MECHANISTIC BEHAVIOUR AND MOLECULAR INTERACTIONS OF HEAT SHOCK PROTEIN 47 (HSP47)

Mohd Firdaus Abdul Wahab

Department of Chemistry, Imperial College London

A thesis submitted for the award of the degree of

DOCTOR OF PHILOSOPHY

of

Imperial College London

NOVEMBER 2011

DECLARATION

I hereby declare that the work presented in this thesis was carried out by me myself unless otherwise stated and acknowledged.

Mohd Firdaus Abdul Wahab

ABSTRACT

This project involves the study of heat shock protein 47 (HSP47), which is a molecular chaperone crucial for collagen biosynthesis. It exhibits a high degree of sequence homology with members of the serine protease inhibitor (serpin) superfamily, though HSP47 does not possess the inhibitory activity. It is a single-substrate chaperone, and binds only to collagen. 'Knock-out' of the *hsp47* gene impairs the secretion of correctly folded collagen triple helix molecules leading to embryonic lethality in mice. Thus the aim of this project was to elucidate the specific mechanism that governs the binding to and release from collagen at the molecular level, known as the 'pH-switch mechanism'. Emphasis is given on histidine (His) residues as the HSP47-collagen dissociation pH is similar to the pK_a of the imidazole side chain of His residues. Site directed mutagenesis was used to mutate surface His residues, based on a mouse HSP47 homology model. The effects of the mutations on the behaviour of HSP47 were then assessed by collagen binding assays and structural analyses with circular dichroism (CD). All mutants were found to have good solubility and retain their binding ability to collagen like wild-type HSP47 in batch assay, but perturbed behaviour was seen in column experiment. Mutation of His residue at position 191 (H191) causes the shift in the collagen dissociation pH, while mutation of H197 and/or 198 disrupt the specific HSP47-collagen interaction. H191, 197 and 198 are predicted to be located in the region near the *C*-terminus of strand 3 of β-sheet A (s3A) in the homology model, a region specifically known as the 'breach cluster' in serpin nomenclature. The extent of conformational rearrangement of this region was further investigated by means of intrinsic tryptophan fluorescence spectroscopy using a series of single tryptophan (Trp) mutants. Results from analyses performed on the mutants did not contradict the observation seen in His mutational work, as Trp residues in the 'breach' cluster are likely to be located in the dynamic region of HSP47 pH-triggered conformational change. In conclusion, this study establishes the importance of His residues in the 'breach cluster' to HSP47 pH-switch behaviour. Finally, a model for HSP47 pH-switch mechanism was proposed from data obtained *via* mutagenesis experiments. The model is hoped to assist future research into HSP47 cellular behaviour and will also be of great use in therapeutic applications involving the molecular chaperone.

ACKNOWLEDGMENTS

First of all, my sincerest gratitude to my parents for their undying support throughout this journey, without which would not have been possible. Also to my beloved family who have always been my strength and motivation to keep me going during challenging times. This thesis is dedicated to them.

My gratitude first goes to Prof Andrew Miller, as a person who introduced me to HSP47 and gave me the advice needed to work with this unpredictable protein. This extends to my supervisors Prof Tony Cass and Prof Robin Leatherbarrow for their support to allow me to continue working on HSP47, through endless discussions and suggestions. Their supports made the transition seamless with very little disruption.

Then to my GTC group members, who are among the first people I knew in London: Dr Michael Wright whose assistance began even before I arrived here, Dr Takayuki Homma for hands-on guidance to HSP47 biology, Dr Nazila Kamaly, Dr Ming Wang, Dr Soumia Kolli, Dr Carla Prata, Melanie, Stefano and many more that I could not list here but you will always be on my mind. Thanks for making this experience enjoyable and less stressful with all the laughters and non-stop parties.

Not to forget my new (and equally fun) Cass group members; Dr Murugesan Muthu for assistance with my peptide work, Dr Sanjiv Sharma for processing the POs, Dr Anna Radomska, Dr Thao Le for the swift but thorough proofreading, Melissa, Sasinee and definitely the whole group members for continuous support scientifically and socially till the end. Thank you very much guys!

I should also mention the people that I met throughout my time here: Dr Timothy Dafforn and Dr Umbreen Ahmed from Birmingham University for sharing their knowledge on HSP47; Prof Kazuhiro Nagata and all Nagata lab members for being such a generous host during my visit to Japan.

A special thank you goes to my Malaysian friends that I met in London; at MARA House, at Imperial, at various events thoughout the city that connects all Malaysians here, and all over the country. I wish you all the best, and do keep in touch no matter where we are. Finally, to my sponsors the MOHE Malaysia and UTM for scholarship and study leave, both are gratefully acknowledged.

PUBLICATIONS, CONFERENCES & PRESENTATIONS

M. F. Abdul-Wahab, T. Homma, M. Wright, D. Olerenshaw, T. R. Dafforn, K. Nagata, A. D. Miller, (2011). "Mutations in the *Breach* Histidine Cluster Alters the pH Sensitivity and Molecular Interaction of Murine Heat Shock Protein 47 (HSP47) to Collagen". *Submitted for publication.*

M. F. Abdul-Wahab, T. Homma, M. Wright, A. D. Miller, T. Cass, R. J. Leatherbarrow (2010). "Locating the Important Histidine Residues of the Collagen Specific Molecular Chaperone HSP47". Presented at the Department of Chemistry Postgraduate Symposium, Imperial College London, UK. 14 July 2010.

M. F. Abdul-Wahab, T. Homma, M. Wright, A. D. Miller (2010). "Locating the Important Histidine Residues of the Collagen Specific Molecular Chaperone HSP47". Presented at the Royal Society of Chemistry (RSC) Bioorganic Meeting, University of Nottingham, UK. 22 April 2010.

T. Homma,Y. Ishikawa, **M. F. Abdul-Wahab,** Y. Masago, Y. Matsuoka, A. D. Miller, H. Kubota, K. Nagata (2009). "Analysing the Interface between Collagen and HSP47, a Collagen Specific Molecular Chaperones". Presented at the 82nd Annual Meeting of the Japanese Biochemical Society, Kobe, Japan. 21-24 October 2009.

M. F. Abdul-Wahab, T. Homma, A. D. Miller (2009). "Protein Mutants: The Importance of Histidines to HSP47 Binding to Collagen". Presented at the Department of Chemistry Postgraduate Symposium, Imperial College London, UK. 10 June 2009.

The GBSF Travel Award to Japan to attend the 32nd Annual Meeting of the Molecular Biology Society of Japan, 11-12th December 2009, Pacifico Yokohama, Yokohama, Japan.

CONTENTS

ABBREVIATIONS

$\Delta\Delta A_{222}$	change in CD signal intensity at 222 nm
ΔA	difference in absorbance between CD L and R components
Abs_{280}, A_{280}	absorbance at 280 nm
ACT	antichymotrypsin
Ala, A	alanine
Arg, R	arginine
Asp, D	aspartic acid
ATP	adenosine triphosphate
AU	arbitrary unit
bis-ANS	4,4′-bis(1-anilinonaphtalene-8-sulfonate)
Boc	*tert*-butyloxycarbonyl
BODIPY	4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-*s*-indacene-3-propionic acid
CBB	Coomassie Brilliant Blue
CD	circular dichroism
cDNA	complementary deoxyribonucleic acid
CEF	chick embryonic fibroblast
CMP	collagen mimetic peptide
CNBr	cyanogen bromide
CNX	calnexin
CRP	collagen-related peptide
CRT	calreticulin

CV	column volume
Cys, C	cysteine
DHFR	dihydrofolate reductase
ECM	extracellular matrix
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide HCl
ER	endoplasmic reticulum
ERGIC	endoplasmic reticulum-Golgi intermediate compartment
ESI	electrospray ionisation
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
HA	histidine to alanine
His, H	histidine
hPCI	human protein C inhibitor
HPLC	high performance liquid chromatography
HSP47	heat shock protein 47
Hyp, O	hydroxyproline
I_f	maximum fluorescence emission intensity
IPTG	isopropyl-β-D-thiogalactopyranoside
k_d	dissociation constant

kDa	kilo Dalton
KDEL	Lys-Asp-Glu-Leu
LDLR	low density lipoprotein receptor
Leu, L	leucine
LH	lysyl hydroxylase
LO	lysyl oxidase
Lys, K	lysine
MALDI-TOF	matrix-assisted laser desorption ionisation-time of flight
mdeg	milidegrees
MEF	mouse embryonic fibroblast
MES	2-(*N*-morpholino)ethane sulfonic acid
MRE	mean residue ellipticity
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NaPi	sodium phosphate buffer
NHS	*N*-hydroxysulfosuccinimide
OI	Osteogenesis imperfecta
P3H	prolyl 3-hydroxylase
P4H	prolyl 4-hydroxylase
PAI-1	plasminogen activator inhibitor-1
PBS	phosphate buffered saline
PCR	polymerase chain reaction

PDI	protein disulfide isomerase
PEDF	pigment epithelium derived factor
Phe, F	phenylalanine
pK_a	acid dissociation constant
POG	proline-hydroxyproline-glycine
PPG	proline-proline-glycine
PPI	polyproline type I
PPIase	peptidyl-prolyl *cis-trans* isomerase
PPII	polyproline type II
PRG	proline-arginine-glycine
Pro, P	proline
RAP	receptor-associated protein
RCL/RSL	reactive centre loop / reactive serpin loop
RDEL	Arg-Asp-Glu-Leu
RET	resonance energy transfer
RNAi	ribonucleic acid interference
rpm	revolution per minute
serpin	serine protease inhibitor
siRNA	small interfering ribonucleic acid
TBG	thyroxine-binding globulin
THP	collagen-like triple helical peptide
T_m	melting temperature

Trp, W	tryptophan
Tyr, Y	tyrosine
WF	tryptophan to phenylalanine
WT	wild-type
α_1-AT	α_1-antitrypsin
α_1-PI	α_1-protein inhibitor
β-ME	β-mercaptoethanol
λ_{em}	maximum fluorescence emission wavelength

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

CHAPTER 1: INTRODUCTION

The extracellular matrix (ECM) is a complex of proteins that defines the structural integrity and carries various physiological functions in organisms. Collagen, being the most abundant protein in ECM, has been extensively studied structurally and physiologically. This includes its biosynthesis and secretion to ECM (Bornstein, 1974; Canty and Kadler, 2005; Prockop et al., 1979). It has been discovered that a group of proteins, called molecular chaperones, assist the formation and transport of collagen triple helix in the endoplasmic reticulum (ER) and the Golgi compartment. Among the most important, but yet least understood of those chaperones is heat shock protein 47 (HSP47). Knock-out of the *hsp47* gene disrupts the formation of collagen triple helices, consequently basement membrane formation leading to embryonic death (Nagai et al., 2000). These demonstrate the importance of this chaperone for the normal development of mice. Although the general function has already been suggested, a specific mechanism that governs the binding to and release from collagen, at the molecular level, is still being debated. Hence, it is the aim of this project to shed light on this mechanism by constructing mutants targeting important residues of HSP47 to investigate the consequent effects on the behaviour of the chaperone.

