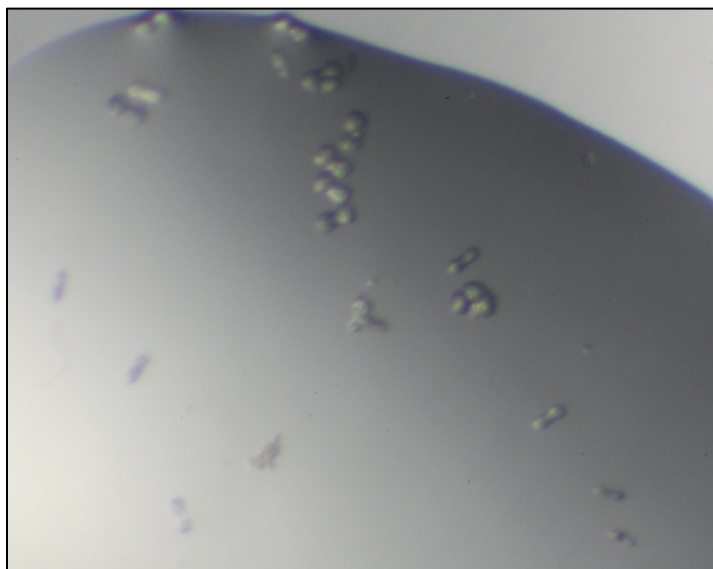


Figure 8.8. Crystal growth in the presence of 0.01 M nickel chloride hexahydrate, 0.1 M Tris and lithium sulphate.

Buffer conditions: 0.2 M magnesium acetate tetrahydrate, 0.1 M Na cacodylate, pH 6.5 and 20% PEG 8000.

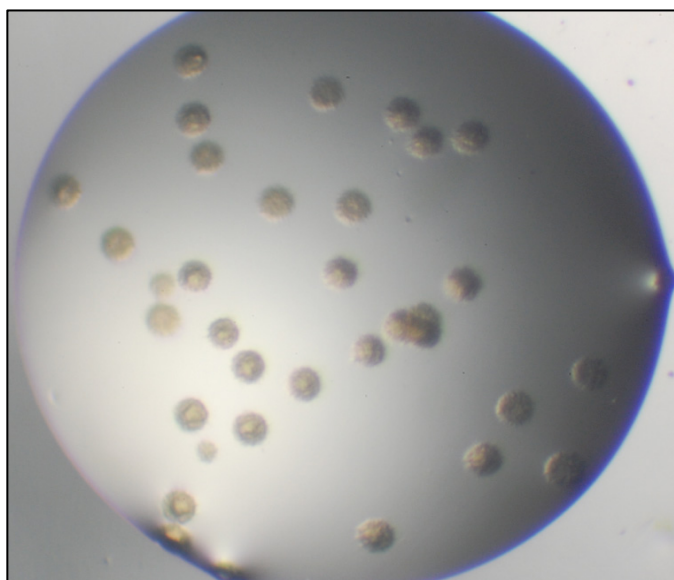


Figure 8.9. Crystal growth in the presence of 0.2 M magnesium acetate tetrahydrate, 0.1 M Na cacodylate and 20% PEG 8000.

Buffer conditions: 0.1 M Na HEPES, pH 7.5 and 1.5 M lithium sulphate monohydrate.

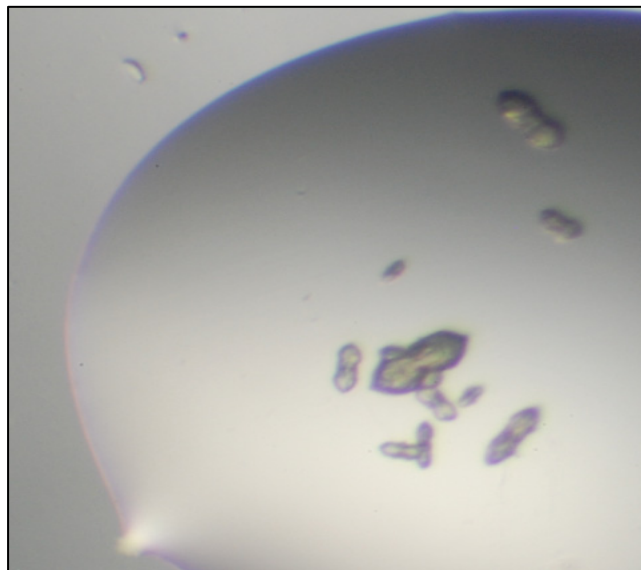


Figure 8.10. Crystal formation in the presence of 0.1 M Na HEPES and 1.5 M lithium sulphate monohydrate.

Due to time restrictions, no further attempts have yet been made on crystallisation of the protein. However, the results obtained look promising and will be used to continue the crystallisation work.

CHAPTER 9: General Discussion, Conclusions and Future work

9.1 General Discussion and Conclusions

This research represents the investigation of the structure, stability and biological activity of the collagen-specific non-inhibitory serpin, Hsp47 [195, 221]. Though the Hsp47-collagen interaction has been strong enough to withstand disruption even in the presence of 2 M NaCl, a reduction in the pH below 6.3 causes the dissociation of Hsp47 from collagen [180]. Under conditions of stress, Hsp47 plays a major part in the quality control system for procollagen, preventing the secretion of procollagen with an abnormal conformation [75]. An RDEL sequence at the C-terminus of Hsp47 acts as an ER-retention sequence, localising the glycoprotein to the ER and the intermediate compartment between the ER and Golgi body [226]. It is the conservation of this carboxyl terminus sequence, along with N-terminus hydrophobic signal sequence and the potential glycosylation sites that provide part of the sequence homology between Hsp47 from different species.

Previous work has examined the expression of the Hsp47 gene in *Mus musculus* (mouse), *Gallus gallus* (chicken), *Ratus* (rat), *Xenopus laevis* (frog), *Homo sapiens* (human) and *Danio rerio* (zebrafish) [110, 227-231]. This study focused on the endothermic *Mus musculus*, poikilothermic *Xenopus laevis* and ectothermic *Danio rerio* species. The difference in properties of these species was considered advantageous as the purification conditions may favour one species more than the others, increasing the possibility of expression of soluble protein. Comparison of structure, stability and activity of the same protein from different species can also provide an insight into the factors that influence their stability, propensity for aggregation and mechanism of action. Previous examination of proteins with similar tertiary folds has provided understanding of the sequence factors involved in their folding mechanism and stability [232-240].

In particular, the folding and stability of members of the serpin superfamily has been extensively studied due to their ability to undergo intrinsic conformational transitions [111-116, 118, 220, 241, 242]. Whilst this conformational change is vital for the inhibitory serpins, a similar structural rearrangement in some human serpins leads to self-association, which subsequently results in various conformational diseases [123, 163, 171, 243, 244]. Non-inhibitory serpins use the conformational transition as part of their function. The native (stressed) forms of the non-inhibitory serpins thyroxine-binding globulin and cortisol binding globulin have a higher affinity for their substrates than their cleaved (relaxed) counterparts. Thus, these serpins do not use cleavage of the RCL and conversion from the stressed to relaxed states for protease inhibition, but instead for the release of ligands [245-247].

The non-inhibitory serpin Hsp47 has not been studied as widely as other members of the superfamily and therefore its biophysical characterisation is lacking detail. Recent studies on the pH dependent conformational switch of Hsp47 and its binding and release of collagen imply that the interaction of the chaperone with its substrate coincides with the pH dependent conformational changes in Hsp47 [87, 97]. Not much focus has been applied to the binding region of collagen on Hsp47 and whether the binding or release is affected by a modification in the amino acid sequence of the glycoprotein. Therefore, this thesis focused on determining the structural regions that were important in the Hsp47-collagen interaction, with the general aim of investigating the ability of the protein to undergo a pH dependent conformational change. Additionally, perturbations in this structural transition following amino acid substitutions by mutagenesis were monitored to observe whether Hsp47 behaved in a manner that was similar to members of other clades of the serpin superfamily.

Cloning and Purification of *Xenopus laevis* and *Danio rerio* Hsp47

The first part of this research concentrated on the optimisation of expression and purification conditions of *Xenopus laevis*, *Danio rerio* and *Mus musculus* Hsp47 in *E. coli*. Much of the previous work on Hsp47 has been hindered by low expression levels and instability of pure protein. Although the above organisms have a highly homologous Hsp47 amino acid sequence resulting in the proteins having a similar fold; characteristics such as protein yield, thermal stability and folding/unfolding pathways may differ from one organism to another. Therefore three different species were used in an attempt to increase the possibility of producing milligramme quantities of pure Hsp47 that could be used to study Hsp47 and its involvement in collagen biosynthesis from a structural perspective.

The majority of the previous work in our lab studied murine Hsp47. However, the expression of this form of the protein has commonly met with aggregation problems, low yields of soluble protein and/or contamination by higher molecular weight oligomers. Therefore, the *Xenopus laevis* and *Danio rerio* gene for Hsp47 was used to try to overcome these obstacles. Difficulties in expressing high yields of recombinant serpins commonly arise due to their tendency to self-associate. For example in *E. coli*, some serpins are expressed in inclusion bodies and consequently require unfolding in the presence of denaturant and subsequent refolding [248]. Taking these characteristics into account, it is likely that the difference in *in vivo* and *in vitro* conditions can affect the expression, solubility and activity of *Xenopus laevis* and *Danio rerio* Hsp47. It is therefore not surprising that these proteins revealed reduced yields and aggregated during expression. In addition to this, the presence of fusion tags in the pET vectors and the inability to cleave the C-terminal His tag interfered with the Hsp47-collagen interaction. Furthermore, the circular dichroism and size exclusion chromatography results implied that the protein was in a partially unfolded state. It is therefore feasible to propose that the C-terminus His tag prevented the correct/complete folding of the protein and

consequently its biological activity. Incomplete or inaccurate protein folding results in residues that would normally be buried within the structure to be exposed to the environment and vice versa, affecting inter- and intra- molecular contacts.

The expression and purification of *Xenopus laevis* and *Danio rerio* Hsp47 in *E. coli* was also attempted using the IMPACT expression system (New England Biolabs) where the gene of interest is fused to an intein with a chitin binding domain, thereby allowing purification on a chitin affinity column. This system was thought to be advantageous as it allowed protein purification in a single step, exploiting the facile cleavage of the guest protein from the fused adduct, intein, in the presence of thiols [249-251]. Thomson and Ananthanarayanan successfully expressed and purified recombinant rat Hsp47 using this technique [249]. They found that induction of the cells at a reduced temperature of 14 °C in comparison to the usual 37 °C resulted in a higher yield of soluble Hsp47. In comparison to the milligramme quantities of pure protein obtained by Thomson and Ananthanarayanan, intein fused-*Danio rerio* Hsp47 from this study did not elute from the chitin affinity beads following on-column cleavage with DTT. However, similar to findings of Thomson and Ananthanarayanan, the majority of the protein was in the insoluble form. Most of the intein fused-*Xenopus laevis* Hsp47 was once again insoluble and although some pure protein was obtained, the concentration was very low, barely visible on SDS-PAGE. Furthermore, as with *Danio rerio*, the majority of the soluble protein applied to the chitin column remained bound following on-column cleavage (even after overnight induction) and subsequent elution. The IMPACT-CN system allows fusion of a self-cleavable intein tag to either the N- or C-terminus of the target protein. Preference for one or the other fusion for different target proteins with structural restraints allows correct folding of the fusion precursor and a higher level of protein expression. pTYB12, an N-terminal fusion vector was used in this study as protein expression was less affected by the target protein and it contained compatible restrictions sites for

cloning. However, this compromised the cleavability of the fused adduct, intein, from the target protein as the disadvantage of this vector was that the on-column cleavage was less efficient than in other vectors, usually requiring a higher induction temperature and a longer incubation time before elution. Unfortunately, this problem with the self-cleavage of the intein tag for this vector could not be overcome by the extended cleavage times or increased temperatures.

Cloning and Purification of *Mus musculus* Hsp47

At this point in the research a decision had to be made between persisting with the *Xenopus laevis* and *Danio rerio* constructs with an N-terminal His tag or moving to a clone of *Mus musculus* Hsp47 with an N-terminal His tag that had become available. Due to the unsuccessful expression of an adequate supply (several milligrammes) of *Xenopus laevis* or *Danio rerio* Hsp47, availability of a successfully cloned construct of *Mus musculus* Hsp47 (by colleagues at Imperial College) that looked promising and the time restraint, it was decided that this form of the protein would be tried for the expression and purification of Hsp47. Advantageously, the mature recombinant mouse Hsp47 gene was cloned into the pCR[®]T7/NT-TOPO[®] vector, which unlike the previously used pET vectors did not contain a C-terminal His tag that could potentially interfere with the correct folding or activity of the protein. As previously published protocols using gelatin [180] or antibody affinity [252] chromatography resulted in low yields and/or aggregation, a novel protocol for protein purification (adapted from [181]) was employed. This procedure proved valuable as it enabled the production of milligramme quantities of Hsp47. The CD spectrum and thermal CD titration results obtained (typical of a serpin fold) were similar to those obtained by Dafforn and coworkers [87], which suggested the presence of a metastable monomer, comprising of a 5-stranded A-sheet and a hyperstable oligomer. A variety of heat-shock proteins and molecular chaperones are known to undergo oligomerisation. For example, the ATPase of BiP

is considered to be due to the conversion of inactive oligomers to active monomers [253]. Heat-shock proteins Hsp12.2 and Hsp12.3 form tetramers [254] while the ER stress protein, Erp29, forms dimers by self-association. Hsp47 has been previously shown to exist in more than one polymerisation state [87, 249, 252]. Dafforn and coworkers identified the oligomeric species as a trimer with a molecular weight of 147 kDa [87]. However, the AUC data for WT Hsp47 in this study demonstrated the oligomer as being only double the weight of the monomer and therefore a dimer. Both the monomeric and oligomeric species were found to be active as each was able to reduce the self-association of collagen into fibrils. This reduction in collagen fibril formation in the presence of Hsp47 is in agreement with the findings of Dafforn and coworkers [87] and Thomson and Ananthanarayanan [249]. The former investigated the interaction between Hsp47 and a collagen mimic peptide (PPG)₁₀ by measuring changes in intrinsic tryptophan fluorescence and reported that Hsp47 prevented collagen fibrillisation and the binding was pH dependent as it was markedly reduced upon reduction of pH from 7.0 to 6.0 [87]. Thomson and Ananthanarayanan examined collagen fibril formation by monitoring turbidity [96, 249] and found that a 1:2 ratio of collagen to Hsp47 resulted in the prevention of collagen fibrillisation [96]. This reduction was observed as being considerably diminished in the presence of reducing agent and Hsp47 together [249]. Thomson and coworkers therefore hypothesised that Hsp47 may need to oligomerise in order to function as a collagen chaperone. Our results however disagree with this proposal as we observed a pronounced reduction in collagen fibril formation by the WT pH dependent species, which consisted only of monomeric Hsp47. Furthermore the inhibition of collagen fibril formation by monomeric Hsp47 was more so than its polymeric counterpart (Figure 4.9). On the other hand, our results are in agreement with those of Nagata and coworkers [222], which demonstrated the binding of Hsp47 monomers to collagen.