1.1 Aims of the Project

HSP47 research at the cellular level has been expanding rapidly with the increasing number of abnormalities and physiological processes reported to have connection with HSP47 expression. Despite the fact that understanding HSP47 mechanism at molecular level is crucial for more advanced research and applications, studies on its molecular behaviour so far seem inadequate. This is especially true for the pH-switch mechanism, which is the only mechanism known to govern the binding and release of the chaperone to collagen. The importance of this mechanism to HSP47 function has previously been established, but has not been fully understood at the structural level. As the release of HSP47 from collagen is an ATP-independent process, His residues have been hypothesised to play a major role in this

mechanism. Thus, the contributions of several His residues to the pH-switch function were therefore set for investigation. Site directed mutagenesis, which is a tool commonly used by molecular biologists, is employed to construct a series of HSP47 mutants targeting His residues of interest. Coupled with circular dichroism (CD) and fluorescence spectroscopy, the importance of those amino acid residues to the protein structure and function could then be investigated.

Hence, the aims of this project were as follows:

1. Construction of His-to-Ala (HA) mutants by site directed mutagenesis on *hsp47* gene previously cloned into an expression vector.
2. Assessment of the effects of HA mutations on the structure and function of HSP47.
3. Investigation of the dynamics of certain regions of HSP47 during pH-transition using intrinsic Trp fluorescence on a series of Trp-to-Phe (WF) mutants.
4. Proposing a molecular model for HSP47 pH-induced structural change based on own data and those already reported in the literature.

1.2 Collagen

Collagen is the most abundant protein component of ECM that gives tissues their specific mechanical and biochemical properties in multicellular animals (Gelse et al., 2003). To date 27 different types of collagen encoded by more than 40 different genes are identified (Koide and Nagata, 2005). All collagens can be characterised by the $(Gly\text{-}X\text{-}Y)_n$ triplets repeats in which position 1 is always occupied by glycyl- residue and positions 2(X) and 3(Y) are usually prolyl- and hydroxyprolyl- residues. Hydroxyproline (Hyp, O) is an amino acid residue formed upon hydroxylation of proline (Pro, P) residue in a reaction catalysed by prolyl hydroxylase (Myllylä et al., 1977; Tuderman et al., 1977). It is based on these molecular characteristics (amino acid repeats) and supramolecular arrangements that give rise to the different types of collagen (I to XXVIII) (Kadler et al., 2007), consisting of specific α-chains encoded by more than 40 different genes (Koide and Nagata, 2005).

Polypeptide chains of collagen molecules are called α-chains while the polypeptides of procollagen are called proα-chains since they contain the additional propeptides at both *N*- and *C*-termini as depicted in Figure 1.1 (Prockop et al., 1979). These α-chains subsequently assemble to form a collagen triple helix which is a right-handed supercoil with three polyproline II-like left-handed helices (Figure 1.2A-C). All peptide bonds in the helix are in *trans*- configuration and the triple helix can exist both as homotrimers or heterotrimers with different α-chains (Beck and Brodsky, 1998). The structure of collagen was first proposed by Ramachandran and co-workers in 1955 (Ramachandran and Kartha, 1955). They suggested that collagen basically consists of three polypeptide chains, winding around each other to form a coiled coil. Each chain was said to be held to its neighbouring chains by two hydrogen bonds (H-bonds). Their proposal was rejected by Rich and Crick for not being stereochemically correct and the impossibility of having two sets of H-bonds to stabilise it (Rich and Crick, 1955). They then went further to suggest two new possible conformations of collagen triple helix based on the (GPO) amino acid repeats, and their own work on polyproline and polyglycine type II. Their new conformations suggested only one H-bond could form between neighbouring chains. Structure I has a glycyl- amino ($-NH_2$) pairing with an adjacent glycyl- carbonyl (-CO) group which creates a bond pointing anticlockwise when observed from the *C*-terminal of the peptide chain, while structure II has the bond pointing clockwise as the pairing is between prolyl- carbonyl group. Analyses of X-ray images of

$(PGP)_n$ peptide suggested that structure II was more viable (Yonath and Traub, 1969). A followed-up investigation using native collagen from kangaroo tail tendon corroborated this finding, and credited structure II of Rich-Crick model for being more accurate (Fraser et al., 1979).

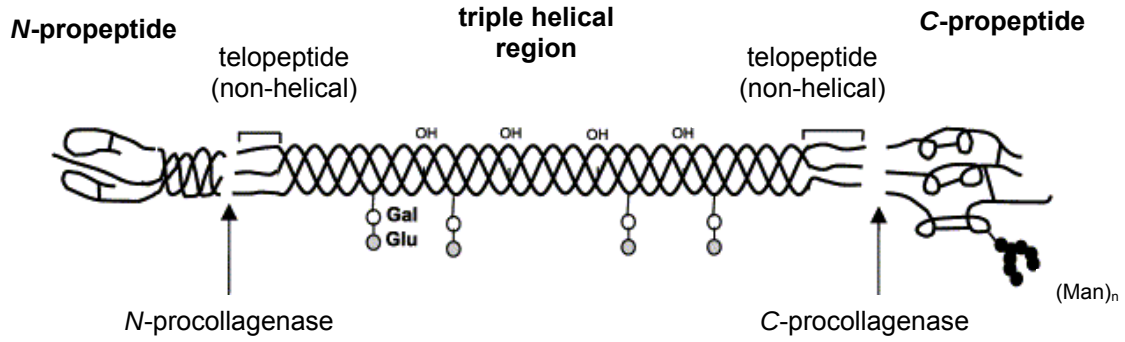


Figure 1.1: General structure of collagen.
Adapted from (Gelse et al., 2003). The structure of procollagen molecule (type I) with the various subdomains and the cleavage sites for *N*- and *C*-proteases (or procollagenases). Also shown are the sugars attached to the helical domain and its non-helical termini. In the *C*-terminal, the propeptide chains are held together by intrachain and interchain disulfide bonds, and have about 10 residues of mannose. These sugar moieties are probably important for secretion. Whereas in the *N*-terminal, only intrachain peptide bonds exist, with a small amount of sugar (Gelse et al., 2003; Guzman et al., 1978; Prockop et al., 1979).

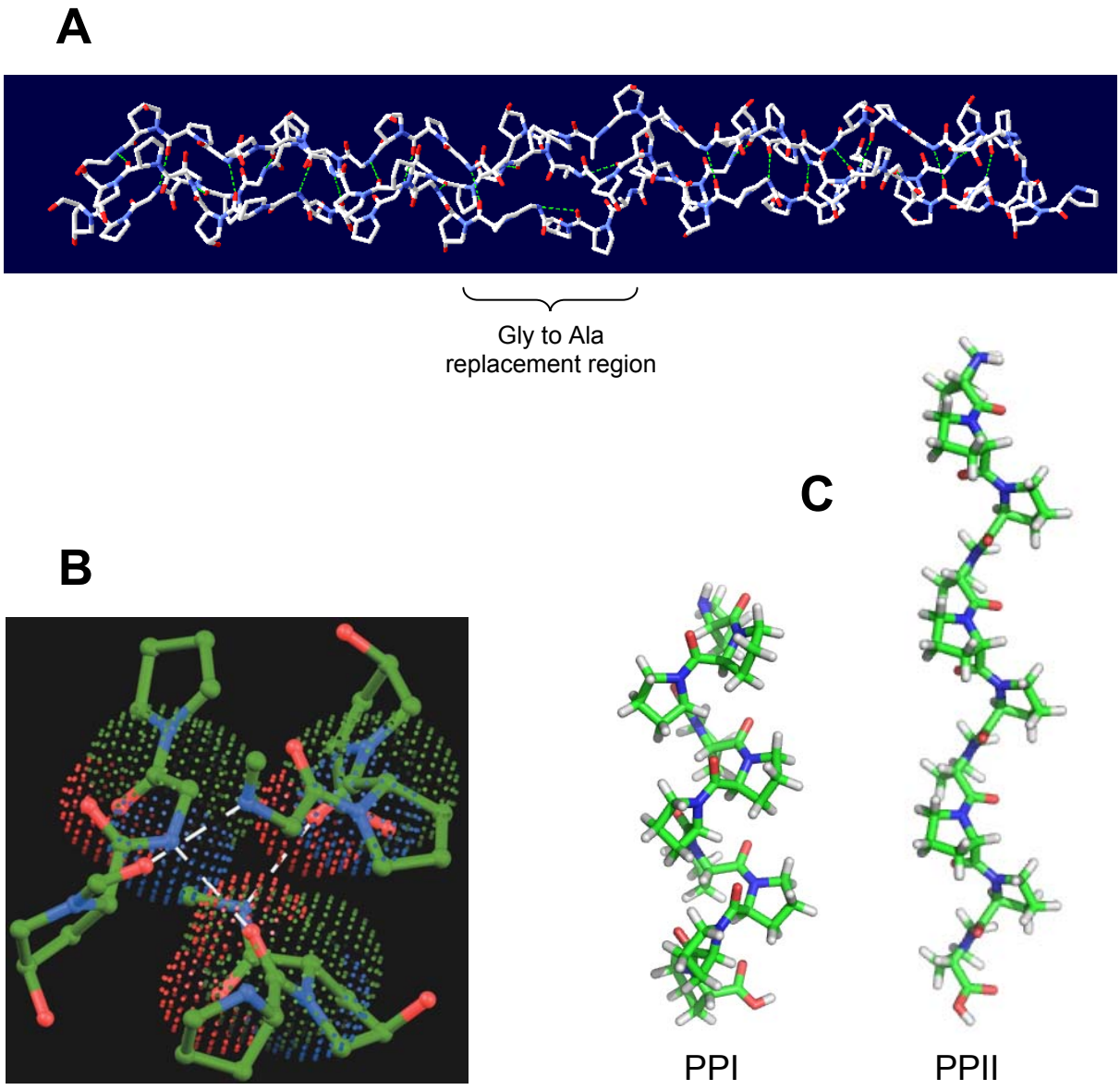


Figure 1.2: The structure of (Pro-Hyp-Gly)$_{10}$ in which the fifth Gly is replaced by Ala. **A.** Full length triple helical peptide, showing the interchain H-bond (yellow dashed line) between Pro carbonyl- and Gly amido- groups. The structure is rendered in ball-and-stick view, PDB ID: 1CAG. It can also be seen from the structure that there is 'local unwinding' in the middle of the triple helix at the Gly replacement site. Visualisation was done with Deepview SwissPDB Viewer v4.0.1. Adapted from (Bella et al., 1994; Voet and Voet, 2005). **B**. The peptide viewed along the helix axis (white dashed lines represent the H-bonds). Adapted from (Bella et al., 1994; Voet and Voet, 2005). **C**. The structures of poly-L-proline type I (PPI, right-handed) and type II (PPII, left-handed) shown in sticks representation. PPI is formed when the backbone dihedral angles are (φ, ψ: -75°, 160°) and are less compact than PPII as the peptide bonds are in *cis* isomers; while PPII have (φ, ψ: -75°, 145°) and the peptide bonds are in *trans* configuration. There is no intrachain H-bond present in either types of helices (Cubellis et al., 2005). Structures adapted from (http://www.cryst.bbk.ac.uk/pps97/assignments/projects/szabo/pphelix.htm).

Because of the cyclic nature and the rigidity of Pro and Hyp, the rotation of the polypeptide backbone is limited and thus contributes to the stability of the triple helix. In addition, the hydroxyl group of Hyp has also been found to help stabilising the triple helix (Prockop et al., 1979). Extensive studies on how Hyp stabilises the triple helix shows that the residue at position-Y is crucial for collagen triple helical conformation. It was earlier proposed that stabilisation is attained through water-mediated hydrogen bonding as the hydroxyl group of Hyp is directed outside the helix (Ramachandran and Chandrasekharan, 1968). Collagen with increased Hyp content also shows increased enthalpic stabilisation. Another mechanism that has been suggested is the inductive effect of the electron withdrawing group (-OH) which prefers *trans*- to *cis*- peptide bond (Brodsky and Persikov, 2005; Eberhardt et al., 1996). Many of these observations were made on short peptides that mimic the sequence of naturally-occurring collagen rather than the full-length native molecule. The reason for this is that these peptides offer the advantages of being easily manageable over natural collagen in terms of the desired amino acid sequence, length and physicochemical properties.

1.2.1 Collagen-mimetic Peptides: Synthesis and Applications

Reseach on collagen-mimetic peptides (CMPs), also known as collagen-related peptides (CRPs) and collagen-like triple helical peptides (THPs) (Cejas et al., 2007; Fields et al., 1996) has been expanding rapidly, ranging from the modelling of collagen structure, to the interaction study between collagen binding clients. The huge amount of information that could be obtained with CMPs has also been utilised in biomedical research, often involving manipulation of collagen-binding receptors acting at various cellular locations (Fields, 2010).

Many CMPs were synthesised using the solid phase procedure, with the earliest peptides synthesised have the repeats of Pro and Gly amino acids (Sakakibara et al., 1972). After that, method to incorporate Hyp into the systematically synthesised peptide was devised, using benzyl- side chain protection of the 4-carbon hydroxyl. The (POG) peptide was found to have higher melting temperature compared to the analogous (PPG) peptide, leading to suggestion that Hyp might have a stabilising effect on the triple helix (Sakakibara et al., 1973). Currently, almost all CMPs are synthesised based on stepwise solid-phase

method using *tert*-butyloxycarbonyl (Boc) or 9-fluorenylmethoxycarbonyl (Fmoc) chemistries. Recent advancement in synthetic techniques enables the making of various CMPs with different chain constituents, not only limited to homotrimer triple helices (Fields, 2010).

The propensity of the peptides to form triple helix is usually monitored using optical rotatory dispersion (ORD) or circular dichroism (CD) spectroscopy. Triple helical peptide gives a distinctive positive molar ellipticity at around 225 nm and a negative ellipticity at around 200 nm. At appropriate wavelengths, these techniques are also used to obtain the temperature at which triple helical to coil structural transition occurs, also known as melting temperature (Fields and Prockop, 1996).

Circular Dichroism Spectroscopy

CD spectroscopy is a very useful technique to study protein folding and as well as the structural elements in both native or unfolded conformation (Schmid, 1997). The CD phenomenon arises from the difference in the absorption of the left-handed (L) and right-handed (R) circularly polarised light. A CD signal will be observed when the optically active molecule differentially absorbs the L and R components giving rise to light that possesses elliptical polarisation. Spectropolarimeters are used to measure the difference in absorbance between the L and R components ($\Delta A = A_L - A_R$) but it is usually reported in ellipticity unit (θ) in degrees. A simple numerical relationship between ΔA and ellipticity is given by Equation 1.1 below.

$$\theta = 32.98\ \Delta A \qquad \text{(Eq. 1.1)}$$

Some of the CD data in this thesis are given in ΔA unit, but mostly in mean residue ellipticity (MRE, deg cm^2 $dmol^{-1}$) unit which normalises the effect of concentration in cases where different samples on different experiments are to be compared. The conversion between these units is given in Chapter 6 (Materials and Method).

Information on the structural composition of proteins can be obtained from their CD spectra. In the far-UV region (170-250 nm), different types of regular secondary structures contribute to distinctive CD spectra (Figure 1.3). Mainly α-helical proteins have negative bands at 222 and 208 nm and a positive band at 193 nm. Proteins with anti-parallel β-sheets have a negative band at 218 nm and a positive band at 195 nm while disordered proteins with irregular structure have a very low ellipticity above 210 nm but have a negative band at around 195 nm. For collagen triple helix, each strand has the conformation that resembles the poly-$_L$-proline in which all of the bonds are *trans* to each other. The CD spectrum of this polypeptide gives a positive band at 218 nm and a negative band at around 195 nm (Figure 1.3). Another type of structure observable with CD is the β-turn (Greenfield, 2007; Kelly et al., 2005). As proteins usually have different types of secondary structures, comparison of a spectrum with those of known three-dimensional structures is usually needed to estimate its structural content. There are several databases that can provide the information on known protein spectra (such as DICHROWEB), and their corresponding structural information based on the experimentally-determined structure (Whitmore and Wallace, 2008).

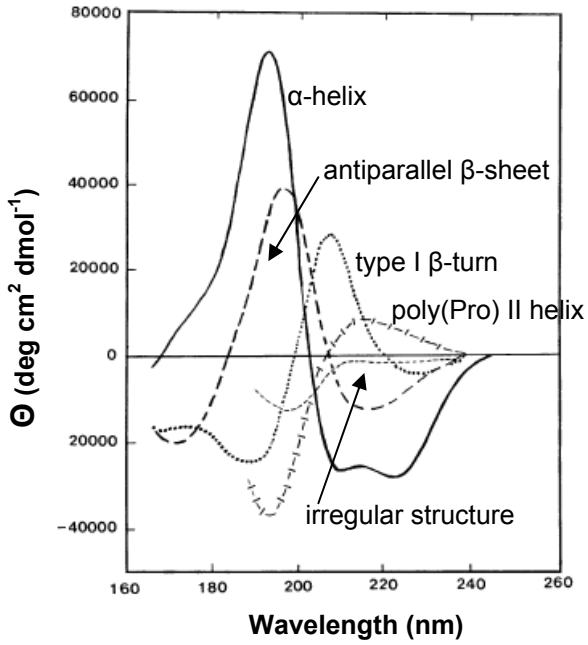


Figure 1.3: Far-UV CD spectra showing the different bands originating from different types of protein secondary structure.
Adapted from (Kelly et al., 2005).

Applications of Collagen Mimetic Peptides

One of the first purposes that led to the synthesis of CMPs is the interest in solving collagen's structure. Many CMPs have been synthesised and characterised for that purpose in the past half a century. A major breakthrough in this path was when Ramachandran and Kartha introduced their model of collagen structure in 1955, quickly followed by Rich-Crick model in the same year and then Yonath-Traub model in 1969. The use of peptides for structural elucidation was initiated by Okuyama and co-workers in 1981. In their study, peptide $(PPG)_{10}$ was crystallised and analysed. Its structure suggested that a hydrogen bonding between the individual peptide strands is contributed by the amino group of Gly to oxygen group of the first Pro residue of the (PPG) repeat in the neighbouring chain. The other specific H-bonds involve water molecules (Okuyama et al., 1981). This was not much different from the Rich-Crick structure II making it the generally acceptable model to date. The first high resolution crystal structure of a collagen triple helix was eventually solved in 1994 using a CMP containing 10 repeats of (POG) amino acids with one replacement of Gly to Ala in the middle of the peptide chains (Bella et al., 1994). The structure allowed the assessment of the effects of Gly substitution, which causes local 'twist relaxation' of the triple helix to accommodate the additional methyl- group from Ala side chain. The high resolution X-ray structure also revealed the specific interaction of the peptide with water molecules in the stabilisation of the disrupted triple helix. This stabilisation is attained by the formation of H-bonds between glycyl- and hydroxyprolyl- residues, with interstitial water molecules, other than the predicted H-bonds between carbonyl- and amide- groups of Pro and Gly respectively (Figure 1.2B) (Bella et al., 1994). The structure of unsubstituted (POG) peptide was eventually solved to explain the implications of proline ring puckering on the triple helix stability (Okuyama et al., 2004).

Recently, focus of CMP synthesis has been towards obtaining new collagen mimetic materials for use in medical applications. The reason for this is to avoid potential rejection of natural collagen by the host (immunogenic responses), especially if it is sourced from a different species, and also on religious basis. There is also the risk of transferring harmful agents, such as prions to humans. The ultimate goal is to design synthetic peptides that are fairly stable *in vitro*, able to form higher order collagenous fibril at certain conditions, non-toxic endogenously, and should also possess properties that can easily be manipulated. The

fact that the field of peptide chemistry is reaching deeper into realm of possibilities makes this aim seemingly within reach. The development of technologies that can covalently link the three strands of triple-helical peptides together, such as solid-phase branching, solution-phase cross-linking, chemoselective and native chemical ligation, disulfide (cystine) knots and various modifications could pave the way for more advanced techniques in peptide design (Fields and Prockop, 1996; Przybyla and Chmielewski, 2010).

One of the most preferred modifications is the linking of the three peptide chains with cystine knots at the *C*-terminal region, arranged in such a way to mimic the staggered arrangement of collagen fibril. This arrangement occurs when collagen triple helices are arranged end-to-end in a tight conformation to minimise surface area by volume ratio (see section 2.7) (Kadler et al., 1996). The 'templated' approach with cystine knots usually enhances the thermal stability of the resulting triple helix (Fields, 2010), but obtaining a much bigger structure proved to be challenging, though possible. Using both $(POG)_n$ and $(PPG)_n$ peptides held together by cystine knots, Kotch and Raines were able to achieve what they called 'self-assembly' of the triple helix *via* intra- and intermolecular triple helix formations (Kotch and Raines, 2006). $(POG)_n$ peptides have a much higher stability than $(PPG)_n$, as apparent from the stabilisation effect of Hyp. They also showed that the peptides self-assembled to form bigger fibrils more than 400 nm. This is longer than that of natural type I collagen (about 300 nm), even though the fibrils formed still contain short-length chains (Kotch and Raines, 2006). An approach giving a gel-like material was also reported by Yamazaki and co-workers using (POG) repeat peptides (Yamazaki et al., 2008). Cystine knots were introduced at different positions along the peptide length with the intention to create a staggered arrangement during self-assembly. The peptides form hydrogel upon incubation at 4°C with varying stability, depending on the knot location (Yamazaki et al., 2008). In all reports, the synthesis routes described to obtain the starting molecules are still considerably long, especially for commercialisation. Hence, research is currently undergoing to find a much shorter route to achieve the highly desirable staggered arrangement *in vitro* for the purpose of gel formation.

1.2.2 Procollagen Assembly in the Endoplasmic Reticulum (ER)

Procollagen biosynthesis occurs in the endoplasmic reticulum (ER) initially as 'preprocollagen' molecule with a signal sequence attached at the *N*-terminus. The presence of this sequence causes the proα- polypeptide to be translocated into the lumen of rough ER. This signal sequence is removed as the chain enters the cisternae of rough ER. Post-translational modifications including hydroxylations of the amino acids at the X- and Y-positions, and glycosylations by the action of hydroxylases and transferases also occur in the ER (Koide and Nagata, 2005). The most extensively characterised of these enzymes is prolyl 4-hydroxylase (P4H) which is responsible for converting prolyl- residues to 4-hydroxyproline (4-Hyp). These hydroxylases act in a specific manner, where the position of the residue hydroxylated in the amino acid sequence is very important. The peptide must also be non-helical, and for P4H and lysyl hydroxylase (LH), only proline or lysine at -Y position is recognised. On the other hand, prolyl 3-hydroxylase (P3H) acts on prolyl residue at position -X. Upon folding to triple helix, these polypeptide chains must possess about 90 4-Hyp residues per chain to be stable at body temperature (out of a total of about 1000 amino acids for fibril forming collagens type I, II and III (Gelse et al., 2003)). The triple helical conformation then prevents further hydroxylation. Glycosylations of the hydroxylysyl residues then follows (Prockop et al., 1979).