Metastability in other proteins

It is not necessary that members of the same protein family, consisting of similar folds and high sequence homology, share similar characteristics such as stability, folding/unfolding pathways and biological activity of a certain fold/state. Frataxin, a small protein involved in iron-sulphur clusters is such an example. The human, yeast and bacterial orthologues of frataxin have high sequence homology and share a similar fold but there is great variability in their stability [255]. The stability of members of the growth hormone family is another example. Bastiras and coworkers studied how the denaturation of porcine growth hormone (pGH) differed from other members of the protein family [256]. Though pGH, human growth hormone (hGH) and bovine growth hormone (bGH) have similar three-dimensional structures [257, 258] and high sequence identities (91% between pGH and bGH and 68% between pGH and hGH) [256], their unfolding pathways differ. Whereby bGH and pGH unfold via a multi-state pathway but hGH exhibits a two-state unfolding mechanism, without an intermediate [259-261]. A final example, which is more closely related to this study, is the difference in activity of members of the serpin superfamily itself. Though all serpins possess essentially the same serpin fold, the non-inhibitory serpins make use of common components in a different manner to that of the inhibitory serpins. As mentioned above, the RCL is a component which is used by inhibitory serpins to trap and inactivate their target protease. Though they have lost their inhibitory activity, the non-inhibitory serpins can utilise the metastability of the fold for the binding and release of ligands [245-247]. These studies reveal how members of the same protein family demonstrate different characteristics. These variations can assist in the identification of factors that cause such changes and site-directed mutagenesis can be used to investigate these variations. Site-directed mutagenesis is a means by which differences in the primary amino acid sequence can be used to examine differences in protein behaviour, highlighting the features that are important for characteristics such as function, stability or activity.

There is compelling evidence that even a single amino acid substitution in a serpin can significantly alter the stability of the protein [21, 123, 163, 171, 262-264]. As described previously, histidine residues are likely to be important in the conformational transition of Hsp47 that occurs upon a reduction in pH from 7.0 to 6.0. This is due to the histidine imidazole group having a pK_a value close to 6.2. It was therefore decided that the importance of histidine residues in the pH dependent conformation of Hsp47 would be examined by substitution to alanine. In the past, Nagata and coworkers have performed site-specific mutagenesis and formed two mutants; a single mutant (H335L) and a double mutant (H197Q:H198N) [97]. Although both variants were reported as having aberrant collagen binding properties (results unpublished), the effect of the substitutions on the structure of Hsp47 was not investigated. We examined the effect of histidine substitutions on the structure, stability and activity of Hsp47 and subsequently investigated the effects of the mutations on collagen binding properties of the chaperone. Mutants of Hsp47 were designed, taking into consideration the regions of other serpins that were thought to be important in the conformational transition. Histidines residues that were highly conserved amongst the serpin superfamily were omitted as it was thought unlikely that these residues would play such an important role in the pH dependent conformational change of Hsp47.

As expected, the strategic rational design approach proved very successful; allowing the identification of residues that are vital for the Hsp47-collagen interaction as well as residues which when mutated, enhance the chaperone activity of Hsp47. Four variants were synthesised: H191A, H197A:H198A, H220A and H255A:H256A (histidine positions illustrated in Figure 9.1). As His 197 and His 255 each had an adjacent His residue (His 198 and His 256, respectively), double mutants were designed to determine the effect of knocking out consecutive histidine residues. It is important to stress that the mutants made in this study were not the same as those made by Nagata and coworkers (unpublished) and they were

ultimately used to probe the pH dependent conformational change of the protein. Fortuitously, the mutants also allowed investigation of the binding site of collagen on Hsp47, examination of structural changes in Hsp47 as well as its ability to act as collagen chaperone.

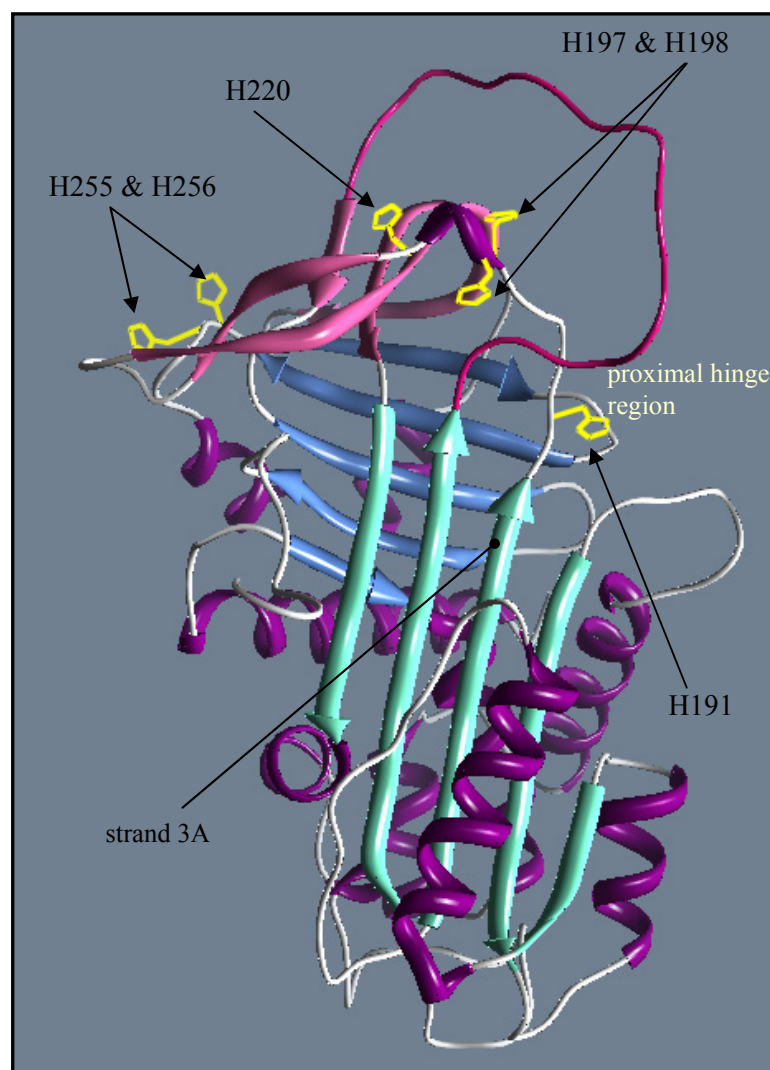


Figure 9.1. Model structure of Hsp47. Illustrating the positions of histidine residues that were mutated.

Firstly, WT Hsp47 revealed results that had not been seen prior to this study. Whilst the pH dependent monomer has been commonly observed, the salt concentration dependent polymer is distinct from any form of Hsp47 identified previously. Though Dafforn and coworkers have identified polymeric Hsp47 in the past [87], they reported a trimeric species whilst this research revealed a dimeric species. More interestingly however, the dimer was released from collagen by an increase in salt concentration rather than a drop in pH, as was the case for the

trimer. This observation of a pH dependent, monomeric and salt concentration dependent, dimeric species simultaneously suggested two forms of Hsp47 that associate with collagen via different interactions. As the polymer dissociated from the gelatin column upon an increase in salt concentration, ionic bonds are likely to prevail between polymeric Hsp47 and collagen. On the other hand, the monomer does not dissociate from collagen after an increase in salt concentration implying that monomeric Hsp47-collagen contact is likely due to hydrophobic contact and/or hydrogen bonds. Additionally, polymeric Hsp47 (identified as a dimer by AUC) possessed beneficial characteristics contrary to the commonly believed hypothesis that in the majority of proteins, β -sheet expansion and subsequent polymerisation is disadvantageous [126, 265]. The results from the collagen fibrillisation assay revealed that dimeric Hsp47 retained some collagen anti-fibrillisation activity, reducing collagen fibril formation in comparison to collagen alone, albeit that this reduction was not as pronounced as the effect of monomeric Hsp47. Therefore, once again, these results disagree with the earlier hypothesis by Thomson and coworkers [249] and Dafforn and coworkers [87] that Hsp47 functions only as a monomer or trimer as the dimer revealed in this study also showed functionality.

Histidines 191, 197 and 198 form a breach cluster near the hinge region of Hsp47. This cluster was particularly noteworthy as it is in the region of serpins that is involved in the conformational change that accompanies insertion of the RCL into the central β -sheet (A-sheet). In the past, this region has been shown to be important for the function of other serpins [21, 125, 198, 200, 201]. Both the single mutant, H191A and the double mutant, H197A:H198A induced a change in the state of the serpin to the latent state. Formation of the latent state led to an increase in the stability of the protein and enhanced the anti-fibrillisation activity of Hsp47. This was a surprising result as it was expected that such a large adjustment in the conformation of Hsp47 would affect the collagen binding site of the chaperone, thereby

reducing its biological activity. This leads to the hypothesis that the binding site is not situated near the region of the serpin that is disrupted during latent state formation. It was also expected that the change to the stabilised structure represented by the latent conformation would constrain the protein in a way that altered the pH dependent conformational transition of the serpin. This effect was also not observed. As both variants retained the ability to undergo the pH induced conformational change, it was evident that histidines 191, 197 or 198 were not the sole residues involved in controlling the pH induced conformational transition of Hsp47.

Implications of the latent state of Hsp47 on structure and function

The induction of an active latent serpin state by these variants is novel. Firstly, this is the initial observation of the latent conformation for Hsp47 and secondly, this is the first sighting of a latent serpin conformation that is active. Although the latent conformation has been observed for PAI-1 [139] and α_1 -AT [220], both these proteins have been inactive [139, 163, 220]. Conversely, the latency conversion of Hsp47 had the opposite effect on the activity of the chaperone. Rather than losing its functionality, latent Hsp47 was more effective at preventing collagen fibril formation. A ruptured hydrogen-bonded network is the likely reason for this increased anti-fibrillisation activity. It may be that the latent form simply has increased biological activity or it could be that the disruption of hydrogen bonds removes some of the structural restraint on the protein, producing a more flexible molecule that can form additional/stronger contacts with collagen chains and consequently reduce collagen fibril formation.

The spontaneous conversion of serpins from the native to latent state is prevented by a hydrogen-bonded network [101] and examination of polar residues that are highly conserved in all serpins has identified a buried polar cluster in the shutter region (beneath the A-sheet)

that is found to stabilise both the stressed and relaxed serpin states via a rearrangement of hydrogen bonds [124]. The critical role of these regions has been discussed with the structure of PAI-2, in which both backbone and sidechain atoms such as Thr, Ser, Asn and His form circles of hydrogen bonds [124]. Additionally, asparagine and histidine residues on β -strands within the A-sheet straddle the site of β -strand insertion. Insertion of the RCL to form s4A requires disruption of this network and consequently some of the hydrogen bonds are broken in PAI-2. In the stressed state, the RCL is held in position by a network of hydrogen bonds in the proximal hinge supported by sidechain electrostatic interactions. The C-termini of s3A and s4A in the central β -sheet are prised apart by the steric effects of Trp194 and the electrostatic interactions restraining the sidechain of Glu342 at the C-terminus of s5A. Interaction with protease results in the breaking of hydrogen bonds holding the proximal hinge (residues P13-P16). Upon formation of a covalent acyl-intermediate with the protease, β -strand s4A partially inserts into the A-sheet up to about P14-P10. Subsequently, there is rearrangement of the hydrogen-bonded pattern by the buried polar residues, allowing separation of the strands in the middle section of the central β -sheet [124]. Examination of this region in the partially inserted form of intact antithrombin revealed many ruptured hydrogen bonds between strands s3A and s5A [266]. In a normal β -sheet, optimal inter-strand interactions occur by cross-strand residue pair interactions [267]. These inter-strand sidechain interactions are important for protein stability [268]. Examination of other stressed proteins revealed that this hydrogen bond network is conserved in α_1 -AT [153] and ovalbumin [269]. The stability of this hydrogen-bonded network is enhanced by the balance of opposite charges on the proximal hinge of the RCL and the loop between s3A and the proceeding residues. The former usually carries a charge of -1 whilst the latter is normally neutral or +1 [124]. According to the Hsp47 model structure (Figure 9.1), histidines 191, 197 and 198 are in close proximity to both the proximal hinge region and the residues proceeding s3A in β -sheet A.

Formation of the latent serpin conformation by the H191A and H197A:H198A variants can be explained in terms of this network of hydrogen bonds. Substitution of histidine residues for alanine in the hinge region would result in a disruption of the hydrogen-bonded network that prevents premature insertion of the RCL in the A-sheet. This is because of the drastic difference in the side chains of histidine and alanine, whereby the former can form hydrogen bonds but the latter cannot. Disruption of hydrogen bonds above the A-sheet would result in concomitant breakage of bonds in the central β -sheet, allowing insertion of the RCL as s4A and consequent latent state formation. Furthermore, the mutations would also compromise the charge balance between the proximal hinge of the RCL and the loop between s3A and the subsequent residues, destabilising the native structure.

Effect of mutations on pH dependent conformational change

Additionally, a difference in the pH at maximum protein elution of the pH dependent species was noted upon further examination of the elution profiles of WT, H191A and H197A:H198A. Whereas WT and H191A mutant pH dependent species eluted at approximately pH 6.1, the H197A:H198A variant equivalent eluted at approximately pH 5.8. This indicated that substitution of histidines 197 and 198 increased the affinity of Hsp47 for collagen as the H197A:H198A variant required a lower pH to be eluted from the gelatin column. This shift in pH suggests that H197 and H198 are in the region that is involved the conformational transition of Hsp47, although these residues do not by themselves account for the complete conformational switch. As the substitutions did not abolish the collagen binding properties of Hsp47; it may be that the alteration in pH dependency of the conformational change was due to the loss/weakening of interactions within the molecule. That is to say that histidines 197 and 198 are important for the structural transition of Hsp47, not because they themselves are involved in the pH dependent conformational change but because of the

intramolecular interactions that they form. The significance of this cluster was to be expected due to the location of the histidines, in the region near the RCL and the A-sheet.

Histidines 255 and 256 form the gate cluster and are located in a turn between β -sheet C and α -helix G. As histidine 256 is conserved within the Hsp47 proteins from different species it was thought likely to be important in the conformational change of the chaperone. The results for the CD spectra and thermal titrations revealed that this mutant behaved similar to its WT counterpart, with the salt dependent species being polymeric and hyperstable and the monomeric pH dependent species being metastable. Interestingly however, the collagen fibrillisation assays (Figure 7.9) demonstrated a difference in the effect of the WT and mutant on collagen fibril formation. Both the H255A:H256A monomer and polymer had a more pronounced effect on the prevention of collagen fibril formation than their WT counterparts, implying that the substitution of amino acids near the gate cluster enhanced the biological activity of the chaperone. Even more interesting was the finding that the mutations did not alter the binding of the polymer or the monomer during gelatin affinity purification. This suggests that the variant altered the binding of the protein to collagen in such a way that influenced its biological activity. However it should also be noticed that this did not disturb its conformational transition upon a reduction in the pH. As both species of Hsp47 produced by the H255A:H256A variant were able to undergo a conformational transition upon a reduction in pH and possessed biological activity by interaction with collagen, it is evident that H255 or H256 are not involved in initiating the pH dependent structural changes or the binding site for collagen.

The H220A mutant also produced an extremely exciting result and probably the most interesting in terms of the collagen binding region on Hsp47. Substitution of the histidine for alanine at the 220 position completely abolished and greatly reduced the binding of the pH

dependent and salt dependent species to gelatin, respectively. This suggested that His 220 was important for the binding of Hsp47 to its substrate. The location of His 220 however, below the hinge region at the top of the A-sheet is over 20 Å from helix A, which is thought to be the likely binding site for collagen, according to previous convincing mutation studies (Nagata, unpublished). However, the deletion mutants, CΔ3 and NΔ1 involved removal of large sections of C- and N- terminals, respectively. Therefore it could be that the diminished binding properties of the variants were an effect of the deletion of vast regions of the protein affecting intramolecular interactions rather than location of the binding cleft. It is also possible that these large deletions led to a large destabilisation of the protein fold as a whole. Interestingly no data was presented in this paper to show that the protein was still intact. However recent experiments by collaborators involving deletion of residues on hA by Miller and coworkers (Olerenshaw and coworkers, unpublished) have shown that this region may indeed be important for collagen binding. For these reasons it was expected that H220 would not be involved in the primary binding site of Hsp47 for collagen. However the results from this work indicate that it does affect collagen binding and may instead be part of a secondary binding site.

Taken together, the results for each of the mutants tell a story about the levels of importance of the different regions in Hsp47. Mutations near the A-sheet and RCL, in the breach cluster (H191, H197 and H198) resulted in formation of the latent serpin state, emphasising the significance of this region in the pH dependent conformational change. Furthermore, another mutation (H220) in this region completely abolished Hsp47-collagen binding and consequently presence of the mutant did not affect collagen fibrillisation. This accentuates the significance of H220 in the collagen binding interaction. Mutations at the bottom of β-sheet C, in the gate cluster (H255 and H256) did not perturb the binding/release of Hsp47 from collagen but greatly enhanced the biological activity of the chaperone, markedly reducing

collagen fibril formation in its presence. These data highlight the importance of this region in anti-fibrillation of collagen. The electron microscopy experiments allowed visualisation of the effect of the absence and presence of Hsp47 on collagen fibril formation and consequently identifying a possible mechanism of action of Hsp47.

In conclusion, this research has provided insight into the behaviour of the non-inhibitory serpin, Hsp47 and enabled the discovery of characteristics that were previously unknown. Most importantly, it has allowed identification of the regions that are vital for its chaperone function, stability and biological behaviour. Moreover, the findings of this thesis are in agreement with previous work on the serpin mechanism of action; providing strength to the data. This study has shown that Hsp47, a non-inhibitory serine protease inhibitor, exists in two different species; each requiring different conditions to be eluted from its substrate, collagen. Furthermore, rationally designed mutants of Hsp47 dramatically affected the ability of Hsp47 to bind collagen and its potency to prevent collagen aggregation via its chaperone activity. The physiological relevance of these findings needs to be investigated.

9.2 Future Work

This thesis has attempted to investigate the regions of importance in the structural transition of the non-inhibitory serpin, Hsp47. Many of the questions have been answered and a number of novel findings have been reported. There is a number of ways in which this research can be extended to emphasize the structural significance of histidine residues in Hsp47.

Chapter 3 identified that *Xenopus laevis* and *Danio rerio* Hsp47 proved difficult to clone and purify a suitable amount (milligramme quantities) of stable protein from either of the species. Due to time constraints, *Xenopus laevis* and *Danio rerio* Hsp47 were not cloned into the

pET[®]T7/NT-TOPO[®] construct that proved extremely successful for *Mus musculus* Hsp47. Cloning of the Hsp47 from the former species and purification following the collagen affinity chromatography method would be particularly interesting as it would allow comparison of the behaviour of the chaperone between different species. Even more interesting would be the examination of whether two distinct species (salt dependent and pH dependent) are produced and investigation of the serpin conformation resulting from histidine to alanine amino acid substitutions at the same positions (for example, His 191, 197, 198, 220, 255 and 256). Additionally, thermal denaturation and the effect on collagen fibril formation of the Hsp47 produced by *Xenopus laevis* and *Danio rerio* could be compared to that of *Mus musculus*, which would provide insight into the stability and anti-fibrillisation potency of each of the species. Consequently, the differences and similarities of Hsp47 between different species would enable comparison of the behaviour of the protein between poikilothermic, endothermic and ectothermic vertebrates. Characteristics of the protein in different species would be interesting to investigate.

The results presented in Chapter 4 revealed the latent Hsp47 conformation and subsequently highlighted the differences in the stability and biological activity of the native conformation compared to the latent conformation. It can therefore be speculated that the enhanced biological activity is a direct consequence of the formation of a more stable serpin structure that binds its substrate in such a way that enhances function. Therefore an interesting extension to this part of the study would be to engineer mutations, other than histidine, in the breach region to establish whether it is the substitution of histidine residues that triggers the formation of stabilising interactions. Both the H191A and H197A:H198A variants produced the latent serpin conformation. For the double mutant, however, it is impossible to distinguish whether it is H197A or H198A mutation that results in the conversion from the native to latent state. Synthesising single mutants; H197A and H198A would allow determination of

the histidine(s) that are responsible for causing the structural transition. To further probe latent state formation in Hsp47, it would be intriguing to investigate whether it is only histidine mutations in the breach region that instigate its formation or whether amino acid substitutions in other regions can also result in increased stability of the protein. Mutation of histidine residues in other regions can be performed to elucidate this.

Alanine substitution of His 220 (Chapter 5) revealed that this side chain was crucial for the interaction between Hsp47 and collagen. Additional amino acid substitutions at this position will reveal structural and functional requirements at this locus for chaperone-substrate binding. Furthermore, as this mutant was unable to bind collagen it was speculated that it formed part of a secondary binding site for collagen. Substitution of other amino acids in close proximity to His 220 would be able to verify/dismiss this suggestion. Mutation of residues within this region may provide further insight into the pH dependent conformational mechanism of Hsp47 and/or the putative binding of chaperone to substrate. Collagen fibril formation did not seem to be affected by the presence of the H220A variant. It would be interesting to visualise the collagen fibrils, by electron microscopy, in the presence of the mutant and compare them to those formed in the absence of Hsp47.

Chapter 6 revealed that although the H255A:H256A double mutant showed similar collagen binding properties to WT Hsp47, its biological activity was greatly enhanced as it significantly reduced fibril formation. Modelling to show bonds formed by His 255 and His 256 and interactions diminished by substitution of these histidines to alanine would be advantageous in determining the contacts that facilitate binding of Hsp47 to collagen and subsequently enhance biological activity of the protein. It was also noted that the salt dependent species of the mutant released from the collagen at a much lower salt concentration in comparison to the WT equivalent. Comparison of molecular models illustrating chemical

binds of mutant and WT species would be beneficial in the identification of interactions responsible for the structural transition and the binding/release to/from collagen.

Chapter 8 illustrated that the results from the attempted protein crystallisation produced some promising results. Firstly, diffraction of a light beam could be used to establish whether the crystals were protein or salt. If the crystals were due to Hsp47, they could be used to construct a structure of the latent conformation of Hsp47. If however, it was realised that the crystals were due to salt, other buffer and precipitant conditions to attempt to crystallise the protein. Comparison of crystallography reagents from a number of different suppliers or increased purity of the latent Hsp47 may prove successful.

CHAPTER 10: References

1. Di Lullo, G.A., et al., *Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen*. J Biol Chem, 2002. **277**(6): p. 4223-31.
2. Ridley, M., Genome, 2006. **New York, NY: Harper Perennial** (ISBN 0-06-019497-9).
3. Vendruscolo, M., et al., *Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems*. Philos Transact A Math Phys Eng Sci, 2003. **361**(1807): p. 1205-22.
4. Jaenicke, R., *Stability and stabilization of globular proteins in solution*. J Biotechnol, 2000. **79**(3): p. 193-203.
5. Dobson, C.M. and M. Karplus, *The fundamentals of protein folding: bringing together theory and experiment*. Curr Opin Struct Biol, 1999. **9**(1): p. 92-101.
6. Ellis, R.J. and F.U. Hartl, *Protein Folding and Chaperones*. **ENCYCLOPEDIA OF LIFE SCIENCES**, 2005.
7. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(96): p. 223-30.
8. Ellis, R.J. and F.U. Hartl, *Principles of protein folding in the cellular environment*. Curr Opin Struct Biol, 1999. **9**(1): p. 102-10.
9. Fersht, A.R. and V. Daggett, *Protein folding and unfolding at atomic resolution*. Cell, 2002. **108**(4): p. 573-82.
10. Shakhnovich, E., V. Abkevich, and O. Ptitsyn, *Conserved residues and the mechanism of protein folding*. Nature, 1996. **379**(6560): p. 96-8.
11. Vendruscolo, M., et al., *Three key residues form a critical contact network in a protein folding transition state*. Nature, 2001. **409**(6820): p. 641-5.
12. Karplus, M., *The Levinthal paradox: yesterday and today*. Fold Des, 1997. **2**(4): p. S69-75.
13. Dobson, C.M., *Protein folding and misfolding*. Nature, 2003. **426**(6968): p. 884-90.
14. Dobson, C.M., *The structural basis of protein folding and its links with human disease*. Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1406): p. 133-45.
15. Wong, K.P. and C. Tanford, *Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions*. J Biol Chem, 1973. **248**(24): p. 8518-23.
16. Baldwin, R.L., *Intermediates in protein folding reactions and the mechanism of protein folding*. Annu Rev Biochem, 1975. **44**: p. 453-75.
17. Hua, Q.X., et al., *Hierarchical protein folding: asymmetric unfolding of an insulin analogue lacking the A7-B7 interchain disulfide bridge*. Biochemistry, 2001. **40**(41): p. 12299-311.
18. Dobson, C.M., M. Karplus, Sali, *Protein Folding: Theory and Experiment* Angew. Chem. Int. Ed. Eng, 1998. **37**: p. 868-893.
19. Onuchic, J.N., Z. Luthey-Schulten, and P.G. Wolynes, *Theory of protein folding: the energy landscape perspective*. Annu Rev Phys Chem, 1997. **48**: p. 545-600.
20. Mirny, L.A., V.I. Abkevich, and E.I. Shakhnovich, *How evolution makes proteins fold quickly*. Proc Natl Acad Sci U S A, 1998. **95**(9): p. 4976-81.
21. Im, H., E.J. Seo, and M.H. Yu, *Metastability in the inhibitory mechanism of human alpha1-antitrypsin*. J Biol Chem, 1999. **274**(16): p. 11072-7.
22. Lee, C., et al., *Regulation of protein function by native metastability*. Proc Natl Acad Sci U S A, 2000. **97**(14): p. 7727-31.

23. Ellis, R.J., *Macromolecular crowding: an important but neglected aspect of the intracellular environment*. Curr Opin Struct Biol, 2001. **11**(1): p. 114-9.
24. Dobson, C.M., *Protein misfolding, evolution and disease*. Trends Biochem. Sci, 1999. **24**: p. 329-332.
25. Bullock, A., Fersht, A. R., *Rescuing the function of mutant p53*. Nat. Rev. Cancer, 2001. **1**: p. 68-76.
26. Sacchettini, J.C.a.K., J. W., *Therapeutic strategies for human amyloid diseases*. Nat. Rev. Drug Disc., 2002. **1**: p. 267-275.
27. Dobson, C.M., *Protein misfolding diseases: getting out of shape*. Nature, 2002. **418**: p. 729-730.
28. Prusiner, S.B., *Prion diseases and the BSE crisis*. Science, 1997. **278**: p. 245-251.
29. Kelly, J.W., *The alternative conformation of amyloidogenic proteins and their multi-step assembly pathways*. Curr Opin Struct Biol, 1998. **8**: p. 101-106.
30. Ellis, R.J. and A.P. Minton, *Protein aggregation in crowded environments*. Biol Chem, 2006. **387**(5): p. 485-97.
31. Walsh, D.M., et al., *Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo*. Nature, 2002. **416**(6880): p. 535-9.
32. Jaenicke, R., Seckler, R., *Protein misassembly in vitro*. Adv. Protein Chem., 1997. **50**: p. 1-53.
33. London, J., Skrzynia, C., Goldberg, M., *Renaturation of Escherchia coli tryptophanase in aqueous urea solutions*. Eur. J. Biochem., 1974. **47**: p. 409-415.
34. Speed, M.A., Wang, D. I. C., King, J., *Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition*. Nat. Biotechnology, 1996. **14**: p. 1283-1287.
35. Lazaridis, T. and M. Karplus, *"New view" of protein folding reconciled with the old through multiple unfolding simulations*. Science, 1997. **278**(5345): p. 1928-31.
36. Dinner, A.R., Sali, A., Smith, L. J., Dobson, C. M. and Karplus, M., *Understanding protein folding via free-energy surfaces from theory and experiment*. Trends Biochem. Sci, 2000. **25**: p. 331-339.
37. Booth, D.R., et al., *Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis*. Nature, 1997. **385**(6619): p. 787-93.
38. Thomas, P.J., B.H. Qu, and P.L. Pedersen, *Defective protein folding as a basis of human disease*. Trends Biochem Sci, 1995. **20**(11): p. 456-9.
39. Stefani, M. and C.M. Dobson, *Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution*. J Mol Med, 2003. **81**(11): p. 678-99.
40. Minton, A.P., *Macromolecular crowding*. Current Biology, 2006. **16**(8): p. R269-R71.
41. Ellis, R.J., *Macromolecular crowding: obvious but underappreciated*. TRENDS in Biochemical Sciences, 2001. **26**(10): p. 597-604.
42. Dobson, C.M., Evans, P. A. and Radford, S. E., *Understanding how proteins fold: the lysozyme story so far*. Trends Biochem. Sci, 1994. **19**: p. 31-37.
43. van den Berg, B., Chung, E.W., Robinson, C. V. and Dobson, C. M., *Characterisation of the dominant oxidative folding intermediate of hen lysozyme*. J Mol Biol, 1999a. **290**: p. 781-796.
44. van den Berg, B., Robinson, C. V., Mateo, P. L. and Dobson, C. M., *The oxidative folding of hen lysozyme and its catalysis by protein disulphide isomerase*. Embo J, 1999b. **18**: p. 4794-4803.
45. van den Berg, B., R.J. Ellis, and C.M. Dobson, *Effects of macromolecular crowding on protein folding and aggregation*. Embo J, 1999. **18**(24): p. 6927-33.

46. Goldberg, M.E., Rudolph, R. and Jaenicke, R., *A kinetic study of the competition between renaturation and aggregation during refolding of denatured-reduced egg white lysozyme*. Biochemistry, 1991. **30**: p. 2790-2797.
47. van den Berg, B., et al., *Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell*. Embo J, 2000. **19**(15): p. 3870-5.
48. Ellis, J., *Proteins as molecular chaperones*. Nature, 1987. **328**(6129): p. 378-9.
49. Ellis, R.J., S.M. van der Vies, and S.M. Hemmingsen, *The molecular chaperone concept*. Biochem Soc Symp, 1989. **55**: p. 145-53.
50. Martin, J. and F.U. Hartl, *Molecular chaperones in cellular protein folding*. Bioessays, 1994. **16**(9): p. 689-92.
51. Laskey, R.A., et al., *Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA*. Nature, 1978. **275**(5679): p. 416-20.
52. Ellis, R.J., *Discovery of molecular chaperones*. Cell Stress Chaperones, 1996. **1**(3): p. 155-60.
53. Ellis, R.J. and S.M. van der Vies, *Molecular chaperones*. Annu Rev Biochem, 1991. **60**: p. 321-47.
54. Ellis, R.J., *The general concept of molecular chaperones*. Philos Trans R Soc Lond B Biol Sci, 1993. **339**(1289): p. 257-61.
55. Munro, S. and H.R. Pelham, *An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein*. Cell, 1986. **46**(2): p. 291-300.
56. Pelham, H.R., *Speculations on the functions of the major heat shock and glucose-regulated proteins*. Cell, 1986. **46**(7): p. 959-61.
57. Gething, M.J., K. McCammon, and J. Sambrook, *Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport*. Cell, 1986. **46**(6): p. 939-50.
58. Martin, J. and F.U. Hartl, *Chaperone-assisted protein folding*. Curr Opin Struct Biol, 1997. **7**(1): p. 41-52.
59. Lodish, H.F., *Molecular cell biology*. 4th ed. ed. 1999, New York ; Basingstoke: W.H. Freeman. xxxvi, 1084, G-17, I-36 p.
60. Langer, T., et al., *Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity*. Embo J, 1992. **11**(13): p. 4757-65.
61. Ranson, N.A., H.E. White, and H.R. Saibil, *Chaperonins*. Biochem J, 1998. **333** (Pt 2): p. 233-42.
62. Saibil, H.R., et al., *ATP induces large quaternary rearrangements in a cage-like chaperonin structure*. Curr Biol, 1993. **3**(5): p. 265-73.
63. Horwich, A.L., G.W. Farr, and W.A. Fenton, *GroEL-GroES-mediated protein folding*. Chem Rev, 2006. **106**(5): p. 1917-30.
64. Georgopoulos, C.P., et al., *Host participation in bacteriophage lambda head assembly*. J Mol Biol, 1973. **76**(1): p. 45-60.
65. Schoffl, F., R. Prandl, and A. Reindl, *Regulation of the heat-shock response*. Plant Physiol, 1998. **117**(4): p. 1135-41.
66. Lele, Z., S. Engel, and P.H. Krone, *hsp47 and hsp70 gene expression is differentially regulated in a stress- and tissue-specific manner in zebrafish embryos*. Dev Genet, 1997. **21**(2): p. 123-33.
67. Hogan, B.L., D.P. Barlow, and M. Kurkinen, *Reichert's membrane as a model for studying the biosynthesis and assembly of basement membrane components*. Ciba Found Symp, 1984. **108**: p. 60-74.
68. Taylor, A., B.L. Hogan, and F.M. Watt, *Biosynthesis of EGF receptor, transferrin receptor and colligin by cultured human keratinocytes and the effect of retinoic acid*. Exp Cell Res, 1985. **159**(1): p. 47-54.