Following association of the three pro-α chains at the *C*-propeptide, formation of interchain disulfide bridges initiates the formation of triple helix in the ER. There is a strict requirement for Gly at the third residue as these residues are small enough to allow them to be buried in the core of the helix. The helix then propagates towards the *N*-terminus, with several chaperones reported to help accelerating this process. The triple helix is actually still unstable at body temperature, leading to suggestion that molecular chaperones in the ER, such as HSP47, act to help to stabilise it *in vivo* (Koide and Nagata, 2005; Leikina et al., 2002).

Similar to other proteins, after being assembled in the ER, procollagen molecules are transported to the Golgi before leaving the cell (Prockop et al., 1979). Beforehand, the globular propeptides at the amino- and carboxy- termini are cleaved by aminoprotease and carboxyprotease respectively. This then triggers fibrillogenesis process, which then prevents further unfolding (Leikina et al., 2002). Eventually, the collagen molecules assemble into

fibrils, either in the plasma membrane or in the transport compartment between Golgi and plasma membrane (Canty and Kadler, 2005). These so-called ‘immature fibrils’ do not possess the necessary tensile strength and require covalent cross-linking to strengthen them. This process is catalysed by lysyl oxidase (LO), which oxidises certain terminal lysyl- and hydroxylysyl- residues. As a result, two types of cross-links are generated. Firstly, intramolecular cross-links whereby two α-chains of the same molecule are joined together by aldol condensation of two aldehydes generated during the oxidation process. Secondly, cross-links between the collagen triple helices which involve bond formations between aldehyde generated from lysine and hydroxylysine and amino group of a second lysine or hydroxylysine. These are then followed by several more stable cross-links (Prockop et al., 1979). These processes are summarised in Figure 1.4.

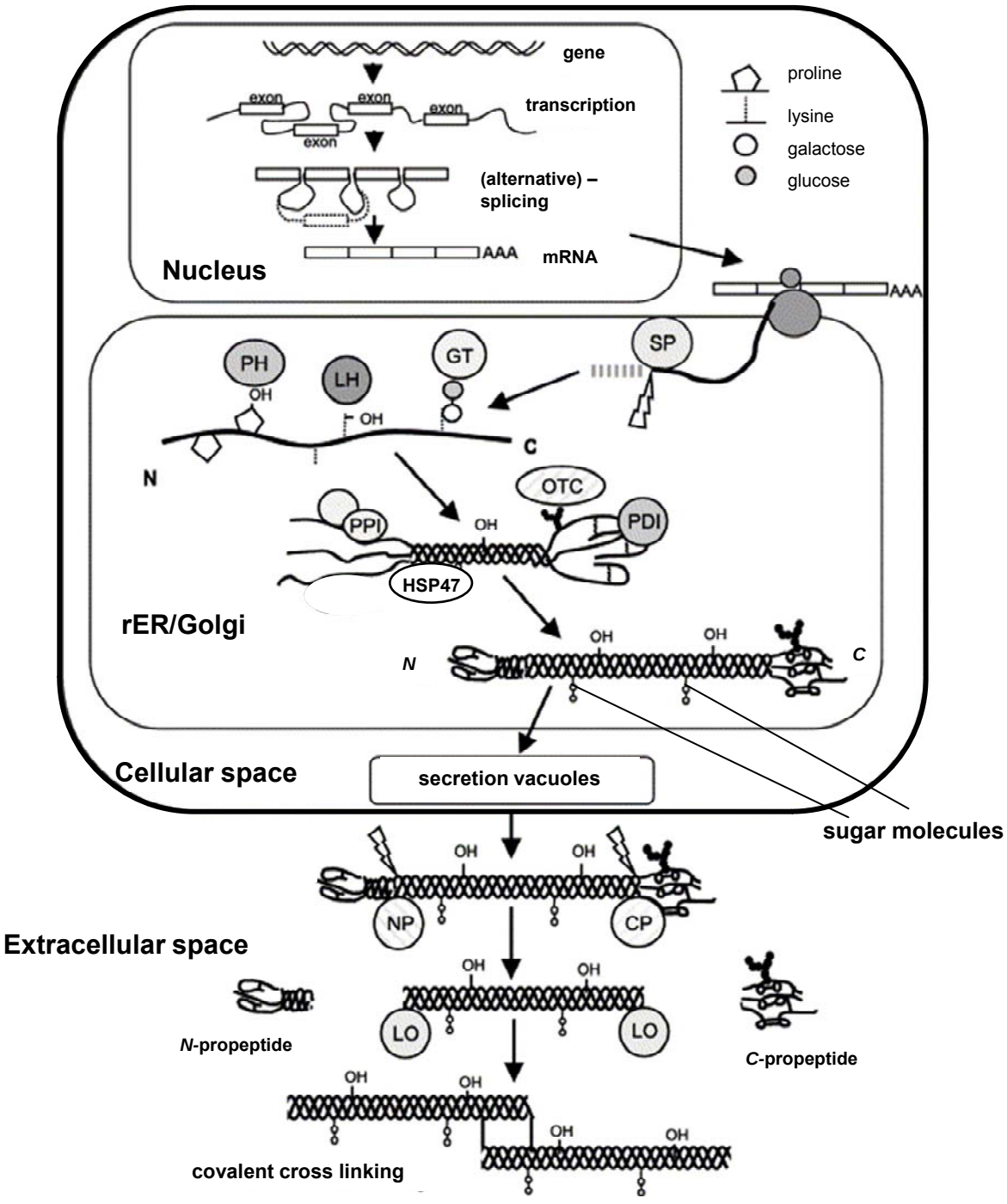


Figure 1.4: Schematic diagram of procollagen assembly showing roughly the enzymes involved during various stages of the biosynthesis.
(SP: signal peptidase; GT: hydroxylysyl galactosyltransferase and galactosylhydroxylysyl glucosyltransferase; LH: lysyl hydroxylase; PH: prolyl hydroxylase; OTC: oligosaccharyl transferase complex; PDI: protein disulfide isomerase; PPI: peptidyl-prolyl *cis-trans*-isomerase; NP: procollagen *N*-protease; CP: procollagen *C*-protease; LO: lysyl oxidase; HSP47: heat shock protein 47). Adapted from (Gelse et al., 2003).

1.2.3 Molecular Chaperones in Collagen Biosynthesis

The process of collagen biosynthesis from procollagen production in the ER to transport to Golgi apparatus, and then finally to ECM involves actions by several general and specific molecular chaperones. Molecular chaperones are a group of proteins that assist the correct folding of other polypeptides or proteins but are not components of the final folded structures (Ellis and van der Vies, 1991). The term 'molecular chaperone' describing proteins that ensure the correct folding of other proteins into functional oligomeric structures was proposed by John Ellis (Ellis, 1987) when describing the chaperoning action of a chloroplast protein in the assembly of a photosynthetic CO_2-fixing enzyme, ribulose bisphosphate carboxylase-oxygenase (Rubisco) from its subunits.

Earlier, the information required for correct folding of protein was said to be specified by the amino acid sequence, as proposed by Anfinsen from the observation made with ribonuclease molecules (Anfinsen, 1973; Anfinsen et al., 1961). Due to this, earlier protein folding studies were based on the knowledge that proteins fold in the cell by the self-assembly process. Until it was found that during some studies, at certain temperatures and polypeptide concentrations, a fraction of the peptide chains failed to fold correctly, and formed aggregates instead (Ellis and Minton, 2006). This is even more significant in cells where the concentration of proteins and others macromolecules are very high (up to 400 g/L) (Hartl and Hayer-Hartl, 2009). The discovery of molecular chaperones then prompted scientists to re-evaluate their initial theory on protein folding, taking into account the crucial function of chaperones in assisting the folding of polypeptide *in vivo* and preventing the undesirable misfolding steps that could lead to aggregations. The chaperones are also reported to assist protein assembly *in vitro* by decreasing the aggregated chain fractions (Ellis and Minton, 2006).

It is not surprising that the assembly and transport of collagen molecules are assisted by a succession of molecular chaperones. The ER alone has many chaperones that are involved in the biosynthesis of various other types of proteins. A review by Koide and Nagata (2005) provides an extensive overview on most chaperones involved in the synthesis of collagen, as early as the production of nascent procollagen chain to the formation of complete collagen triple helical molecule. The chaperones can be divided into 'general' and 'specific'

chaperones. Among the general chaperones are the immunoglobulin heavy-chain binding protein (BiP) or also known as 78 kDa glucose-regulated protein (GRP78), GRP94, calnexin and calreticulin, peptidyl-prolyl *cis-trans* isomerase (PPIase) and protein disulfide isomerase (PDI), while HSP47 is classified as a specific chaperone due to its single substrate recognition. Figure 1.5 lists these general and specific chaperones and their suggested binding sites on collagen molecule during biosynthesis.