69. Cates, G.A., A.M. Brickenden, and B.D. Sanwal, *Possible involvement of a cell surface glycoprotein in the differentiation of skeletal myoblasts*. J Biol Chem, 1984. **259**(4): p. 2646-50.
70. Nagata, K. and K.M. Yamada, *Phosphorylation and transformation sensitivity of a major collagen-binding protein of fibroblasts*. J Biol Chem, 1986. **261**(16): p. 7531-6.
71. Yasuda, K., et al., *The Kruppel-like factor Zf9 and proteins in the Sp1 family regulate the expression of HSP47, a collagen-specific molecular chaperone*. J Biol Chem, 2002. **277**(47): p. 44613-22.
72. Hosokawa, N., et al., *Structure of the gene encoding the mouse 47-kDa heat-shock protein (HSP47)*. Gene, 1993. **126**(2): p. 187-93.
73. Thomson, C.A., R. Tenni, and V.S. Ananthanarayanan, *Mapping Hsp47 binding site(s) using CNBr peptides derived from type I and type II collagen*. Protein Sci, 2003. **12**(8): p. 1792-800.
74. Nagata, K., *HSP47 as a collagen-specific molecular chaperone: function and expression in normal mouse development*. Semin Cell Dev Biol, 2003. **14**(5): p. 275-82.
75. Nagata, K., *Hsp47: a collagen-specific molecular chaperone*. Trends Biochem Sci, 1996. **21**(1): p. 22-6.
76. Takechi, H., et al., *Alternative 5' splice site selection induced by heat shock*. Mol Cell Biol, 1994. **14**(1): p. 567-75.
77. Nagata, K., *Expression and function of heat shock protein 47: a collagen-specific molecular chaperone in the endoplasmic reticulum*. Matrix Biol, 1998. **16**(7): p. 379-86.
78. Hendershot, L.M. and N.J. Bulleid, *Protein-specific chaperones: the role of hsp47 begins to gel*. Curr Biol, 2000. **10**(24): p. R912-5.
79. Lodish, H.F., *Molecular cell biology*. 5th ed. ed. 2003, New York: W. H. Freeman ; Basingstoke : [Palgrave]. xxxiii, 973 , [79] p.
80. Branden, C. and J. Tooze, *Introduction to protein structure*. 2nd ed. ed. 1999, New York: Garland Pub. xiv, 410 p.
81. Maier, A. and R. Mayne, *Distribution of connective tissue proteins in chick muscle spindles as revealed by monoclonal antibodies: a unique distribution of brachionectin/tenascin*. Am J Anat, 1987. **180**(3): p. 226-36.
82. McLaughlin, S.H. and N.J. Bulleid, *Thiol-independent interaction of protein disulphide isomerase with type X collagen during intra-cellular folding and assembly*. Biochem J, 1998. **331** (Pt 3): p. 793-800.
83. Sharma, S.R., et al., *Effect of vitamin C on collagen biosynthesis and degree of birefringence in polarization sensitive optical coherence tomography (PS-OCT)*. African Journal of Biotechnology 2008. **7**(12): p. 2049-2054.
84. Groff, J.L. and S.S. Gropper, *Advanced Nutrition and Human Metabolism*, 1999. 3rd Edn.(West Publishing Company, Belmont).
85. Voet and Voet, *Biochemistry*, 2nd Edn. John Wiley and Sons, Inc. New York, NY, 1995.
86. Koide, T., et al., *Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47*. J Biol Chem, 2000. **275**(36): p. 27957-63.
87. Dafforn, T.R., M. Della, and A.D. Miller, *The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis*. J Biol Chem, 2001. **276**(52): p. 49310-9.
88. Koide, T., S. Asada, and K. Nagata, *Substrate recognition of collagen-specific molecular chaperone HSP47. Structural requirements and binding regulation*. J Biol Chem, 1999. **274**(49): p. 34523-6.

89. Koide, T., et al., *Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation*. J Biol Chem, 2006. **281**(6): p. 3432-8.
90. Koide, T., et al., *Specific recognition of the collagen triple helix by chaperone HSP47. II. The HSP47-binding structural motif in collagens and related proteins*. J Biol Chem, 2006. **281**(16): p. 11177-85.
91. Koide, T., et al., *Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47*. J Biol Chem, 2002. **277**(8): p. 6178-82.
92. Macdonald, J.R. and H.P. Bachinger, *HSP47 binds cooperatively to triple helical type I collagen but has little effect on the thermal stability or rate of refolding*. J Biol Chem, 2001. **276**(27): p. 25399-403.
93. Nagai, N., et al., *Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis*. J Cell Biol, 2000. **150**(6): p. 1499-506.
94. Matsuoka, Y., et al., *Insufficient folding of type IV collagen and formation of abnormal basement membrane-like structure in embryoid bodies derived from Hsp47-null embryonic stem cells*. Mol Biol Cell, 2004. **15**(10): p. 4467-75.
95. Tasab, M., M.R. Batten, and N.J. Bulleid, *Hsp47: a molecular chaperone that interacts with and stabilizes correctly-folded procollagen*. Embo J, 2000. **19**(10): p. 2204-11.
96. Thomson, C.A. and V.S. Ananthanarayanan, *Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro*. Biochem J, 2000. **349 Pt 3**: p. 877-83.
97. T. S. H. El-Thaher, A.F.D., S. Yokota, A. Nakai, K. Nagata & A. D. Miller, *The pH-dependent, ATP-independent interaction of collagen specific serpin/stress protein HSP47*. Prot. Pept. Lett., 1996. **3**: p. 1-8.
98. Kini, R.M., *Serine proteases affecting blood coagulation and fibrinolysis from snake venoms*. Pathophysiol Haemost Thromb, 2005. **34**(4-5): p. 200-4.
99. Huntington, J.A. and R.W. Carrell, *The serpins: nature's molecular mousetraps*. Sci Prog, 2001. **84**(Pt 2): p. 125-36.
100. Lomas, D.A. and R.W. Carrell, *Serpinopathies and the conformational dementias*. Nat Rev Genet, 2002. **3**(10): p. 759-68.
101. Whisstock, J.C., et al., *Conformational changes in serpins: I. The native and cleaved conformations of alpha(1)-antitrypsin*. J Mol Biol, 2000. **295**(3): p. 651-65.
102. Carrell, R.W. and M.C. Owen, *Plakalbumin, alpha 1-antitrypsin, antithrombin and the mechanism of inflammatory thrombosis*. Nature, 1985. **317**(6039): p. 730-2.
103. Hunt, L.T. and M.O. Dayhoff, *A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha 1-proteinase inhibitor*. Biochem Biophys Res Commun, 1980. **95**(2): p. 864-71.
104. Irving, J.A., et al., *Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function*. Genome Res, 2000. **10**(12): p. 1845-64.
105. van Gent, D., et al., *Serpins: structure, function and molecular evolution*. International Journal of Biochemistry & Cell Biology, 2003. **35**(11): p. 1536-1547.
106. Hammond, G.L., et al., *Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors*. Proc Natl Acad Sci U S A, 1987. **84**(15): p. 5153-7.
107. Tsujimoto, M., et al., *Purification, cDNA cloning, and characterization of a new serpin with megakaryocyte maturation activity*. J Biol Chem, 1997. **272**(24): p. 15373-80.

108. Dickinson, J.L., et al., *The C-D interhelical domain of the serpin plasminogen activator inhibitor-type 2 is required for protection from TNF-alpha induced apoptosis*. Cell Death Differ, 1998. **5**(2): p. 163-71.
109. Sugimoto, Y., et al., *Ovalbumin in developing chicken eggs migrates from egg white to embryonic organs while changing its conformation and thermal stability*. J Biol Chem, 1999. **274**(16): p. 11030-7.
110. Hirayoshi, K., et al., *HSP47: a tissue-specific, transformation-sensitive, collagen-binding heat shock protein of chicken embryo fibroblasts*. Mol Cell Biol, 1991. **11**(8): p. 4036-44.
111. Bruch, M., V. Weiss, and J. Engel, *Plasma serine proteinase inhibitors (serpins) exhibit major conformational changes and a large increase in conformational stability upon cleavage at their reactive sites*. J Biol Chem, 1988. **263**(32): p. 16626-30.
112. James, E.L., et al., *Probing the unfolding pathway of alpha1-antitrypsin*. J Biol Chem, 1999. **274**(14): p. 9482-8.
113. Kwon, K.S., et al., *Single amino acid substitutions of alpha 1-antitrypsin that confer enhancement in thermal stability*. J Biol Chem, 1994. **269**(13): p. 9627-31.
114. Cabrita, L.D., J.C. Whisstock, and S.P. Bottomley, *Probing the role of the F-helix in serpin stability through a single tryptophan substitution*. Biochemistry, 2002. **41**(14): p. 4575-81.
115. Wang, Z., J. Mottonen, and E.J. Goldsmith, *Kinetically controlled folding of the serpin plasminogen activator inhibitor 1*. Biochemistry, 1996. **35**(51): p. 16443-8.
116. Powell, L.M. and R.H. Pain, *Effects of glycosylation on the folding and stability of human, recombinant and cleaved alpha 1-antitrypsin*. J Mol Biol, 1992. **224**(1): p. 241-52.
117. Christensen, H. and R.H. Pain, *Molten globule intermediates and protein folding*. Eur Biophys J, 1991. **19**(5): p. 221-9.
118. Kim, D. and M.H. Yu, *Folding pathway of human alpha 1-antitrypsin: characterization of an intermediate that is active but prone to aggregation*. Biochem Biophys Res Commun, 1996. **226**(2): p. 378-84.
119. Dafforn, T.R., et al., *A kinetic mechanism for the polymerization of alpha1-antitrypsin*. J Biol Chem, 1999. **274**(14): p. 9548-55.
120. Whisstock, J., R. Skinner, and A.M. Lesk, *An atlas of serpin conformations*. Trends Biochem Sci, 1998. **23**(2): p. 63-7.
121. Silverman, G.A., et al., *The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature*. J Biol Chem, 2001. **276**(36): p. 33293-6.
122. Wright, H.T. and J.N. Scarsdale, *Structural basis for serpin inhibitor activity*. Proteins, 1995. **22**(3): p. 210-25.
123. Stein, P.E. and R.W. Carrell, *What do dysfunctional serpins tell us about molecular mobility and disease?* Nat Struct Biol, 1995. **2**(2): p. 96-113.
124. Harrop, S.J., et al., *The crystal structure of plasminogen activator inhibitor 2 at 2.0 Å resolution: implications for serpin function*. Structure, 1999. **7**(1): p. 43-54.
125. Lee, K.N., S.D. Park, and M.H. Yu, *Probing the native strain in alpha1-antitrypsin*. Nat Struct Biol, 1996. **3**(6): p. 497-500.
126. Lomas, D.A., et al., *The mechanism of Z alpha 1-antitrypsin accumulation in the liver*. Nature, 1992. **357**(6379): p. 605-7.
127. Picard, V., et al., *Antithrombin Phe229Leu: a new homozygous variant leading to spontaneous antithrombin polymerization in vivo associated with severe childhood thrombosis*. Blood, 2003. **102**(3): p. 919-25.
128. Fraizer, G.C., et al., *In-frame single codon deletion in the Mmalton deficiency allele of alpha 1-antitrypsin*. Am J Hum Genet, 1989. **44**(6): p. 894-902.

129. Seyama, K., et al., *Siiyama (serine 53 (TCC) to phenylalanine 53 (TTC)). A new alpha 1-antitrypsin-deficient variant with mutation on a predicted conserved residue of the serpin backbone*. J Biol Chem, 1991. **266**(19): p. 12627-32.
130. Perry, D.J., et al., *Antithrombin Cambridge II, 384 Ala to Ser. Further evidence of the role of the reactive centre loop in the inhibitory function of the serpins*. FEBS Lett, 1991. **285**(2): p. 248-50.
131. Perry, D.J., et al., *Antithrombin Cambridge, 384 Ala to Pro: a new variant identified using the polymerase chain reaction*. FEBS Lett, 1989. **254**(1-2): p. 174-6.
132. Devraj-Kizuk, R., et al., *Antithrombin-III-Hamilton: a gene with a point mutation (guanine to adenine) in codon 382 causing impaired serine protease reactivity*. Blood, 1988. **72**(5): p. 1518-23.
133. Hopkins, P.C., R.W. Carrell, and S.R. Stone, *Effects of mutations in the hinge region of serpins*. Biochemistry, 1993. **32**(30): p. 7650-7.
134. Lawrence, D.A., et al., *Serpin reactive center loop mobility is required for inhibitor function but not for enzyme recognition*. J Biol Chem, 1994. **269**(44): p. 27657-62.
135. Davis, A.E., 3rd, et al., *C1 inhibitor hinge region mutations produce dysfunction by different mechanisms*. Nat Genet, 1992. **1**(5): p. 354-8.
136. Levy, N.J., et al., *Type II hereditary angioneurotic edema that may result from a single nucleotide change in the codon for alanine-436 in the C1 inhibitor gene*. Proc Natl Acad Sci U S A, 1990. **87**(1): p. 265-8.
137. Hopkins, P.C. and S.R. Stone, *The contribution of the conserved hinge region residues of alpha1-antitrypsin to its reaction with elastase*. Biochemistry, 1995. **34**(48): p. 15872-9.
138. Perani, P., et al., *Mutations on the hinge region of leukocyte elastase inhibitor determine the loss of inhibitory function*. Biochem Biophys Res Commun, 2000. **274**(3): p. 841-4.
139. Mottonen, J., et al., *Structural basis of latency in plasminogen activator inhibitor-1*. Nature, 1992. **355**(6357): p. 270-3.
140. Chang, W.S., et al., *Importance of the release of strand 1C to the polymerization mechanism of inhibitory serpins*. Protein Sci, 1997. **6**(1): p. 89-98.
141. Carrell, R.W. and P.E. Stein, *The biostructural pathology of the serpins: critical function of sheet opening mechanism*. Biol Chem Hoppe Seyler, 1996. **377**(1): p. 1-17.
142. Carrell, R.W., D.L. Evans, and P.E. Stein, *Mobile reactive centre of serpins and the control of thrombosis*. Nature, 1991. **353**(6344): p. 576-8.
143. Gettins, P.G.W., P.A. Patston, and S.T. Olson, *Serpins : structure, function and biology*. Molecular biology intelligence unit. 1996, Austin: R.G. Landes ; New York ; London : Chapman & Hall. 202p.
144. Ye, S., et al., *The structure of a Michaelis serpin-protease complex*. Nat Struct Biol, 2001. **8**(11): p. 979-83.
145. Peterson, F.C., N.C. Gordon, and P.G. Gettins, *Formation of a noncovalent serpin-proteinase complex involves no conformational change in the serpin. Use of 1H-15N HSQC NMR as a sensitive nonperturbing monitor of conformation*. Biochemistry, 2000. **39**(39): p. 11884-92.
146. Lawrence, D.A., et al., *Serpin-protease complexes are trapped as stable acyl-enzyme intermediates*. J Biol Chem, 1995. **270**(43): p. 25309-12.
147. Stratikos, E. and P.G. Gettins, *Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A*. Proc Natl Acad Sci U S A, 1999. **96**(9): p. 4808-13.
148. Loebermann, H., et al., *Human alpha 1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function*. J Mol Biol, 1984. **177**(3): p. 531-57.

149. Plotnick, M.I., et al., *Distortion of the active site of chymotrypsin complexed with a serpin*. Biochemistry, 1996. **35**(23): p. 7586-90.
150. Huntington, J.A., R.J. Read, and R.W. Carrell, *Structure of a serpin-protease complex shows inhibition by deformation*. Nature, 2000. **407**(6806): p. 923-6.
151. Plotnick, M.I., et al., *Heterogeneity in serpin-protease complexes as demonstrated by differences in the mechanism of complex breakdown*. Biochemistry, 2002. **41**(1): p. 334-42.
152. Dafforn, T.R., R.N. Pike, and S.P. Bottomley, *Physical characterization of serpin conformations*. Methods, 2004. **32**(2): p. 150-8.
153. Stein, P.E., et al., *Crystal structure of uncleaved ovalbumin at 1.95 Å resolution*. J Mol Biol, 1991. **221**(3): p. 941-59.
154. Gettins, P., *Absence of large-scale conformational change upon limited proteolysis of ovalbumin, the prototypic serpin*. J Biol Chem, 1989. **264**(7): p. 3781-5.
155. Stein, P.E., D.A. Tewkesbury, and R.W. Carrell, *Ovalbumin and angiotensinogen lack serpin S-R conformational change*. Biochem J, 1989. **262**(1): p. 103-7.
156. Sancho, E., et al., *Conformational studies on plasminogen activator inhibitor (PAI-1) in active, latent, substrate, and cleaved forms*. Biochemistry, 1995. **34**(3): p. 1064-9.
157. Wardell, M.R., et al., *Crystallization and preliminary X-ray diffraction analysis of two conformations of intact human antithrombin*. J Mol Biol, 1993. **234**(4): p. 1253-8.
158. Carrell, R.W., et al., *Biological implications of a 3 Å structure of dimeric antithrombin*. Structure, 1994. **2**(4): p. 257-70.
159. Lawrence, D.A., et al., *Characterization of the binding of different conformational forms of plasminogen activator inhibitor-1 to vitronectin. Implications for the regulation of pericellular proteolysis*. J Biol Chem, 1997. **272**(12): p. 7676-80.
160. Wardell, M.R., et al., *Preparative induction and characterization of L-antithrombin: a structural homologue of latent plasminogen activator inhibitor-1*. Biochemistry, 1997. **36**(42): p. 13133-42.
161. Im, H., et al., *Interactions causing the kinetic trap in serpin protein folding*. J Biol Chem, 2002. **277**(48): p. 46347-54.
162. Im, H. and M.H. Yu, *Role of Lys335 in the metastability and function of inhibitory serpins*. Protein Sci, 2000. **9**(5): p. 934-41.
163. Lomas, D.A., et al., *Preparation and characterization of latent alpha 1-antitrypsin*. J Biol Chem, 1995. **270**(10): p. 5282-8.
164. Bottomley, S.P., P.C. Hopkins, and J.C. Whisstock, *Alpha 1-antitrypsin polymerisation can occur by both loop A and C sheet mechanisms*. Biochem Biophys Res Commun, 1998. **251**(1): p. 1-5.
165. Huntington, J.A., et al., *A 2.6 Å structure of a serpin polymer and implications for conformational disease*. J Mol Biol, 1999. **293**(3): p. 449-55.
166. Zhou, A., et al., *Polymerization of plasminogen activator inhibitor-1*. J Biol Chem, 2001. **276**(12): p. 9115-22.
167. Marszal, E., D. Danino, and A. Shrake, *A novel mode of polymerization of alpha1-proteinase inhibitor*. J Biol Chem, 2003. **278**(22): p. 19611-8.
168. Wilczynska, M., et al., *A redox-sensitive loop regulates plasminogen activator inhibitor type 2 (PAI-2) polymerization*. Embo J, 2003. **22**(8): p. 1753-61.
169. Risse, B.C., et al., *Differentiating cells of murine stratified squamous epithelia constitutively express plasminogen activator inhibitor type 2 (PAI-2)*. Histochem Cell Biol, 1998. **110**(6): p. 559-69.
170. Gettins, P.G., *Serpin structure, mechanism, and function*. Chem Rev, 2002. **102**(12): p. 4751-804.
171. Lomas, D.A., et al., *Alpha 1-antitrypsin Siiyama (Ser53-->Phe). Further evidence for intracellular loop-sheet polymerization*. J Biol Chem, 1993. **268**(21): p. 15333-5.

172. Fabbretti, G., et al., *Genetic variants of alpha-1-antitrypsin (AAT)*. Liver, 1992. **12**(4 Pt 2): p. 296-301.
173. Stein, P. and C. Chothia, *Serpin tertiary structure transformation*. J Mol Biol, 1991. **221**(2): p. 615-21.
174. Yamasaki, M., et al., *Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization*. Nature, 2008. **455**(7217): p. 1255-8.
175. Sivasothy, P., et al., *Pathogenic alpha 1-antitrypsin polymers are formed by reactive loop-beta-sheet A linkage*. J Biol Chem, 2000. **275**(43): p. 33663-8.
176. Corral, J., et al., *Mutations in the shutter region of antithrombin result in formation of disulfide-linked dimers and severe venous thrombosis*. J Thromb Haemost, 2004. **2**(6): p. 931-9.
177. Skinner, R., et al., *The 2.6 Å structure of antithrombin indicates a conformational change at the heparin binding site*. J Mol Biol, 1997. **266**(3): p. 601-9.
178. Sharp, A.M., et al., *The active conformation of plasminogen activator inhibitor 1, a target for drugs to control fibrinolysis and cell adhesion*. Structure, 1999. **7**(2): p. 111-8.
179. Mast, A.E., J.J. Enghild, and G. Salvesen, *Conformation of the reactive site loop of alpha 1-proteinase inhibitor probed by limited proteolysis*. Biochemistry, 1992. **31**(10): p. 2720-8.
180. Saga, S., et al., *pH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein*. J Cell Biol, 1987. **105**(1): p. 517-27.
181. Olerenshaw, D., A.D. Miller, and T.R. Dafforn, *Investigations into the Mechanism of Heat Shock Protein 47 (Hsp47) by Site-Directed Mutagenesis and Biophysical Analyses*. Unpublished, 2006.
182. Davids, J.W., et al., *Modeling the three-dimensional structure of serpin/molecular chaperone HSP47* Bioorganic Chemistry, 1995. **23**(4): p. 427-438.
183. Skinner, R., et al., *Implications for function and therapy of a 2.9 Å structure of binary-complexed antithrombin*. J Mol Biol, 1998. **283**(1): p. 9-14.
184. Roskams, J. and L. Rodgers, *Lab ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench*. 2002.
185. Williams, B.R., et al., *Collagen fibril formation. Optimal in vitro conditions and preliminary kinetic results*. J Biol Chem, 1978. **253**(18): p. 6578-85.
186. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids*. J Mol Biol, 1983. **166**(4): p. 557-80.
187. Yamaguchi-Iwai, Y., et al., *Differentiation of F9 embryonal carcinoma cells induced by the c-jun and activated c-Ha-ras oncogenes*. Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8670-4.
188. Kane, J.F., *Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli*. Curr Opin Biotechnol, 1995. **6**(5): p. 494-500.
189. Ranatunga, W., et al., *Human RAD52 protein has extreme thermal stability*. Biochemistry, 2001. **40**(29): p. 8557-62.
190. Wood, G.C. and M.K. Keech, *The formation of fibrils from collagen solutions. 1. The effect of experimental conditions: kinetic and electron-microscope studies*. Biochem J, 1960. **75**: p. 588-98.
191. Wood, G.C., *The formation of fibrils from collagen solutions. 2. A mechanism of collagen-fibril formation*. Biochem J, 1960. **75**: p. 598-605.
192. Wood, G.C., *The formation of fibrils from collagen solutions. 3. Effect of chondroitin sulphate and some other naturally occurring polyanions on the rate of formation*. Biochem J, 1960. **75**: p. 605-12.
193. Benesusan, H.B. and B.L. Hoyt, *The effect of various parameters on the rate of formation of fibers from collagen solutions*. J. Amer. Chem. Soc, 1958. **80**.

194. Comper, W.D. and A. Veis, *The mechanism of nucleation for in vitro collagen fibril formation*. Biopolymers, 1977. **16**(10): p. 2113-2131.
195. Nagata, K., S. Saga, and K.M. Yamada, *A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein*. J Cell Biol, 1986. **103**(1): p. 223-9.
196. Lee, K.N., et al., *Characterization of a human alpha1-antitrypsin variant that is as stable as ovalbumin*. J Biol Chem, 1998. **273**(5): p. 2509-16.
197. Ryu, S.E., et al., *The native strains in the hydrophobic core and flexible reactive loop of a serine protease inhibitor: crystal structure of an uncleaved alpha1-antitrypsin at 2.7 Å*. Structure, 1996. **4**(10): p. 1181-92.
198. Hekman, C.M. and D.J. Loskutoff, *Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants*. J Biol Chem, 1985. **260**(21): p. 11581-7.
199. Schechter, I. and A. Berger, *On the size of the active site in proteases. I. Papain*. Biochem Biophys Res Commun, 1967. **27**(2): p. 157-62.
200. De Taeye, B., et al., *Immobilization of the distal hinge in the labile serpin plasminogen activator inhibitor 1: identification of a transition state with distinct conformational and functional properties*. J Biol Chem, 2003. **278**(26): p. 23899-905.
201. Tucker, H.M., et al., *Engineering of plasminogen activator inhibitor-1 to reduce the rate of latency transition*. Nat Struct Biol, 1995. **2**(6): p. 442-5.
202. Sivasothy, P.e.a., Respir. Crit. Care Med., 1999. **159**: p. A481.
203. Lomas, D.A., et al., *alpha 1-Antitrypsin Mmalton (Phe52-deleted) forms loop-sheet polymers in vivo. Evidence for the C sheet mechanism of polymerization*. J Biol Chem, 1995. **270**(28): p. 16864-70.
204. Lomas, D.A., et al., *Effect of the Z mutation on the physical and inhibitory properties of alpha 1-antitrypsin*. Biochemistry, 1993. **32**(2): p. 500-8.
205. Eriksson, S., J. Carlson, and R. Velez, *Risk of cirrhosis and primary liver cancer in alpha 1-antitrypsin deficiency*. N Engl J Med, 1986. **314**(12): p. 736-9.
206. Takeda, K., et al., *Greater than normal expression of the collagen-binding stress protein heat-shock protein-47 in the infarct zone in rats after experimentally-induced myocardial infarction*. Coron Artery Dis, 2000. **11**(1): p. 57-68.
207. Rocnik, E., L.H. Chow, and J.G. Pickering, *Heat shock protein 47 is expressed in fibrous regions of human atheroma and is regulated by growth factors and oxidized low-density lipoprotein*. Circulation, 2000. **101**(11): p. 1229-33.
208. Williams, R.S., *Heat shock protein 47 : a chaperone for the fibrous cap?* Circulation, 2000. **101**(11): p. 1227-8.
209. Masuda, H., et al., *Coexpression of the collagen-binding stress protein HSP47 gene and the alpha 1(I) and alpha 1(III) collagen genes in carbon tetrachloride-induced rat liver fibrosis*. J Clin Invest, 1994. **94**(6): p. 2481-8.
210. Sunamoto, M., et al., *Expression of heat shock protein 47 is increased in remnant kidney and correlates with disease progression*. Int J Exp Pathol, 1998. **79**(3): p. 133-40.
211. Becerril, C., et al., *Acidic fibroblast growth factor induces an antifibrogenic phenotype in human lung fibroblasts*. Am J Respir Cell Mol Biol, 1999. **20**(5): p. 1020-7.
212. Ishii, H., et al., *Increased expression of collagen-binding heat shock protein 47 in murine bleomycin-induced pneumopathy*. Am J Physiol Lung Cell Mol Physiol, 2003. **285**(4): p. L957-63.
213. Kakugawa, T., et al., *Pirfenidone attenuates expression of HSP47 in murine bleomycin-induced pulmonary fibrosis*. Eur Respir J, 2004. **24**(1): p. 57-65.
214. Nishino, T., et al., *Antisense oligonucleotides against collagen-binding stress protein HSP47 suppress peritoneal fibrosis in rats*. Kidney Int, 2003. **64**(3): p. 887-96.

215. Iwashita, T., et al., *Involvement of collagen-binding heat shock protein 47 and procollagen type I synthesis in idiopathic pulmonary fibrosis: contribution of type II pneumocytes to fibrosis*. Hum Pathol, 2000. **31**(12): p. 1498-505.
216. Shiohita, K., et al., *Expression of heat shock proteins 47 and 70 in the peritoneum of patients on continuous ambulatory peritoneal dialysis*. Kidney Int, 2000. **57**(2): p. 619-31.
217. Abe, K., et al., *Interstitial expression of heat shock protein 47 and alpha-smooth muscle actin in renal allograft failure*. Nephrol Dial Transplant, 2000. **15**(4): p. 529-35.
218. Kim, J.H., et al., *Noninvasive measurement of the pH of the endoplasmic reticulum at rest and during calcium release*. Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2997-3002.
219. Kim, J.H., et al., *Dynamic measurement of the pH of the Golgi complex in living cells using retrograde transport of the verotoxin receptor*. J Cell Biol, 1996. **134**(6): p. 1387-99.
220. Gooptu, B., et al., *Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease*. Proc Natl Acad Sci U S A, 2000. **97**(1): p. 67-72.
221. Nakai, A., et al., *Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum*. J Cell Biol, 1992. **117**(4): p. 903-14.
222. Natsume, T., et al., *Interactions between collagen-binding stress protein HSP47 and collagen. Analysis of kinetic parameters by surface plasmon resonance biosensor*. J Biol Chem, 1994. **269**(49): p. 31224-8.
223. Wiman, B., et al., *Plasminogen activator inhibitor 1 (PAI) is bound to vitronectin in plasma*. FEBS Lett, 1988. **242**(1): p. 125-8.
224. Andreasen, P.A., et al., *Plasminogen activator inhibitors: hormonally regulated serpins*. Mol Cell Endocrinol, 1990. **68**(1): p. 1-19.
225. Lawrence, D.A., et al., *Engineering plasminogen activator inhibitor 1 mutants with increased functional stability*. Biochemistry, 1994. **33**(12): p. 3643-8.
226. Satoh, M., et al., *Intracellular interaction of collagen-specific stress protein HSP47 with newly synthesized procollagen*. J Cell Biol, 1996. **133**(2): p. 469-83.
227. Hamilton, A.M. and J.J. Heikkila, *Examination of the stress-induced expression of the collagen binding heat shock protein, hsp47, in Xenopus laevis cultured cells and embryos*. Comp Biochem Physiol A Mol Integr Physiol, 2006. **143**(1): p. 133-41.
228. Pearson, D.S., et al., *Cloning and characterization of a cDNA encoding the collagen-binding stress protein hsp47 in zebrafish*. DNA Cell Biol, 1996. **15**(3): p. 263-72.
229. Clarke, E.P., et al., *A collagen-binding protein in the endoplasmic reticulum of myoblasts exhibits relationship with serine protease inhibitors*. J Biol Chem, 1991. **266**(26): p. 17230-5.
230. Clarke, E.P. and B.D. Sanwal, *Cloning of a human collagen-binding protein, and its homology with rat gp46, chick hsp47 and mouse J6 proteins*. Biochim Biophys Acta, 1992. **1129**(2): p. 246-8.
231. Wang, S.Y. and L.J. Gudas, *A retinoic acid-inducible mRNA from F9 teratocarcinoma cells encodes a novel protease inhibitor homologue*. J Biol Chem, 1990. **265**(26): p. 15818-22.
232. Plaxco, K.W., et al., *A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules*. J Mol Biol, 1997. **270**(5): p. 763-70.
233. Clarke, J., et al., *Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway*. Structure, 1999. **7**(9): p. 1145-53.
234. Ferguson, N., et al., *Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9*. J Mol Biol, 1999. **286**(5): p. 1597-608.