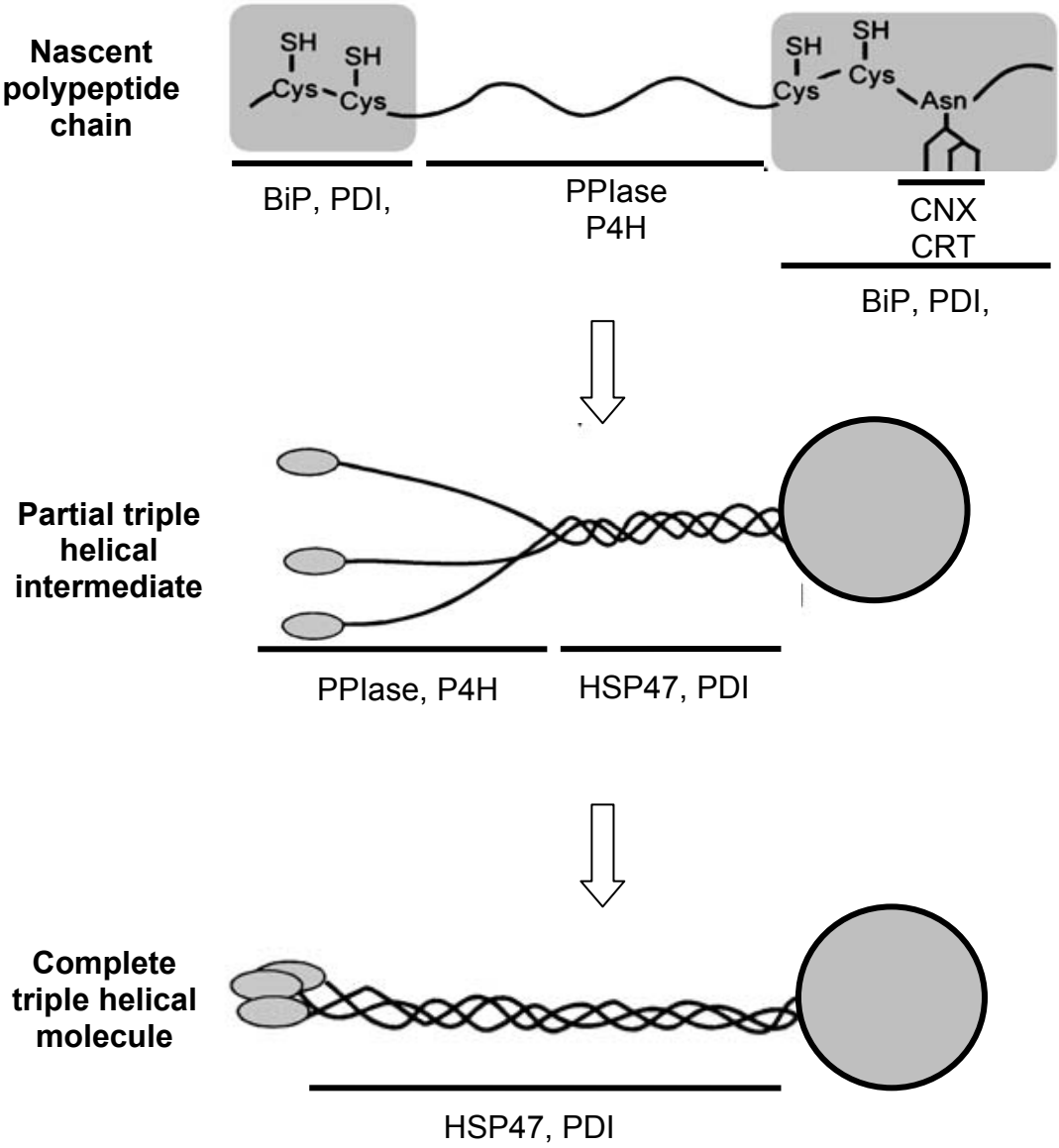


Figure 1.5: Collagen chaperones binding regions (reported and predicted) at various stages of collagen biosynthesis in the ER.
These include general ER chaperones like BiP, PDI, Grp94, CNX, CRT and PPIase, while P4H and HSP47 are classified as specific collagen chaperones. Adapted from (Koide and Nagata, 2005).

BiP/GRP78 and GRP9

The immunoglobulin heavy-chain binding protein (BiP), also known as 78-kDa glucose-regulated protein (GRP78), is the only member of heat shock protein 70 (Hsp70) family that resides in the ER. The chaperone is required for translocation of newly synthesised polypeptides across the ER membrane and subsequently for their folding and assembly in the ER lumen. BiP recognises a wide variety of nascent polypeptides that have no sequence similarity but can accurately discriminate between properly folded and unfolded structures (Blond-Elguindi et al., 1993). In collagen biosynthesis, BiP is proposed to promote protein assembly especially at the early stage of triple helical formation. It has been shown to bind to abnormal type I procollagen, specifically the pro-α chains that possess *C*-terminal propeptide mutation that interferes chain association and interchain disulfide bond formation (Lamandé et al., 1995). BiP binding to secretory proteins is transient before they have folded correctly, but it remains bound to misfolded and aberrant proteins (Lamandé et al., 1995). Hence, it is also possible that BiP is involved in the quality control of the terminal propeptide of the newly-synthesised procollagen molecules. GRP94 has been shown to form a complex with BiP in the stabilisation of nascent collagen chains (Ferreira et al., 1994; Ferreira et al., 1996).

Calnexin and calreticulin

Calnexin (CNX) was discovered in 1991 in several forms but was mainly related to multimeric T-cell antigen receptor (TCR), a chaperone that interacts with partially folded protein complex in the ER and is absent from the fully folded one (Degen and Williams, 1991; Hochstenbach et al., 1992). CNX was found to share sequence motifs with calreticulin (CRT), a major calcium-binding intraluminal resident ER protein. The name CNX was proposed as it represents a membrane-bound homologue of CRT (Wada et al., 1991). CNX and CRT have been shown to interact transiently with a wide array of nascent proteins in the ER, especially the folding or assembly intermediates but not with fully folded structures (Ou et al., 1993). Hence, these chaperones are associated with the quality control by prolonging the retention of unfolded protein in the ER. Specifically, CNX and CRT interact with the glycan moieties of substrate glycoproteins after they have been trimmed by glucosidases I

and II to the monoglucosylated form (Helenius and Aebi, 2001). The critical function of these chaperones in collagen biosynthesis is yet to be demonstrated, but they have been postulated to bind to the *C*-terminal propeptide region that has been glycosylated with *N*-oligosaccharide right at the early stage of procollagen synthesis (Harwood et al., 1975). In type II collagen for example, glycosylations have been suggested to mediate the interaction with proteoglycans (Gelse et al., 2003).

PPIase and PDI

Peptidyl-prolyl *cis-trans* isomerase (PPIase) catalyses the *cis-trans* isomerisation of proline peptide bonds, which can be a rate-limiting reaction in many slow molecular conversions (Lang et al., 1987). Unfolded type I procollagen was shown to have 16% of X-Pro peptide bonds (where X is any amino acid) in *cis* conformation, and 8% in X-Hyp bonds (Sarkar et al., 1984) while the rest are in *trans*. Hence the newly-folded procollagen molecules require a chaperone to assist the formation of *trans* peptide bond for increased stability. This chaperone is important in the synthesis of correctly-folded collagen molecules as inhibition of its action in the presence of cyclosporin A results in the collagen produced by fibroblasts are over-modified (especially lysine hydroxylation) with raised intracellular degradation (Steinmann et al., 1991).

Protein disulfide isomerase (PDI) is a chaperone that catalyses the correct pairing of disulfide bonds during folding processes assisted by disulfide bridges. It involves a redox reaction to form reduced/oxidised forms of PDI and the protein substrate (Creighton et al., 1980; Darby and Creighton, 1995). In procollagen molecules, PDI is involved in the formation of disulfide bonds within the α-chains (intra-chain) and among the chains (inter-chains) (Wilson et al., 1998). Another collagen chaperone in the ER, which is normally associated with PDI, is prolyl 4-hydroxylase (P4H). P4H catalyses the formation of 4-hydroxyprolines by the hydroxylation of proline residues. The formation of hydroxyproline helps to stabilise the triple helical conformation of collagen, through water-mediated hydrogen bonding (Ramachandran and Chandrasekharan, 1968) or inductive effect of the hydroxyl group attached to the proline residues (Eberhardt et al., 1996). PDI also exists as a subunit of the active oligomeric complex of P4H (Koivu et al., 1987). As a subunit of P4H, the chaperone complex prevents the formation of insoluble aggregates of procollagen

molecules by binding to the unhydroxylated chains in the ER. Additionally, the complex also binds the incompletely folded parts of the collagen triple helices, increasing their solubility and preventing immature secretion by longer ER retention (Wilson et al., 1998).

1.3 Heat Shock Protein 47 (HSP47)

HSP47 has been identified by several names including colligin (Kurkinen et al., 1984), CB48 (Yannariello-Brown and Madri, 1990), J6 (Wang and Gudas, 1990) and gp46 (Clarke et al., 1991) based on the cell type by which it was discovered. All other nomenclatures have rarely been used to identify this chaperone, and hence for standardisation the name HSP47 is used throughout this study. Historically, HSP47 was first found in 1986 in heat-inducible chick embryo fibroblasts and was shown to be able to bind type I collagen (Nagata et al., 1986). HSP47 synthesis is increased upon cell incubation at 42°C, hence the name. The isoelectric point (p*I*) of HSP47 is around 9.0 and it is an unusually basic protein. The localisation of HSP47 in the ER was demonstrated using immunocytochemical studies with monoclonal and polyclonal antibodies against HSP47 *in vitro* (Satoh et al., 1996). The presence of an ER retention sequence, Arg-Asp-Glu-Leu (RDEL) can also signify that HSP47 is recycled back to the ER upon dissociation in the Golgi *via* the action of receptors to maintain the concentration of HSP47 inside the ER (Satoh et al., 1996).

HSP47 is a cell-specific protein, and only expressed in collagen-synthesising cells (Saga et al., 1987), such as in fibroblasts. The complementary DNA (cDNA) has been cloned independently from F9 mouse carcinoma cells (Wang and Gudas, 1990) and from a cDNA library of heat-shocked chick embryo fibroblasts (Hirayoshi et al., 1991).

1.3.2 pH-dependent Conformational Change of Proteins

pH is a critical factor in determining a correct folding and optimum function of proteins. Enzymes for example, have certain narrow pH ranges at which they become active. pH affects the binding affinity of the substrate to the enzyme, either due to the ionisation of amino acid residues of the enzyme; or due to the ionisation of the substrate itself. A classic example is haemoglobin, where its affinity to oxygen depends on pH and carbon dioxide concentration.

Haemoglobin is a well-characterised protein made up of four subunits (tetramer). In vertebrates, the subunits are called α and β subunits. Each subunit also has a heme group and hence one oxygen binding site. Upon binding to oxygen, haemoglobin undergoes changes in the quarternary structure that brings the β subunits closer and narrows its central cavity. pH plays a crucial role in haemoglobin action as an increase in proton concentration enhances its tendency to release oxygen, a phenomenon known as the Bohr effect. This process reportedly involves the protonation of a His residue (H146) located at the *C*-terminus of each β subunit. In the non oxygen-binding form, this residue forms a salt bridge with an Asp residue (D94) thus stabilises the protonated form against dissociation and increases its pK_a. The salt bridge is not observed in the binding form. This means that at low pH (high proton concentration), the non binding form is preferable and this increases the tendency of haemoglobin to release oxygen. This is important in the event of high metabolic activity when tissues produce more lactic acid, hence higher proton concentration. This in turn causes haemoglobin to release more oxygen molecules to cope with the higher metabolic requirement (Mathews et al., 1999; Stryer, 1995).

Like haemoglobin, many proteins change their conformations at different pH values and most of which are critical for their functions. It is also well known that pH can greatly affect the tendency of proteins to precipitate, causing collapse in their structural integrity. Notwithstanding, milder pH variation can trigger measureable changes, that can either be localised to certain regions of the protein (particularly the ligand binding site), or the whole protein itself. In most cases, the changes are triggered by the ionisation of amino acid side chains strategically located in the protein interior that are accessible to solvent. In the cellular pH range, several amino acid side chains are ionised (Figure 1.6).

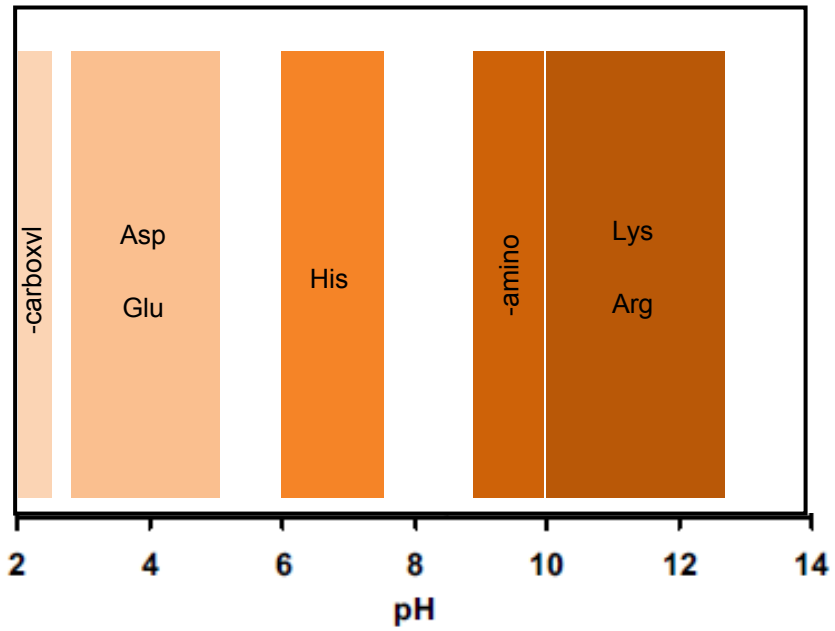


Figure 1.6: Major ionisation zones of several amino acid side chains.
The side chains can be divided into acidic (Asp has a side chain pK_a of about 3.9, Glu 4.2) and basic (Lys pK_a of 10.0, Arg 12.5). His has a pK_a of around 6.0 and is loosely classified as basic. Cys and Tyr are omitted as the side chains are mostly neutral at physiological pH range (Chan and Warwicker, 2009; Mathews et al., 1999). The ionisation regions for the –amino and –carboxyl groups are also shown for comparison.