235. Morozova-Roche, L.A., et al., *Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme*. J Mol Biol, 1999. **289**(4): p. 1055-73.
236. Cota, E., et al., *Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability*. J Mol Biol, 2000. **302**(3): p. 713-25.
237. Hamill, S.J., et al., *Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac*. J Mol Biol, 2000. **295**(3): p. 641-9.
238. Luke, K., M. Perham, and P. Wittung-Stafshede, *Kinetic folding and assembly mechanisms differ for two homologous heptamers*. J Mol Biol, 2006. **363**(3): p. 729-42.
239. Geierhaas, C.D., et al., *Comparison of the transition states for folding of two Ig-like proteins from different superfamilies*. J Mol Biol, 2004. **343**(4): p. 1111-23.
240. Wirmer, J., et al., *Characterization of the unfolded state of bovine alpha-lactalbumin and comparison with unfolded states of homologous proteins*. Protein Sci, 2006. **15**(6): p. 1397-407.
241. Liu, T., P.A. Pemberton, and A.D. Robertson, *Three-state unfolding and self-association of maspin, a tumor-suppressing serpin*. J Biol Chem, 1999. **274**(42): p. 29628-32.
242. Pearce, M.C., H. Rubin, and S.P. Bottomley, *Conformational change and intermediates in the unfolding of alpha 1-antichymotrypsin*. J Biol Chem, 2000. **275**(37): p. 28513-8.
243. Yu, M.H., K.N. Lee, and J. Kim, *The Z type variation of human alpha 1-antitrypsin causes a protein folding defect*. Nat Struct Biol, 1995. **2**(5): p. 363-7.
244. Carrell, R.W. and D.A. Lomas, *Conformational disease*. Lancet, 1997. **350**(9071): p. 134-8.
245. Zhou, A., et al., *Structural mechanism for the carriage and release of thyroxine in the blood*. Proc Natl Acad Sci U S A, 2006. **103**(36): p. 13321-6.
246. Klieber, M.A., et al., *Corticosteroid-binding globulin, a structural basis for steroid transport and proteinase-triggered release*. J Biol Chem, 2007. **282**(40): p. 29594-603.
247. Pemberton, P.A., et al., *Hormone binding globulins undergo serpin conformational change in inflammation*. Nature, 1988. **336**(6196): p. 257-8.
248. Bird, P.I., et al., *Production of recombinant serpins in Escherichia coli*. Methods, 2004. **32**(2): p. 169-76.
249. Thomson, C.A. and V.S. Ananthanarayanan, *A method for expression and purification of soluble, active Hsp47, a collagen-specific molecular chaperone*. Protein Expr Purif, 2001. **23**(1): p. 8-13.
250. Evans, T.C., Jr., et al., *Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803*. J Biol Chem, 2000. **275**(13): p. 9091-4.
251. Evans, T.C. and M.-Q. Xu, *Intein-mediated protein ligation: Harnessing nature's escape artists*. Biopolymers, 2000. **51**: p. 355-362.
252. Jain, N., et al., *Interaction of procollagen I and other collagens with colligin*. Biochem J, 1994. **304** (Pt 1): p. 61-8.
253. Blond-Elguindi, S., et al., *Peptide-dependent stimulation of the ATPase activity of the molecular chaperone BiP is the result of conversion of oligomers to active monomers*. J Biol Chem, 1993. **268**(17): p. 12730-5.
254. Kokke, B.P., et al., *Caenorhabditis elegans small heat-shock proteins Hsp12.2 and Hsp12.3 form tetramers and have no chaperone-like activity*. FEBS Lett, 1998. **433**(3): p. 228-32.
255. Adinolfi, S., et al., *The factors governing the thermal stability of frataxin orthologues: how to increase a protein's stability*. Biochemistry, 2004. **43**(21): p. 6511-8.

256. Bastiras, S. and J.C. Wallace, *Equilibrium denaturation of recombinant porcine growth hormone*. Biochemistry, 1992. **31**(38): p. 9304-9.
257. Carlacci, L., K.C. Chou, and G.M. Maggiora, *A heuristic approach to predicting the tertiary structure of bovine somatotropin*. Biochemistry, 1991. **30**(18): p. 4389-98.
258. Parkinson, E.J., M.B. Morris, and S. Bastiras, *Acid denaturation of recombinant porcine growth hormone: formation and self-association of folding intermediates*. Biochemistry, 2000. **39**(40): p. 12345-54.
259. Brems, D.N., et al., *Stabilization of an associated folding intermediate of bovine growth hormone by site-directed mutagenesis*. Proc Natl Acad Sci U S A, 1988. **85**(10): p. 3367-71.
260. Brems, D.N. and H.A. Havel, *Folding of bovine growth hormone is consistent with the molten globule hypothesis*. Proteins, 1989. **5**(1): p. 93-5.
261. Brems, D.N., P.L. Brown, and G.W. Becker, *Equilibrium denaturation of human growth hormone and its cysteine-modified forms*. J Biol Chem, 1990. **265**(10): p. 5504-11.
262. Bruce, D., et al., *Thromboembolic disease due to thermolabile conformational changes of antithrombin Rouen-VI (187 Asn-->Asp)*. J Clin Invest, 1994. **94**(6): p. 2265-74.
263. Davis, R.L., et al., *Familial dementia caused by polymerization of mutant neuroserpin*. Nature, 1999. **401**(6751): p. 376-9.
264. Seo, E.J., et al., *Distribution of the native strain in human alpha 1-antitrypsin and its association with protease inhibitor function*. J Biol Chem, 2000. **275**(22): p. 16904-9.
265. Lomas, D.A., et al., *Polymerisation underlies alpha1-antitrypsin deficiency, dementia and other serpinopathies*. Front Biosci, 2004. **9**: p. 2873-91.
266. Schreuder, H.A., et al., *The intact and cleaved human antithrombin III complex as a model for serpin-proteinase interactions*. Nat Struct Biol, 1994. **1**(1): p. 48-54.
267. Wouters, M.A. and P.M. Curmi, *An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs*. Proteins, 1995. **22**(2): p. 119-31.
268. Smith, C.K. and L. Regan, *Guidelines for protein design: the energetics of beta sheet side chain interactions*. Science, 1995. **270**(5238): p. 980-2.
269. Elliott, P.R., et al., *Inhibitory conformation of the reactive loop of alpha 1-antitrypsin*. Nat Struct Biol, 1996. **3**(8): p. 676-81.

CHAPTER 11: Appendices

Appendix A – Primer sequences

Primer sequences for PCR

Cloning into pET101/D-TOPO[®]:

Danio rerio - Forward: 5'-CAC-CAT-GGA-CAA-AAA-3'

Backward: 5'-CAA-TTC-ATC-TCT-CAT-TTT-3'

Cloning into pET102/D-TOPO[®]:

Xenopus laevis - Forward: 5'-CAC-CGC-AGA-GAA-AAA-AGT-G-3'

Reverse: 5'-CAC-CAT-GGC-AGA-GAA-A-3'

Danio rerio - Forward: 5'-CAC-CGA-CAA-AAA-GTT-GAG-CAC-T-3'

Reverse: 5'-CAA-TTC-ATC-TCT-CAT-TTT-3'

Cloning into pET-32 Xa/LIC:

The production of compatible sticky stranded when cloning into Xa/LIC required the inclusion of the following sequences on the 5' end of each primer:

Forward: 5'-GGT-ATT-GAG-GGT-CGC

Reverse: 5'-AGA-GGA-GAG-TTA-GAG-CC

Xenopus laevis - Forward: 5'-GCT-ATT-GAG-GCT-CGC-TGG-ATG-ATC-AAG-CTT-CTA-3'

Reverse: 5'-AGA-GGA-GAG-TTA-GAG-CCT-AAT-TCA-TCT -CGC-ATT-TTG-TCT-CC-3'

Danio rerio - Forward: 5'-GGT-ATT-GAG-GGT-CGC-GAC-AAA-AAG-TTG-AGC-ACT-3'

Reverse: 5'-AGA-GGA-GAG-TTA-GAG-CCC-AAT-TCA-TCT-CTC-
ATT-TT-3'

Cloning into pTYB12:

Xenopus laevis - Forward: 5'-GGT-ATT-GAG-GGT-CGC-ATG-CAT-GCA-GAG-AAA-
AAA-3'

Reverse: 5'-AGA-GGA-GAG-TTA-GAG-CCT-TAT-AAT-TCA-TCT-
CGC-ATT-TTG-TCT-CC-3'

Danio rerio - Forward: 5'-GGT-ATT-GAG-GGT-CGC-ATG-CAT-GAG-GAC-AAA-
AAG-3'

Reverse: 5'-AGA-GGA-GAG-AGT-TAG-AGC-CTT-ACA-ATT-CAT-
CTC-TCA-TTT-TT-3'

Appendix B – Hsp47 sequence identity between different organisms.

Organism	Sequence
Homo	
Pan	
Bos	
Mus musculus	
Rattus	
Cricetulus	
Gallus	
Alligator	
Xenopus	
Danio	
Carassius	
Oncorhynchus	
Macaca	MWVAPGVLGDTDASTNLLTPLQECAIRSGLACAESLGVGGVS H SEG H AGPRGDSRRPAAW
Gasterosteus	
Branchiostoma	
Equus	
Ciona	
Sus	
Canis	
Fugu	
Monodelphis	

Organism	Sequence
Homo	
Pan	
Bos	
Mus musculus	
Rattus	
Cricetulus	
Gallus	
Alligator	
Xenopus	
Danio	
Carassius	
Oncorhynchus	
Macaca	AAEGGSRTQGPGGAGRGRGRGLGRFEGGLWLFLAELGRPPEAFPTFQKFLGTGRRGWGLP
Gasterosteus	
Branchiostoma	
Equus	
Ciona	
Sus	
Canis	
Fugu	
Monodelphis	

Organism	Sequence
Homo	
Pan	MRSLLLLSAFCLLEAA
Bos	
Mus musculus	
Rattus	
Cricetulus	
Gallus	
Alligator	
Xenopus	
Danio	
Carassius	
Oncorhynchus	
Macaca	YIDPRSRGAGEEQNRVVARERESRPGASPARPRPTVVHENHLPAMHSLLLLLSAFCLLAVA
Gasterosteus	
Branchiostoma	
Equus	
Ciona	
Sus	
Canis	MRLLLLINTCCLLAVV
Fugu	
Monodelphis	MSRSSPSLHPQVPSGGKQGFAPRTAMWPTQLLSALCLLAVA

Organism	Sequence
Homo	AEVKKPAAAAAPGTAEKLSPKAATLAERSAGLAFSLYQAMAKDQAVENILVSPVVVAS
Pan	LAAEVKKPAAAAAPGTAEKLSPKAATLAERSAGLAFSLYQAMAKDQAVENILVSPVVVAS
Bos	AEVKKPAAAAAPGTAEKLSPKAATLAERSAGLAFSLYQAMAKDQAVENILLSPVVVAS
Mus musculus	AEVKKPLEAAAPGTAEKLSSKATTLAERSTGLAFSLYQAMAKDQAVENILLSPLVVAS
Rattus	AEVKKPVEAAAPGTAEKLSSKATTLAERSTGLAFSLYQAMAKDQAVENILLSPLVVAS
Cricetulus	AEVKKPVEAAAPGTAEKLSSKATTLAERSTGLAFSLYQAMAKDQAVENILLSPLVVAS
Gallus	KLSDKATTLADRSTTLAFNLYHAMA KDKNMENILLSPVVVAS
Alligator	KKLSDKANALADRSATLAFNLYHAMA KDKNMENILVSPVVVAS
Xenopus	QKMSQHANVLADKSAGLAFNLYQIMAKDKKVENILLSPVVVAS
Danio	GEDKKLSTHATSMADTSANLAFNLYHNVAKEKGLENILISPVVVAS
Carassius	LSSHASILADNSANFAFNLYHNLAKEKDIENTIVISPVVVAS
Oncorhynchus	KKLSSHATTMADKSANLAFSLYHTVAKEKGLDNILISPVVVAS
Macaca	LTAEVKKPAAAAAPGTAEKLSPKAATLAERSAGLAFSLYQAMAKDQAVENILVSPVVVAS
Gasterosteus	KKLSSYATALADHSANLAFSLYHNMAKDKDTENILLSPVVVAS
Branchiostoma	SQESTPLADINSEFALELYKTLHKDHP-ENIFFSPFSIST
Equus	PRAAFAMEQLSTANTHFAVDLFRALNESDPTGNIFISPLSISS
Ciona	
Sus	LSEANGTFALRLLKILCQDDPSHNVFYSPVSISS
Canis	LAAEVKKPAAAAAPGSAEKLSPKAATLAERSAGLAFSLYQAMAKDQAVENILLSPVVVAS
Fugu	ANIVVLSELLALLASAEDKKLSNHATTADNSANLAFSLYHNMAKDKNVENILISPVVLAS
Monodelphis	GAAEVKKPATAGKSG-EAKLSEKASTLAERSAGLAFNLYRVMAQDKGVENILLSPVVVAS

Organism	Sequence
Homo	SLGLVSLGGKATTASQAKAVLSAEQLRDEEVHAGLGELLRSLSNSTARNVTWKLGSRLYG
Pan	SLGLVSLGGKATTASQAKAVLSAEQLRDEEVHAGLGELLRSLSNSTARNVTWKLGSRLYG
Bos	SLGLVSLGGKAATASQAKAVLSAEQLRDDEVHAGLGELLRSLSNSTARNVTWKLGSRLYG
Mus musculus	SLGLVSLGGKATTASQAKAVLSAEKLRDEEVHTGLGELLRSLSNSTARNVTWKLGSRLYG
Rattus	SLGLVSLGGKATTASQAKAVLSAEKLRDEEVHTGLGELVRSLSNSTARNVTWKLGSRLYG
Cricetulus	SLGLVSLGGKATTASQAKAVLSAEKLRDEEVHKGLGELLRSLSNSTARNVTWKLGSRLYG
Gallus	SLGLVSLGGKATTASQAKAVLSADKLNDDYVHSGLSELLNEVSNSTARNVTWKIGNRLYG
Alligator	SLGLVSLGGKAATASQAKAVLSADKLNDDYVHSGLSELLNEVSNSTARNVTWKIGNRLYG
Xenopus	SLGLVSMGGQASTAAQAKTVLNAEKLSDHEIHSGLAELLNEVSNSTARNVTWKMGNRLYG
Danio	SLGMVAMGSKSSTASQVKSILKADALKDEHLHTGLSELLTEVSDPQTRNVTWKISNRLYG
Carassius	SLGLVALGGKSNTASQVKTVLSATTVKDEQLHSGLSELLTEVSNSTARNVTWKISNRLYG
Oncorhynchus	SLGMVALGGKASTASQVKSIVLSADALNDEHLHTGLSELLTEVSDPKTRNVTWKISNRLYG
Macaca	SLGLVSLGGKATTASQAKAVLSAEQLRDEEVHTGLGELLRSLSNSTARNVTWKLGSRLYG
Gasterosteus	SLGMVALGGKASTASQVKTVLSADKCLKDEHLHAGLSELLSEVSDAKSRNTTWKINNRLYG
Branchiostoma	CLAMAYLGARNDTAQQMSRVLRFHKMDASDLHVLFDLLTQLHHSD-RPYTLKTANRLFG
Equus	ALAMIFLGTRGNTAAQVSKALYFDTV--EDIHSRFQSLNADINKPGAPYIL-KLANRLYG
Ciona	KGNTAKQIDDAFMFSKIEDGRFHSAFGELHGLLFDKATEKVTAKSSNRVFA
Sus	ALAMVLLGAKGDTAAQLAQVLSLNTEKD--IHQDFQALLAELNKPSTRYLL-RTANKLFG
Canis	SLGLVSLGGKATTASQAKAVLSAEQLRDEEVHAGLGELLRSLSNSTARNVTWKLGSRLYG
Fugu	SLGMVALGGKASTASQVKTVLSADKCLKDEHLHAGLSELLTELSADAKRNTTWKINNRLYG
Monodelphis	SLGLVSLGGKATTASQAKAVLSADKLRDDDVHTGLAELLASVSNNTARNVTWKLGSRLYG