An example of a side chain-triggered local conformational change is in the case of a protein that forms part of viral infection machinery. A His residue (H263) located in a fusion loop (called FL2) of glycoprotein B of herpes simplex virus type 1 (HSV-1) has been reported to trigger a local conformational change upon exposure to low pH. This rearrangement is suggested to play a role during endosomal entry of HSV probably *via* enhanced interaction with other glycoproteins of the acidic state (Stampfer et al., 2010). The local conformational changes of FL2 are illustrated in Figure 1.7. In this case, H263 is suggested to be a 'trigger' residue that controls the change in conformation that largely affects the function of the virus.

A more profound pH-triggered conformation change was observed in monomer-to-dimer transition of dynein light chain protein (DLC8). DLC8 is the smallest subunit of the microtubule-associated cytoplasmic dynein motor assembly, usually involves in intracellular transport of materials within cell cytoplasm (King, 2000). Very much like in glycoprotein B of HSV, the dimer dissociation of DLC8 is also controlled by the protonation of His residues. Instead of just a single residue, this process is suggested to involve the ionisation of several residues, which gradually increases the flexibility of the β-sheet loops at the dimer interface. Three His residues which are predicted to have slightly different pK_a by NMR assignments, are ionised at different pH values and subsequently trigger strand repulsions. The repulsions finally cause the abolition of salt bridges before the dimer fully dissociate into monomer at pH as low as 3.0 (Mohan et al., 2006).

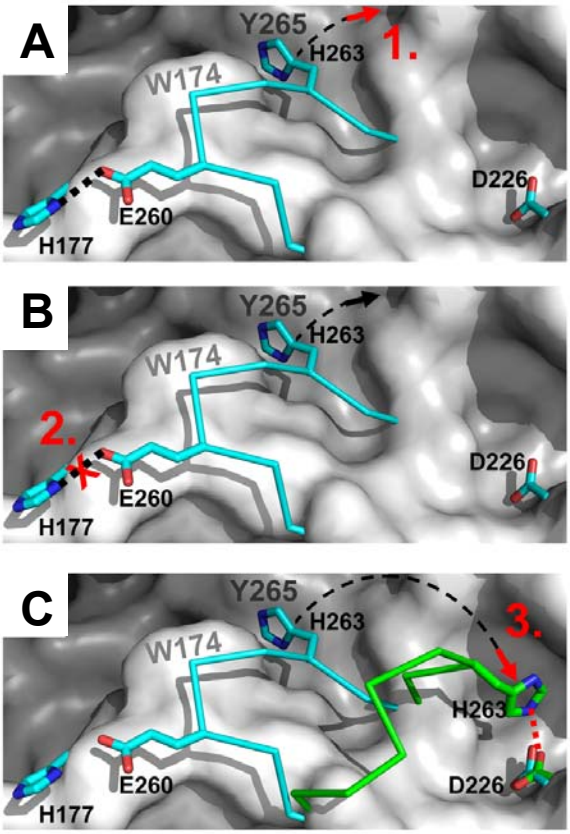


Figure 1.7: Proposed pH-induced conformational changes of loop FL2 of glycoprotein B of HSV.
Firstly in (**A**), as illustrated by *step 1*, His 263 (H263) becomes protonated and tends to repel the uncharged region formed by Trp 174 (W174) and Tyr 265 (Y265). Then in (**B**) the loop is pulled into more hydrophilic environment, as a result the hydrogen bond between H177 and Glu 260 (E260) is abolished (*step 2*). Finally in (**C**) H263 forms a new salt bridge with Asp 226 (D226) keeping the new conformation stable and in place (*step 3*) (Stampfer et al., 2010). Neutral WT protein is shown as white surface representation, residues 258-264 shown as sticks representation in cyan and the acidic conformation is shown in green stick representation.

1.3.3 HSP47 Binding and Releasing Properties

The expression of HSP47 correlates with collagen expression indicating that HSP47 plays an important role in the processing, secretion and higher order assembly of collagen (Dafforn et al., 2001; Ishida et al., 2006; Satoh et al., 1996). Binding is also observed with gelatin, which is a denatured form of type I and III collagen. Unlike some proteins, glycosylation and phosphorylation of HSP47 are not essential for the binding event (Jain et al., 1994b; Kurkinen et al., 1984). HSP47 was earlier shown to be able to bind both single-stranded procollagen chains and triple helical collagen (Satoh et al., 1996). A more recent studies demonstrated that it is more likely to bind triple helical collagen instead of the single-stranded ones (Koide et al., 2006a). Dafforn and co-workers (Dafforn et al., 2001) has established that under normal folding conditions, HSP47 is likely to exist as a native, five-stranded β-sheet A (similar to the structure of α_1-AT shown in Figure 1.8A, B) and is metastable. Metastability in serpins exists because the native state of serpins is not the most thermodynamically stable, and this increases the tendency of the protein to form more stable conformations (Im and Yu, 2000). This is achieved either after cleavage of the centre loop (RCL) upon binding with protease (Engh et al., 1989), or through conformational rearrangement at suitable conditions (Mottonen et al., 1992).

Certain chaperones, such as heat shock protein 70 (Hsp70), DnaJ, GroEL and GroES complex, requires ATP for energy supply during chaperoning action. Hsp70 family members have the basic function of binding and releasing hydrophobic segments of unfolded polypeptide chain in a reaction cycle involving the hydrolysis of ATP. This class of chaperone can be found in the ER, cytosol, mitochondria and chloroplast (Georgopoulos and Welch, 1993). Hsp70 homologue in bacteria is called DnaK. Another chaperone, DnaJ is found to interact closely with DnaK by stimulating its ATP hydrolysis activity, which is significantly higher in the presence of another chaperone GrpE. It was suggested that DnaJ binds DnaK, causing the rate of ATP hydrolysis to accelerate. GrpE then binds and subsequently accelerates the rate of release of the bound ADP (Liberek et al., 1991). These chaperones are involved in bacterial protein assemblies during stress conditions and also some other normal cellular functions. Likewise, GroEL/ES are chaperone complex originally the products of *groE* gene locus of *E. coli*. GroEL, which is the large sub-unit possesses a cylindrical-shaped molecular arrangement with a central hole about 6 nm in diameter,

constituted of 14 identical subunits that are arranged in two stacked rings of seven sub-units each. Electron micrograph revealed that GroEL has a 'double-doughnut' shape. At one end of the double-doughnut is GroES, which is a smaller subunit, heptameric ring-like molecule and known to bind only to GroEL (Lodish et al., 1995). In short, the complex assists protein folding when ATP and GroES bind the same GroEL ring (called *cis* ring) as misfolded polypeptide. The polypeptide is allowed to fold correctly in the ring cavity until dissociation from the ring is triggered by ATP hydrolysis in the other ring (*trans* ring) that allows ATP entry and another polypeptide. This process then continues back and forth (Horwich et al., 1999). So these chaperone complexes utilises ATP as energy source while assisting the folding of polypeptides. HSP47, unlike these chaperone systems, does not require the presence of ATP for association with and dissociation from collagen, prompting a number of studies to investigate its mechanism of action (Nakai et al., 1992).

Binding of HSP47 to collagen *in vitro* was reported to be in a pH-sensitive manner. The binding was observed at pH 8.0 and abolished at pH 6.3 (Saga et al., 1987). The structural aspects of this pH-change property studied using circular dichroism (CD) and fluorescence spectroscopy suggested that HSP47 undergoes a two-step conversion from high to low pH state *via* an intermediate structure. The β-sheet content is estimated to increase during the transition by up to 10-20%, based on CD spectra curve-fitting procedure (Thomson and Ananthanarayanan, 2000). This property is attributed to the protonation and deprotonation of histidine (His) imidazole side chain due to the transition midpoint around pH 6.2, which resembles that of the pK_a of His side chain (El-Thaher et al., 1996). In the presence of HSP47, fibril formation of collagen is prevented at neutral but not at low pH (Thomson and Ananthanarayanan, 2000). Further research reported on the pH-dependent binding of HSP47, including to a synthetic collagen-like peptide $(PPG)_{10}$, showed HSP47 undergoes conformational change as pH is reduced from 7 to 6 causing dissociation from the peptide (Dafforn et al., 2001).