: ** * : . * : * : . :

Organism	Sequence
Homo	PSSVSFADDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWAAQT TDGKLPEVTKD--VE
Pan	PSSVSFADDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWAAQT TDGKLPEVTKD--VE
Bos	PSSVSFAEDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWAAQT TDGKLPEVTKD--VE
Mus musculus	PSSVSFADDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWASQT TDGKLPEVTKD--VE
Rattus	PSSVSFADDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWASQT TDGKLPEVTKD--VE
Cricetulus	PSSVNFVEDFVHSSKQHYNCEH SKINFR-DKRSALQS INEWASQT TDGKLPEVTKD--VE
Gallus	PASINFADDFVKNSKKHYNIEH SKINFR-DKRSALKS INEWAAQT TDGKLPEVTKD--VE
Alligator	PSSISFADDFVKNSKKHYNIEH SKINFR-DKRSALKS INEWASQT TNGKLPEVTKD--VE
Xenopus	PSSISFSDNFVKDSKKHYNIEH SKINFR-DKRSTLRS INEWAAQT TDGKLPEVTS--VE
Danio	PSSVSFAEDFVKNSKKHYNIEH SKINFR-DKRSAINS INEWAAKT TDGKLPEITKD--VK
Carassius	PSSVSFVDNFLKSSKKHYNCEH SKINFR-DKRSAVKA INDWASKSTDGKLPEVTKD--VE
Oncorhynchus	PSSVTFADNFVKSSKKHYNIDH SKINLR-DKRSAVNS INEWAAKSTDGKLPEITKD--VQ
Macaca	PSSVSFADDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWAAQT TDGKLPEVTKD--VE
Gasterosteus	PSSVSFADDFVKNSKKHYNIDH SKINFR-DKRSAVNS INEWAAKSTDGKLTEITKD--VQ
Branchiostoma	QNSFEFSQKFLDETSRH YRAQLAPVDFSGNTEGARQT INSWVEEQTENKI QDLLAPGTVT
Equus	EKTYNFLADFLASTQKMYGAELASVDFQQAPEDARKE INEWVKGQTEGKI PELLVKGMVD
Ciona	DKHITVLEDY-QDSL SVYSATVESVDFK-MPKSAVKKINDWASDATNGVIKSMLEEDGVN
Sus	EKSREFLSTFKESCLRFYDAELEQLSFASAAEASRKQ INAWVSKKTEGKI PEVLPWNSID
Canis	PSSVSFAEDFVRSSKQHYNCEH SKINFR-DKRSALQS INEWAAQT TDGKLPEVTKD--VE
Fugu	PSSVSFSDDFVKSSKKHYKYDH SKINFR-DKRSAVNS INEWAAKATDGLPEITKD--VQ
Monodelphis	PSSVSFAEDFVQSSKKHYNIEH SKINFR-DKRSALQS INEWASQT TDGKLPEVTKD--VE

. : . * :: . : . ** *. *: . : . :

Organism	Sequence
Homo	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVMMMHRTGLYNYYDDEKEKL
Pan	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVMMMHRTGLYNYYDDEKEKL
Bos	RTDGALLVNAMFFKPHWDERFHHKMVDNRGFMVTRSYTVGVTTMMHRTGLYNYYDDEKEKL
Mus musculus	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTTMMHRTGLYNYYDDEKEKL
Rattus	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTTMMHRTGLYNYYDDEKEKL
Cricetulus	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTTMMHRTGLYNYYDDEKEKL
Gallus	KTDGALIVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVPMMHRTGLYNYYDDEAEKL
Alligator	KTDGALIVNAMFFKPHWDEKFHHTMVDNRGFMVTRSYTVGVPMMHRTGLYNYFDDETEKL
Xenopus	KTDGALIVNAMFFKPHWDERFHHQMVDNRGFMVTRSYTVSVPMMHRTGLYKYTDDETNNL
Danio	NTDGAMIVNAMFFKPHWDEKSHHKMVDNRGFLVTRSHTVSVPMMHRTGIYGFYEDTENRF
Carassius	KTDGAMIINAMFYKPHWDEQFHHKMVDNRGFLVHRSYTVSVPMMHRTGIYGLFDDTTNNL
Oncorhynchus	NADGATIANAMFFKPHWDEKFHEKMVDNRGFLVTRSYTVSVPMMHRTGLYKFHDDTENKV
Macaca	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVMMMHRTGLYNYYDDEKEKL
Gasterosteus	NTDGAMIVNAMFYKPHWFETFNDQMVDNRGFLVTRSYTVGVPMMHRTGLYDFHEDKENRL
Branchiostoma	PATMLVLVNAIYFKGSWERKFEESTRRLGTFHISRDEKVEVPMMHQQGRFKLAYDEDLNC
Equus	NMTKLVLVNAIYFKGNWQEKFMKEATRDAPFRLNKKDTKTVMKMYQKKKFPYNYIEDLKC
Ciona	NDTALLIINALYFRGNWDYEFDEGRTRRPFYVSKDKAVETSFMFQNEHFKYAYISELTL
Sus	EQTRLVLVNAVYFKGRWDQQFDKKYTREMPFRVNQKEQRPVQMMFQEATFRLGRVEEVPA
Canis	RTDGALLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTTMMHRTGLYNYYDDEKEKL
Fugu	NADGAMIVNAMFFKPHWDERFHDKMVDTRGFLVTRSHTIGISMHRTGLYDFYDDEVNRI
Monodelphis	KTDGALIVNAMFFKPHWDERFHHKMVDNRGFMVTRSYTVGVPMMHRTGLYNYYDDETEKL

: **::: * . . * : :. :*. : :

Organism	Sequence
Homo	QIVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKIWMGK--MQKKAV AISLP
Pan	QIVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKIWMGK--MQKKAV AISLP
Bos	QMVEMPLA H KLSSLIIIMP HH VEP----LERLEKLLTKEQLKVWMGK--MQKKAV AISLP
Mus musculus	QMVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKAWMGK--MQKKAV AISLP
Rattus	QLVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKTWMGK--MQKKAV AISLP
Cricetulus	QILEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKAWMGK--MQKKAV AISLP
Gallus	QVVEMPLA H KLSSMIFIMP NH VEP----LERVEKLLNREQLKTWASK--MKKRSV AISLP
Alligator	QIVEMPLA H KLSSMIFIMP NH VEP----LERVEKLLTREQLKTWISK--LKKRSV AISLP
Xenopus	QILEMPLA H KLSSMIIIMPY H VEP----LERLEKLLTREQVNAWDGK--MKKRAV AVSLP
Danio	LIVSMPLA H KKSSMIFIMPY H VEP----LDRLENLLTRQQLDTWISK--LEERAV AISLP
Carassius	LVLDMALA H KMSSIVFIMPY H VES----LERVEKLLTRQQNLTWISK--MEQRAV AVSLP
Oncorhynchus	FVLDMPLGQKQSSLVFIMPY H LEP----LDRLEKLLTRKQLETWMGK--MEERAV AISLP
Macaca	QIVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKIWMGK--MQKKAV AISLP
Gasterosteus	FVLNMPLGKKEASMIILIMPY H LEP----LARLEKLLTRKQVDTWISK--MENIAV AISLP
Branchiostoma	QILEMPYQGK H LSMLLVLP EKMD----ALSTIETSLTPDILRRWQKS--MDEVSTMVQIP
Equus	RVLELPYQGKELSMIILLPDDIEDESTGLEKIEKQLTLEKLREWTKPENLYLAENV H LP
Ciona	QVLEMDYAGKDYSMVLLMPENFE-----LPKVEANLN H ANLTKWLSA--LK H ESVDLTIP
Sus	QVLELPYEDREL SMVLLPDD H ----VALSEVERQLTFEKLLAWTKPERMQSLDVEVFLP
Canis	QIVEMPLA H KLSSLIIILMP HH VEP----LERLEKLLTKEQLKIWMGK--MQKKAV AISLP
Fugu	YVLNMPLGQKQASMIILIMPY H LEP----LERLEKLLSKKQVDTWISK--MTNKAV AISLP
Monodelphis	QMVEMPLA H KLSSLIIIMP HH VEP----LERLEKLLTKEQLKTWL GK--MKKRAV AISLP

: : : : : * : : : : * . * : * * . : * : . : : *

Organism	Sequence
Homo	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFELDTDGN-
Pan	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFELDTDGN-
Bos	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFEWDTDGN-
Mus musculus	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFEWDTEGN-
Rattus	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFEWDTEGN-
Cricetulus	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFEWDTDGN-
Gallus	K--VVLEVS H DLQ K HLADLGLTEAIDKTKADLSKISGKKDLYLSNVF H AAALEWDTDGN-
Alligator	K--VSLEVS H DLQ K HLADLGLTEAIDKNKADLSKISGKKDLYLSNVF H AAALEWDTEGN-
Xenopus	K--VSLEVS H DLQ K HLGDLGLTEAIDKSKADLSKISGKKDLYLASMF H AAALEWDTEGN-
Danio	K--VSMEVS H DLQ K HLGELGLTEAVDKPKADLSNISGKKDLYLSNVF H ASSLEWDTEGN-
Carassius	K--VSVEVS H DLQ K HLTELGLTEAVDKAKADLSNISGKKDLYLSNVF H SAMEWDTEGN-
Oncorhynchus	K--VSMEVS H NLQ K TLGELGLTDAVDKTKADLSNISGKKDLYLSNVF H ASSMEWDIEGN-
Macaca	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFELDTDGN-
Gasterosteus	K--IALEVS H NLQ K HLAELGLTEAVDKSKADLSNISGKKDLYLSNVF H ASALELDVKGN-
Branchiostoma	K--FKLV H DFVLNEKLADMGMTDLFSMADADLSGITGSRDL H VSQVI H KAFVEVNEEGSE
Equus	R--FKLEESYDLTS H LARLGVDLFDNRGKADLSGMSGARDLFDVSKI I H KSFVDLNEEGTE
Ciona	K--FKLEETLQLQEVLPMGVADLFDQRQACDLNGIANRNDLFVDQIV H KTVLDVNEQGSE
Sus	R--FKLDASYDLELLLG H LGVVDAFQQGKADFSAMAPERDLSLSTFV H KSVVEVNEEGSE
Canis	K--GVVEVT H DLQ K HLAGLGLTEAIDKNKADLSRMSGKKDLYLASVF H ATAFEWDTEGN-
Fugu	K--ISVDVS H NIQKYLSELGLTEAVDKAKADLSNISGKKDLYLSNVF H ASAVELDVDGN-
Monodelphis	K--GVVEVT H DLQ K HLAGLGLTEAMDKNKADLSRMSGKKDLYLASVF H ATAFEWDTEGN-

: : : * :* : : .. .* : : .** : ..* : . : : .*.

Organism	Sequence
Homo	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDTQSGSLLFIGRLVRPKGDKMRDEL
Pan	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDTQSGSLLFIGRLVRPKGDKMRDEL
Bos	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDTQSGSLLFIGRLVRPKGDKMRDEL
Mus musculus	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDNQSGSLLFIGRLVRPKGDKMRDEL
Rattus	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDNQSGSLLFIGRLVRPKGDKMRDEL
Cricetulus	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDNQSGSLLFIGRLVRPKGDKMRDEL
Gallus	--PYDADIYGREEMRN-PKLFYADHPFIFMIKDSKTNSILFIGRLVRPKGDKMRDEL
Alligator	--PFDADIYGREEIRN-PRLFYADHPFIFLIKDNKTNSILFIGRLVRPKGDKMRDEL
Xenopus	--PFDQDIYGREELRS-PKLFYADHPFIFLVQDAQSGSLLFIGRLVRPKGDKMRDEL
Danio	--PFDPSIFGSEKMRN-PKLFYADHPFIFLVKDNKTNSILFIGRLVRPKGDKMRDEL
Carassius	--PPDTSIYGTDKTKT-PKLFYADHPFIFLVKDKTNSILFMGRVQPKGDKMRDEL
Oncorhynchus	--PFDTSIFGSEKLRN-PKLFYADHPFIFLVKDNKTNSILFIGRMVRPKGDKMRDEL
Macaca	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDTQSGSLLFIGRLVRPKGDKMRDEL
Gasterosteus	--PFDTSVYGSGKLTN-PRLFYADHPFVFLVKDNKTNSILYIGRVVKPKGDKMRDEL
Branchiostoma	AAAATAVNMMKRSLD--GETFFADHPFLFLIRDNDNSNSILFLGRLVRPEGLTTKDEL
Equus	AAAATAGTIMLAMLMPEENFNADHPFIFFFIRHNPSANILFLGRFSSP-----
Ciona	AAATTSVRSQCDSVAFNPISFVADHPFLWAIRHRQSELLVFMGRFSRPEGPLLGRDE
Sus	AAAASALVLMECCMES-GPRFCADHPFLFFIRHNKAKSILFCGRFSSP-----
Canis	--PFDQDIYGREELRS-PKLFYADHPFIFLVRDTQSGSLLFIGRLVRPKGDKMRDEL
Fugu	--PYDTSIFGTEKLN-PKLFYVDHPFIFLVKDNKTNSIMYIGRVVKPKGDKMRDEL
Monodelphis	--PFDQDIYGREELRS-PKLFYADHPFIFLVQDAQSGSLLFIGRLVRPKGDKMRDEL

: * .*****:: ::. : ::: **. *

Appendix C

WT *Mus musculus* Hsp47

atgcgctctc	tccttctg	caccttatgc	ctcttggccg	tggccctggc	agccgaggtg	60
*****	*****	*****	*****	*****	*****	
aagaaacccc	tagaggcggc	agcccctggt	actgcggaga	agctaagttc	caaggcgacc	120
*****	*****	*****	*****	*****	*****	
acactggcag	agcgcagcac	aggcctggcc	ttcagcctat	atcaggcgat	ggccaaagac	180
*****	*****	*****	*****	*****	*****	
caggcggtgg	agaacatcct	cctgtcaccc	ttggtggtgg	cctcatccct	gggtcttgtg	240
*****	*****	*****	*****	*****	*****	
tactgggtg	gtaaagccac	cacagcgtcg	caggcgaagg	cagtgctgag	cgctgagaag	300
*****	*****	*****	*****	*****	*****	
ctgcgcgatg	aggaggtgca	cacggggctg	ggtgagctgc	tccgctccct	cagcaactcc	360
*****	*****	*****	*****	*****	*****	
actgcgcgca	acgtgacctg	gaaactgggc	agccgcctgt	acggggccag	ctccgtgagc	420
*****	*****	*****	*****	*****	*****	
ttcgccgatg	acttcgtgcg	cagcagcaag	caacactaca	actgcgaaca	ctccaagatc	480
*****	*****	*****	*****	*****	*****	
aacttccgag	acaagcgcag	cgccctgcag	tccatcaacg	agtgggcctc	gcagaccacg	540
*****	*****	*****	*****	*****	*****	
gacggcaagc	tgcttgaggt	caccaaggat	gtggagcgca	cgatggggc	actgcttgtg	600
*****	*****	*****	*****	*****	*****	
aacgccatgt	tctttaagcc	acactgggat	gagaagtttc	accacaagat	ggtggacaac	660
*****	*****	*****	*****	*****	*****	
cgtggcttca	tggtgacctg	ctcctatact	gtgggtgtta	cgatgatgca	ccggacaggc	720
*****	*****	*****	*****	*****	*****	
ctgtacaact	actatgacga	cgagaaggag	aagctgcaga	tggtggagat	gcccctggct	780
*****	*****	*****	*****	*****	*****	
cacaagctct	ccagcctcat	catcctcatg	ccccaccatg	tggagccgct	cgagcgcttg	840
*****	*****	*****	*****	*****	*****	
gagaagctgc	tgaccaagga	gcagctgaag	gcctggatgg	gaaagatgca	gaagaaggct	900
*****	*****	*****	*****	*****	*****	
gtcgccatct	ccctgccccaa	gggcgtggtg	gaggtgaccc	atgacctgca	gaaacatctg	960
*****	*****	*****	*****	*****	*****	
gcaggactgg	gcctgaccga	agccatcgac	aagaacaagg	cagacctatc	gcgcatgtct	1020
*****	*****	*****	*****	*****	*****	
ggcaagaagg	acctgtacct	ggccagtgtg	ttccacgcca	ctgccttcga	gtgggacacc	1080
*****	*****	*****	*****	*****	*****	
gagggcaacc	cctttgacca	agacatctac	gggcgcgagg	agctgcgcag	ccccaaagctg	1140
*****	*****	*****	*****	*****	*****	
ttctatgccg	accaccctt	catcttcctg	gtgcgagata	atcagagcgg	ctccttgctc	1200
*****	*****	*****	*****	*****	*****	
ttcattggcc	gcctgggtccg	gccaaggga	gacaagatgc	gagatgagtt	gtag	1254
*****	*****	*****	*****	*****	*****	