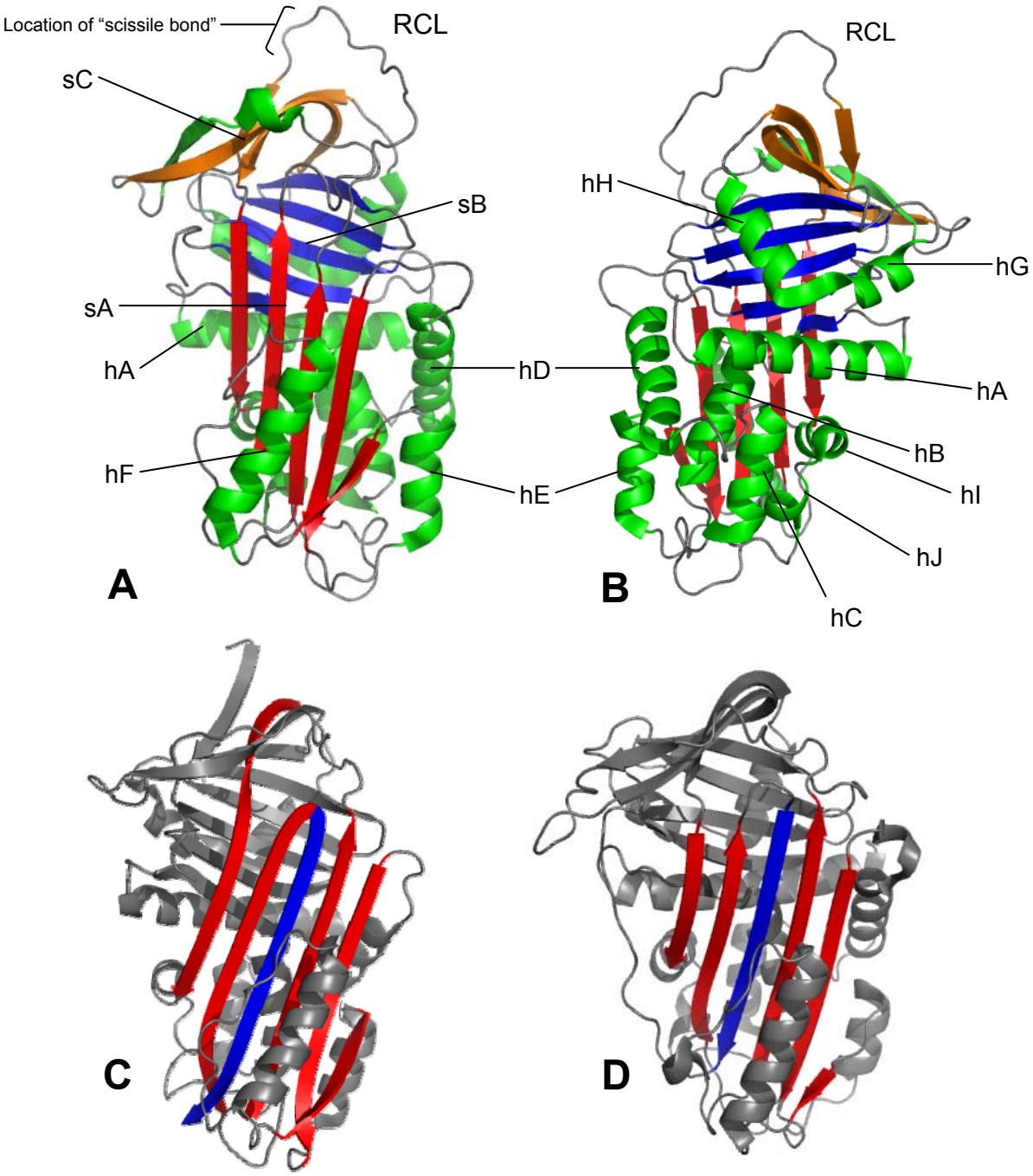


Figure 1.8: Typical structure of serpin family members.
A. The 'front' view of α_1-AT (PDB ID: 1QLP) showing the positions of central β-sheet A (sA), sheet B (sB) and sheet C (sC) above sB. **B**. 'Back' view of α_1-AT showing that all the helices (A to I, labelled hA to hI) except hF lie on the relative back side of the protein. The serpin reactive centre loop (RCL) is exposed at the top of the molecule. **C**. The cleaved (single chain) structure of α_1-AT (PDB ID: 7API). The RCL is shown in blue and the rest of A-sheet is in red. **D**. The latent structure of PAI-1 (PDB ID: 1C5G). Figures adapted from (Elliott et al., 2000; Engh et al., 1989; Gettins, 2002; Tucker et al., 1995).

More detailed binding studies suggested that the dissociation of HSP47 from procollagen occurs as the HSP47/procollagen complex enters *cis*-Golgi compartment when the concentration of free HSP47 decreases as the ER-retention receptors retrieve HSP47 (Natsume et al., 1994). Using surface plasmon resonance, it was shown that this rapid decrease of free HSP47 causes dissociation because of its high dissociation rate constant (k_{diss}) from collagen of more than 10^{-2} s^{-1} (Natsume et al., 1994; Satoh et al., 1996). Despite this, the occurrence of pH-dissociation *in vivo* could not yet be confirmed. The conformational requirements of collagen/collagenous peptide for HSP47 recognition has been studied extensively by Koide and co-coworkers (Koide et al., 2000). Using a library of collagenous peptides, they demonstrated that the hydroxylation of the second Pro residue in the sequence $(Pro\text{-}Pro\text{-}Gly)_n$ reduced the interaction between HSP47 and the peptide. As tested in yeast two-hybrid systems, Pro residues in the X and Y positions do not contribute equally to the interactions with HSP47 as single Pro to Ala substitution at the X position abolishes the binding but this does not occur at the Y position. Peptide library screening showed that Arg at position Y is preferred for HSP47 binding, followed by Pro, up to 5-6 times more. This is related to the triple helix stability of the former peptides compared to the other amino acids (Koide et al., 2000).

The tendency of HSP47 to bind the triple helical part of collagen molecules was also reported by other studies (Thomson and Ananthanarayanan, 2000). Using CNBr fragments of collagen type I and II, the region in collagen molecule which has got the highest affinity for HSP47 binding was determined to lie towards the *N*-terminus (Thomson et al., 2003). Synthetic peptides, like $(PPG)_{10}$ were also used to study the interaction with HSP47 leading to the finding that the binding of HSP47 to $(PPG)_{10}$ is less tight when compared to full-length collagen, given that $(PPG)_{10}$ is monomeric. The binding event also promotes the conformational ordering of the peptide bonds in Pro residues leading to P_{II} helix formation (Dafforn et al., 2001). An investigation has also revealed that HSP47 recognises triple-helical collagen clients, and is unlikely to bind to the molecules before the triple helix is formed (Koide et al., 2006a). By using short (POG) heterotrimeric collagen peptides, it was demonstrated that at least one Arg residue per triple helix at Y position is sufficient for HSP47 recognition. Enhanced interaction was also observed when the amino acids at -3 position (Y^{-3}) from Arg at the Y position is Thr, with the highest affinity observed when both Arg and Thr are located at the same polypeptide in the triple helix. These suggest that HSP47

helps stabilising the triple helix when the cell is under stressed condition, by increasing binding to the lower affinity sites under stressed environment, other than the usual high affinity ones during normal chaperoning function (Koide et al., 2006b).

As in the case of other pH-dependent proteins, pH changes are suggested to be particularly important for HSP47 as it is a collagen-specific chaperone and its release from collagen is triggered by protonation. This means that the only mechanism of action for HSP47 is *via* pH-induced conformational change and thus justifies the needs to fully understand the general and specific structural rearrangement during the process.

1.3.4 Functions in Collagen Biosynthesis

As a unique chaperone that only binds to collagen, various functions have been proposed for HSP47. Although the actual mechanism of the chaperone is still being debated, among the suggested functions include:

a) Protecting the newly-synthesised pro-collagen chains from being folded or aggregated until translation is complete (Satoh et al., 1996).

b) Stabilising newly-synthesised pro-collagen in the event of metabolic stress (Ferreira et al., 1994).

c) Inhibiting the degradation of pro-collagen in the ER (Jain et al., 1994a).

d) Helping the formation of three-dimensional assemblies of collagen molecules (Dafforn et al., 2001; Kurkinen et al., 1984; Nakai et al., 1992) by inducing the formation of a polyproline type II conformation (Dafforn et al., 2001).

e) Assisting the formation of correct trimers among different α-chains in heterotrimeric collagens, probably with other collagen chaperones, and preventing the secretion of incorrectly-folded heterotrimers (Kobayashi and Uchiyama, 2010).

f) Stabilising the correctly-folded triple helix against external stress, such as elevated temperature (Koide et al., 2006b; Makareeva and Leikin, 2007).

g) Stabilising partially folded triple helical intermediates that would be unstable at body temperature (Koide et al., 2000; Koide et al., 2006b).

h) Binding to misfolded areas within fully formed collagen to reduce aggregation (Dafforn et al., 2001).

i) Preventing the secretion of immature or misfolded procollagen molecules (Nakai et al., 1992).

j) Retaining the underhydroxylated procollagen in the ER longer (Satoh et al., 1996).

The advent of many sophisticated techniques in biological research has enabled HSP47 researchers to further support or refutes all the suggested functions listed above. For example, the function of HSP47 to help stabilise the collagen chain in the event of cellular stress has

been receiving more attention as recent study on HSP47 binding motives on collagen molecules suggests the presence of high-affinity and low affinity binding sites. These low affinity binding sites are thought to be the sites where HSP47 molecules bind to in stress conditions, apart of the normal binding site during regular cellular events (Koide et al., 2006a; Koide et al., 2006b). Furthermore, HSP47 involvement as early as pro-collagen synthesis is now heavily debated as it is recently found to bind to single stranded collagen peptide significantly less than the triple helical molecules (Koide et al., 2006a; Koide et al., 2002). These then suggest that other collagen chaperones are involved in stabilising the single stranded nascent pro-collagen molecules, like BiP, CNX and CRT. The ability of HSP47 to retain underhydroxylated collagen triple helices is also supported by further research on the importance of hydroxylation for HSP47 binding and collagen transport to the Golgi (Koide et al., 1999; Koide and Nagata, 2005; Walmsley et al., 1999). Despite these, there are many other functions that are still being investigated, especially involving the molecular changes induced upon HSP47 binding to collagen molecules.

1.3.5 HSP47 and Diseases

Wound Healing

HSP47 is a chaperone that is expressed constitutively in collagen-synthesising cells and tissues. Its expression has been shown to increase during induced skin and gastrointestinal wound. In neonatal rat for example, a few days after wound introduction, increased collagen expression and thus an increase of HSP47 expression could be observed in dermis and subcutaneous tissues (Wang et al., 2002). Detection of HSP47 expression was performed using immunoblotting while collagen expression was probed using RT-PCR of collagen type I mRNA. Similarly, in rat model induced with gastric ulcer, an increase in HSP47 expression was recorded, especially from the third day to twelfth day after ulcer induction. Most of the increased expression was observed in fibroblast cells around the ulcer wound, probably involved in ulcer healing. Healthy gastric tissues also showed HSP47 expression, but at a much lower level (Guo et al., 2002). Other than just in skin and gastric ulcers, HSP47 expression was also detected in induced oral ulcer. In normal tissues, HSP47 is usually found in the suprabasal layer of the oral epithelium and the lamina propria, which is a layer of connective tissue under the epithelium. Upon ulcer induction, HSP47 expression increases significantly especially at the stage of wound healing, before decreasing to the initial level after complete regeneration. Hence, like in the case of gastric ulcer healing, this suggests that HSP47 also assists collagen deposition during remodelling phase of the induced oral ulcerative wound (Vasques et al., 2010).

Fibrosis

HSP47 has been known to be involved in many types of fibrotic diseases, characterised by the extensive depositions of collagens. A large number of studies have reported the up-regulation of HSP47 in the event of fibrosis. A recent observation suggested that the increase in HSP47 expression promotes excessive collagen assembly in patients with chronic graft-versus-host disease (cGVHD), which involves uncontrolled fibrosis in the lacrimal glands (Ogawa et al., 2007). Furthermore, HSP47's abnormal expression has also been associated with oral submucosal fibrosis, which is a pre-cancerous condition (Yang et al., 2008).

Another type of fibrosis that would be a potential target for treatment is liver fibrosis. It is a result of the chronic damage to the liver due to the accumulation of extracellular matrix (ECM) proteins. Cirrhosis mortality, which is the end stage of fibrosis, has been rising in numbers in the UK lately (Henderson and Iredale, 2007) suggesting an urgent need for treatment. It is worthy to note that in industrialised nations, the main reasons for this fibrosis are hepatitis C infection, alcohol abuse and non-alcoholic steatohepatitis (NASH). The relationship of HSP47 expression and the progression of fibrosis have been well researched (Masuda et al., 1994; Razzaque et al., 1998), and in cirrhotic liver, HSP47 is shown to be expressed intensely (Brown et al., 2005; Sato et al., 2008). Nonetheless, HSP47 is also present in control liver tissue suggesting that the chaperone might serve as an important function in normal healthy liver. This poses a real challenge to knock-down therapy approaches (Brown et al., 2005).