Key: *: Primer F2_C01 *: Primer F3_D01 *: Primer F4_D03 *: Primer F1_B01

*: Primer R1_F03

H191A *Mus musculus* Hsp47

atgcgctctc	tccttctg	caccttatgc	ctcttgccg	tggccctggc	agccgaggtg	60
*****	*****	*****	*****	*****	*****	
aagaaacccc	tagaggcggc	agcccctggt	actgcggaga	agctaagttc	caaggcgacc	120
*****	*****	*****	*****	*****	*****	
acactggcag	agcgcagcac	aggcctggcc	ttcagcctat	atcaggcgat	ggccaaagac	180
*****	*****	*****	*****	*****	*****	
caggcggtgg	agaacatcct	cctgtcaccc	ttggtggtgg	cctcatccct	gggtcttgtg	240
*****	*****	*****	*****	*****	*****	
tcactgggtg	gtaaagccac	cacagcgtcg	caggcgaagg	cagtgtctgag	cgctgagaag	300
*****	*****	*****	*****	*****	*****	
ctgcgcgatg	aggaggtgca	cacggggctg	ggtgagctgc	tccgctccct	cagcaactcc	360
*****	*****	*****	*****	*****	*****	
actgcgcgca	acgtgacctg	gaaactgggc	agccgcctgt	acggggccag	ctccgtgagc	420
*****	*****	*****	*****	*****	*****	
ttcgccgatg	acttcgtgcg	cagcagcaag	caacactaca	actgcgaaca	ctccaagatc	480
*****	*****	*****	*****	*****	*****	
aacttccgag	acaagcgcag	cgccctgcag	tccatcaacg	agtgggcctc	gcagaccacg	540
*****	*****	*****	*****	*****	*****	
gacggcaagc	tgcctgaggt	caccaaggat	gtggagcgca	cggatggggc	actgcttgtg	600
*****	*****	*****	*****	*****	*****	
aacgccatgt	tctttaagcc	acactgggat	gagaagtttc	accacaagat	ggtggacaac	660
*****	*****	*gca*****	*****	*****	*****	
cgtaggcttc	tggtgacctg	ctcctatact	gtgggtgtta	cgatgatgca	cgggacaggc	720
*****	*****	*****	*****	*****	*****	
ctgtacaact	actatgacga	cgagaaggag	aagctgcaga	tggtggagat	gcccctggct	780
*****	*****	*****	*****	*****	*****	
cacaagctct	ccagcctcat	catcctcatg	ccccaccatg	tggagccgct	cgagcgcttg	840
*****	*****	*****	*****	*****	*****	
gagaagctgc	tgaccaagga	gcagctgaag	gcctggatgg	gaaagatgca	gaagaaggct	900
*****	*****	*****	*****	*****	*****	
gtcgccatct	ccctgcccac	ggcggtgggt	gaggtgacct	atgacctgca	gaaacatctg	960
*****	*****	*****	*****	*****	*****	
gcaggactgg	gcctgaccga	agccatcgac	aagaacaagg	cagacctatc	gcgcatgtct	1020
*****	*****	*****	*****	*****	*****	
ggcaagaagg	acctgtacct	ggccagtgtg	ttccacgcca	ctgccttcga	gtgggacacc	1080
*****	*****	*****	*****	*****	*****	
gagggcaacc	cctttgacca	agacatctac	gggcgcgagg	agctgcgcag	ccccaagctg	1140
*****	*****	*****	*****	*****	*****	
ttctatgccg	accacccctt	catcttcctg	gtgcgagata	atcagagcgg	ctccttgctc	1200
*****	*****	*****	*****	*****	*****	
ttcattggcc	gcctggtccg	gcccaaggga	gacaagatgc	gagatgagtt	gtag	1254
*****	*****	*****	*****	*****	*****	

Key: *: Primer F2_G03 *: Primer F3_H03 *: Primer F4_D05

*: Primer R2_A04 *: Primer R1_E05

H197A:H198A *Mus musculus* Hsp47

atgcgctctc	tccttctggg	caccttatgc	ctcttggccg	tggccctggc	agccgaggtg	60
*****	*****	*****	*****	*****	*****	
aagaaacccc	tagaggcggc	agcccctggt	actgcggaga	agctaagttc	caaggcgacc	120
*****	*****	*****	*****	*****	*****	
acactggcag	agcgcagcac	aggcctggcc	ttcagcctat	atcaggcgat	ggccaaagac	180
* ** *	*****	*****	*****	*****	*****	
caggcggtgg	agaacatcct	cctgtcaccc	ttggtggtgg	cctcatccct	gggtcttgtg	240
* ** *	*****	*****	*****	*****	*****	
tcactgggtg	gtaaagccac	cacagcgtcg	caggcgaagg	cagtgtctgag	cgctgagaag	300
* ** *	*****	*****	*****	*****	*****	
ctgcgcgatg	aggaggtgca	cacggggctg	ggtgagctgc	tccgctccct	cagcaactcc	360
*****	*****	*****	*****	*****	*****	
actgcgcgca	acgtgacctg	gaaactgggc	agccgcctgt	acgggcccag	ctccgtgagc	420
* ** *	*****	*****	*****	*****	*****	
ttcgccgatg	acttcgtgcg	cagcagcaag	caacactaca	actgcgaaca	ctccaagatc	480
*****	*****	*****	*****	*****	*****	
aacttccgag	acaagcgcag	cgccctgcag	tccatcaacg	agtgggcctc	gcagaccacg	540
*****	*****	*****	*****	*****	*****	
gacggcaagc	tgcttgaggt	caccaaggat	gtggagcgca	cggatggggc	actgcttgtg	600
*****	*****	*****	*****	*****	*****	
aacgccatgt	tctttaagcc	acactgggat	gagaagtttc	accacaagat	ggtggacaac	660
*****	*****	*****	*****g	cagca*****	*****	
cgtaggcttca	tggtgacccg	ctcctatact	gtgggtgtta	cgatgatgca	ccggacaggc	720
*****	*****	*****	*****	*****	*****	
ctgtacaact	actatgacga	cgagaaggag	aagctgcaga	tgggtggagat	gcccctggct	780
*****	*****	*****	*****	*****	*****	
cacaagctct	ccagcctcat	cacccctcatg	ccccaccatg	tggagecget	cgagecgettg	840
*****	*****	*****	*****	*****	*****	
gagaagctgc	tgaccaagga	gcagctgaag	gcctggatgg	gaaagatgca	gaagaaggct	900
*****	*****	*****	*****	*****	*****	
gtcgccatct	ccctgccccaa	ggcgtggtg	gaggtgaccc	atgacctgca	gaaacatctg	960
*****	*****	*****	*****	*****	*****	
gcaggactgg	gcctgaccga	agccatcgac	aagaacaagg	cagacctatc	gcgcatgtct	1020
*****	*****	*****	*****	*****	*****	
ggcaagaagg	acctgtacct	ggccagtgtg	ttccacgcca	ctgccttcga	gtgggacacc	1080
*****	*****	*****	*****	*****	*****	
gagggcaacc	cctttgacca	agacatctac	gggcgcgagg	agctgcgcag	ccccaaagctg	1140
*****	*****	*****	*****	*****	*****	
ttctatgccg	accaccctt	catcttcctg	gtgcgagata	atcagagcgg	ctccttgctc	1200
*****	*****	*****	*****	*****	*****	
ttcattggcc	gcctggtccg	gccaaggga	gacaagatgc	gagatgagtt	gtag	1254
*****	*****	*****	*****	*****	*****	

Key: *: Primer F2_F09 *: Primer F3_G11 *: Primer F1_H11

*: Primer R1_B12 *: Primer R2_C12

H220A *Mus musculus* Hsp47

atgcgctctc	tcctttctggg	caccttatgc	ctcttggccg	tggccctggc	agccgaggtg	60
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
aagaaacccc	tagaggcggc	agcccctggt	actgcggaga	agctaagttc	caaggcgacc	120
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
acactggcag	agcgcagcac	aggcctggcc	ttcagcctat	atcaggcgat	ggccaaagac	180
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
caggcgggtg	agaacatcct	cctgtcacc	ttggtggtgg	cctcatccct	gggtcttgtg	240
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
tactgggtg	gtaaagccac	cacagcgtcg	caggcgaagg	cagtgtctgag	cgctgagaag	300
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ctgcgcgatg	aggaggtgca	cacggggctg	ggtgagctgc	tccgctccct	cagcaactcc	360
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
actgcgcgca	acgtgacctg	gaaactgggc	agccgcctgt	acgggcccg	ctccgtgagc	420
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ttcgccgatg	acttcgtgcg	cagcagcaag	caacactaca	actgcgaaca	ctccaagatc	480
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
aacttccgag	acaagcgcag	cgccctgcag	tccatcaacg	agtgggcctc	gcagaccacg	540
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
gacggcaagc	tgcctgaggt	caccaaggat	gtggagcgca	cggatggggc	actgcttgtg	600
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
aacgccatgt	tctttaagcc	acactgggat	gagaagtttc	accacaagat	ggtggacaac	660
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
cgtggcttca	tggtgacctg	ctcctatact	gtgggtgtta	cgatgatgca	cgggacaggc	720
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ctgtacaact	actatgacga	cgagaaggag	aagctgcaga	tgggtggagat	gccccctggct	780
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
cacaagctct	ccagcctcat	catcctcatg	ccccaccatg	tggagccgct	cgagcgcttg	840
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
gagaagctgc	tgaccaagga	gcagctgaag	gcctggatgg	gaaagatgca	gaagaaggct	900
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
gtcgccatct	ccctgcccac	gggcgtgggtg	gaggtgacct	atgacctgca	gaaacatctg	960
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
gcaggactgg	gcctgaccga	agccatcgac	aagaacaagg	cagacctatc	gcgcatgtct	1020
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ggcaagaagg	acctgtacct	ggccagtgtg	ttccacgcca	ctgccttcga	gtgggacacc	1080
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
gagggcaacc	cctttgacca	agacatctac	gggcgcgagg	agctgcgcag	ccccaagctg	1140
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ttctatgccg	accacccctt	catcttcctg	gtgcgagata	atcagagcgg	ctccttgctc	1200
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	
ttcattggcc	gcctggtccg	gccaaggga	gacaagatgc	gagatgagtt	gtag	1254
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	

Key: * : Primer F2_B03 * : Primer F3_C03 * : Primer F4_H05 * : Primer F5_C05
* : Primer R0_B05 * : Primer R1_A05 * : Primer R2_D03

H255A:H256A *Mus musculus* Hsp47

atgcgctctc	tccttctg	caccttatgc	ctcttgccg	tggccctggc	agccgaggtg	60
*****	*****	*****	*****	*****	*****	
aagaaacccc	tagaggcggc	agcccctggt	actgcggaga	agctaagtcc	caaggcgacc	120
*****	*****	*****	*****	*****	*****	
acactggcag	agcgcagcac	aggcctggcc	ttcagcctat	atcaggcgat	ggccaaagac	180
*****	*****	*****	*****	*****	*****	
caggcggtgg	agaacatcct	cctgtcacc	ttggtggtgg	cctcatccct	gggtcttgtg	240
*****	*****	*****	*****	*****	*****	
tcactgggtg	gtaaagccac	cacagcgctg	caggcgaagg	cagtgtctgag	cgctgagaag	300
*****	*****	*****	*****	*****	*****	
ctgcgcgatg	aggaggtgca	cacggggctg	ggtgagctgc	tccgctccct	cagcaactcc	360
*****	*****	*****	*****	*****	*****	
actgcgcgca	acgtgacctg	gaaactgggc	agccgcctgt	acggggccag	ctccgtgagc	420
*****	*****	*****	*****	*****	*****	
ttcgccgatg	acttcgtgcg	cagcagcaag	caacactaca	actgcgaaca	ctccaagatc	480
*****	*****	*****	*****	*****	*****	
aacttccgag	acaagcgcag	cgccctgcag	tccatcaacg	agtgggcctc	gcagaccacg	540
*****	*****	*****	*****	*****	*****	
gacggcaagc	tgcctgaggt	caccaaggat	gtggagcgca	cggatggggc	actgcttgtg	600
*****	*****	*****	*****	*****	*****	
aacgccatgt	tctttaagcc	acactgggat	gagaagtttc	accacaagat	ggtggacaac	660
*****	*****	*****	*****	*****	*****	
cgtaggctta	tggtgacctg	ctcctatact	gtgggtgtta	cgatgatgca	cgggacaggc	720
*****	*****	*****	*****	*****	*****	
ctgtacaact	actatgacga	cgagaaggag	aagctgcaga	tggtggagat	gcccctggct	780
*****	*****	*****	*****	*****	*****	
cacaagctct	ccagcctcat	catcctcatg	ccc caccat	tggagccgct	cgagcgcttg	840
*****	*****	*****	*** cgacga	*****	*****	
gagaagctgc	tgaccaagga	gcagctgaag	gcctggatgg	gaaagatgca	gaagaaggct	900
*****	*****	*****	*****	*****	*****	
gtcgccatct	ccctgcccaa	ggcggtggtg	gaggtgacct	atgacctgca	gaaacatctg	960
*****	*****	*****	*****	*****	*****	
gcaggactgg	gcctgaccga	agccatcgac	aagaacaagg	cagacctatc	gcgcatgtct	1020
*****	*****	*****	*****	*****	*****	
ggcaagaagg	acctgtacct	ggccagtgtg	ttccacgcca	ctgccttcga	gtgggacacc	1080
*****	*****	*****	*****	*****	*****	
gagggcaacc	cctttgacca	agacatctac	gggcgcgagg	agctgcgcag	ccccaagctg	1140
*****	*****	*****	*****	*****	*****	
ttctatgccg	accacccctt	catcttcctg	gtgcgagata	atcagagcgg	ctccttgctc	1200
*****	*****	*****	*****	*****	*****	
ttcattggcc	gcctggtccg	gccaaggga	gacaagatgc	gagatgagtt	gtag	1254
*****	*****	*****	*****	*****	*****	

Key: * : Primer F2_E02 * : Primer F3_F02 * : Primer F4_D04 * : Primer F5_G04

* : Primer R0_F04 * : Primer R1_E04 * : Primer R2_G02