Osteogenesis imperfecta (OI)

As has been anticipated by HSP47 researchers and collagen biologists in general, the first direct connection between HSP47 and osteogenesis imperfecta (OI) was reported in 2009, based on the study on OI Dachshunds dogs (Drögemüller et al., 2009). A mutation at Leu position 326 to Pro (L326P) in HSP47 was found in OI individuals. The amino acid is highly conserved across all vertebrates carrying HSP47 protein. Later on, a similar mutation was also found in human individuals suffering from OI (Christiansen et al., 2010). The abnormal HSP47 is degraded *via* proteasome pathway and the absence of normal HSP47 appears to have several effects on type I procollagen production. Firstly, procollagen trimers are rapidly transported from the ER to the Golgi. Then the overall transit time from the ER to the extracellular space is also increased. The trimers were also found to be hypersensitive to protease degradation suggesting that they are immaturely expressed. These in the end affect the collagen fibril assembly in the ECM and lead to irregular bone mineralisation pertinent to OI (Christiansen et al., 2010).

Connection with Cancer Progression

HSP47 gene expression (also called *SERPINH1*), has been detected in many types and stages of cancer. Recently, a comparative study of public gene expression data of seven types of cancer found that *SERPINH1* is one of the top marker genes for stomach cancer, particularly at the early stage. Collagen type XII gene (*COL12A1*), is also one of the possible markers in the aforementioned cancer type (Xu et al., 2010). Similarly, HSP47 antigen has been detected in sera of breast cancer patients, among 27 other biomarkers (Anderson et al., 2011). The gene has also been shown to be highly upregulated in neuroblastoma tumour cells after treatment with demethylating agent 5'-aza-2'-deoxycytidine (5-aza-dC). This is because hypermethylation of various genes has been reported to be closely-related to NB. In the presence of 5-aza-dC, *SERPINH1* gene expression is induced more than 80-fold thus the growth of neuroblastoma is impaired. It was also found that, in the NB cell lines, the higher the level of collagen expression, the higher the tumorigenicity (Yang et al., 2004). In some cases of carcinomas, HSP47 is highly expressed. For example in the epithelial cells of ulcerative colitis associated carcinoma, that relates to the increase in collagen type I production. The function of collagen from epithelial cells is somewhat unclear (Araki et al., 2009). Overexpression is also observed in oral squamous cell carcinoma induced by areca nut chewing, which is a famous pastime among East, South, and South East Asian communities. This habit has been declared as unhealthy due to the carcinogenicity of the compounds present in the nut. The high expression of HSP47 was found to accompany collagen expression in the tissues surrounding the tumour, probably to prevent metastasis (Lee et al., 2011). Similar overexpression has also reported in head and neck squamous cell carcinoma (Li et al., 2008), and pancreatic carcinoma (Maitra et al., 2002). These findings suggest the growing significance of HSP47 as biomarker and potential target for therapy for various types of cancerous cells.

Potential HSP47-related Treatments

Treatments of various fibrotic diseases by targeting HSP47 mechanism are actively being developed due to its binding specificity to collagen. Among these include small molecule inhibitors and RNA interference (RNAi), using viral and non-viral gene delivery system (Li

et al., 2008; Sato et al., 2008; Thomson et al., 2005). But there has been another treatment method reported to be able to suppress HSP47 expression thus suppressing collagen deposition in fibrotic diseases, namely employing the antisense oligonucleotides. Oligonucleotides are short (13-25 nucleotides) unmodified or chemically modified single stranded DNA molecules that hybridises to a unique sequence in the total pool of targets in cells. Generally, the mechanism of inhibition of gene expression by oligonucleotides can be divided into two: 1. *Via* the RNase H-dependent oligonucleotides, which induces mRNA degradation. 2. The steric-blocker oligonucleotides which prevents or inhibit splicing or translational machinery progression. The majority of drugs employing oligonucleotides are based on the first mechanism (Dias and Stein, 2002). Antisense oligonucleotides has been shown to abrogate the expression of HSP47 in peritoneal fibrosis in rats, thus suppressing the expression of collagen types I and III (Nishino et al., 2003). Later on, the oligonucleotides have been shown to suppress pulmonary fibrosis induced by paraquat (a herbicide) ingestion. The level of HSP47 expression was halved and fibrotic lesions was inhibited based on the difference in connective tissue deposition (Hagiwara et al., 2007c). Furthermore, experiments on a rat model suffering from pulmonary fibrosis induced with bleomycin and lipopolysaccharide was also found to inhibit or decrease HSP47 expression (Hagiwara et al., 2007a; Hagiwara et al., 2007b). Bleomycin is a cytostatic drug commonly used in the treatment of cancer, with an undesirable side effect of inducing chronic pulmonary inflammation that may progress to fibrosis in some cases. Compared to other techniques, oligonucleotides administration does not require gene transfection.

Vector systems on the other hand, also have their own potential. Adenoviral vector targeting HSP47 as tumour-specific biomarker has been developed, and used to target the head and neck squamous cell carcinoma (HNSCC) (Li et al., 2008). Liposomes, as an alternative to viral delivery system, have also been applied to deliver small interference RNA (siRNA) against gp46 (the rat homolog of HSP47). The vitamin A-coupled liposome system carrying the siRNA has been shown to resolve collagen deposition in cirrhotic rat liver by suppressing HSP47 expression (Sato et al., 2008). Hepatic stellate cells, which are the principal cell type responsible for liver fibrosis, were particularly selected as the cells take up vitamin A from circulation before storing it. For this reason, vitamin A was coupled to the surface of liposome to achieve a more specific targeting to the cells, before the siRNA targets gp46 and suppresses its expression. Finally, the use of small molecules is also appealing for

pharmacological targeting of HSP47. Four small molecules have been identified to inhibit HSP47 function *in vitro*, but the efficacy of these compounds in cells and *in vivo* is yet to be tested (Thomson et al., 2005).

Inhibiting excessive collagen synthesis in various fibrotic diseases by blocking HSP47 function is a promising mechanism for the treatment of such diseases. Therefore, it is essential to understand the mechanism by which HSP47 binds to collagen. This will facilitate the development of drug design strategies so that the best treatment method can be devised.

1.4 Serpins

HSP47 is a slightly unusual member of the heat shock proteins as it belongs to the *ser*ine *p*rotease *in*hibitor (serpin) family as well. Serpins generally have a highly conserved secondary structure that is usually made up of a core of three β-sheets surrounded by nine α-helices (Huber and Carrell, 1989). They can be found widely across eukaryotes, in their related viruses and also in higher plants (Gettins, 2002). As a member of the serpin superfamily, HSP47 has been shown to possess several of the archetypal serpin characteristics. Hence, it is helpful to consider the structure and behaviour of the members of this family to assist in studying HSP47 functions.

Serpins are single chain proteins with conserved domains of around 370-390 amino acid residues. Members of this group of proteins have a highly conserved secondary structure. Serpins are characterised by the presence of three groups of β-sheets (A to C) and eight to nine α-helices (hA to hI) (Figure 1.8A, B) named according to the Huber and Carrell nomenclature of α_1-antitrypsin (α_1-AT) (Huber and Carrell, 1989). Many members of this group are serine protease inhibitors, but some are non-inhibitors such as ovalbumin, pigment epithelium derived factor (PEDF), thyroid- and corticosteroid binding globulins (T- and CBG) and angiotensinogen (AGT) (Silverman et al., 2001). In inhibitory serpins (native form), the interaction with their target proteases is at a reactive site within a loop structure (reactive centre loop or reactive serpin loop, RCL or RSL) of about 30-40 amino acids from the *C*-terminus (Figure 1.8A, B). The loop is exposed on the surface making it susceptible to proteolysis by non-target proteases (Potempa et al., 1994). The bond cleaved during

proteolysis is known as the 'scissile bond' which linked residues called P_1 and P'_1 on the RCL. Upon binding to target proteases, the residues *N*-terminal to the scissile bond then insert into β-sheet A forming the 'cleaved' form of serpin (Engh et al., 1989). The protein is said to undergo a 'stressed' to 'relaxed' (S to R) transition which gives rise to an increase in thermostability and loss of inhibitory activity (Figure 1.8C). In the case of α_1- AT, the binding of trypsin protease at the active site causes the protease to be translocated to the distal end from the point of initial docking, the distance of which was suggested to be more than 70 Å (Stratikos and Gettins, 1999).

The native conformation of serpins required for their inhibitory activity consists of a five stranded β-sheet A with the RCL exposed to the environment. A more stable but non-inhibitory conformation, called the 'latent' form can also be achieved *via* the insertion of RCL into the β-sheet A giving a fully antiparallel β-sheet (Figure 1.8D). This conformation explains the inactivity of the latent form as the RCL is inaccessible to the cognate proteases. The structural rearrangement also involves the extraction of the first strand of β-sheet C (s1C). The latent form can be observed in plasminogen activator inhibitor-1 (PAI-1) and antithrombin. In PAI-1, the newly-inserted β-strand forms hydrogen bonds and hydrophobic interactions with the third (s3A) and the fourth strand of β-sheet A (s5A) leading to the hyperthemostability of the latent form (melting temperature of up to 70°C). The native metastable form usually melts around 60°C (Mottonen et al., 1992; Munch et al., 1991).

As more serpins are discovered, their functions are also moving away from the classical serpin function of protease inhibition. In fact the first structure of uncleaved serpin was the structure of a noninhibitory serpin ovalbumin (Stein et al., 1990). Ovalbumin in the most abundant protein found in egg white. In the ovalbumin structure, the peptide loop analogous to the RCL of inhibitory serpin forms a protruding α-helix (Stein et al., 1990). This is particularly important as structures of serpins before that were only available are in cleaved form, especially that of cleaved α_1-AT (Loebermann et al., 1984). The structure of thyroxine-binding globulin (TBG) has also been solved (Zhou et al., 2006). Thyroxine is the major hormone controlling cellular development as well as the rate of body metabolism. It is carried in the blood by TBG, which has a high binding affinity to it. The structure of thyroxine-bound TBG revealed that the binding pocket of thyroxine molecule is formed by strands 3-5 of β-sheet B and helices H and A (Zhou et al., 2006) (Figure 1.9A, B). The pocket provides an enclosed hydrophobic environment, and allows formation of H-bond between thyroxine

and the adjacent residues in the pocket. The reversible release of thyroxine from the pocket is governed by the partial insertion of the side chain of P14, a residue situated 14 residues before the 'scissile' residue P1 of the RCL, into β-sheet A. The release is triggered by the insertion of the RCL up to P12, which causes contraction of the binding pocket, hence the release of thyroxine. What triggers this insertion is still unclear; heat is among the factors proposed. Using the model of δ-ACT as model for TBG with unbound thyroxine, TBG is shown to be in equilibrium between the high-affinity binding and the low affinity forms, distinguished by the partial RCL insertion into β-sheet A (Figure 1.9C) (Zhou et al., 2006).

In another case, the crystal structure of another noninhibitory serpin PEDF shows that at the RCL region, there are Pro residues adjacent to the scissile bond which probably do not facilitate loop insertion. The RCL region of PEDF was also predicted to project outward of the protein body. Another possible explanation of the loss of inhibitory activity of PEDF is the inability of the main β-sheet A to open up and accept the RCL insertion. PEDF is found as a component of the retinal inter-photoreceptor matrix as well as in the vitreous and aqueous liquid in the adult eye (Simonovic et al., 2001).

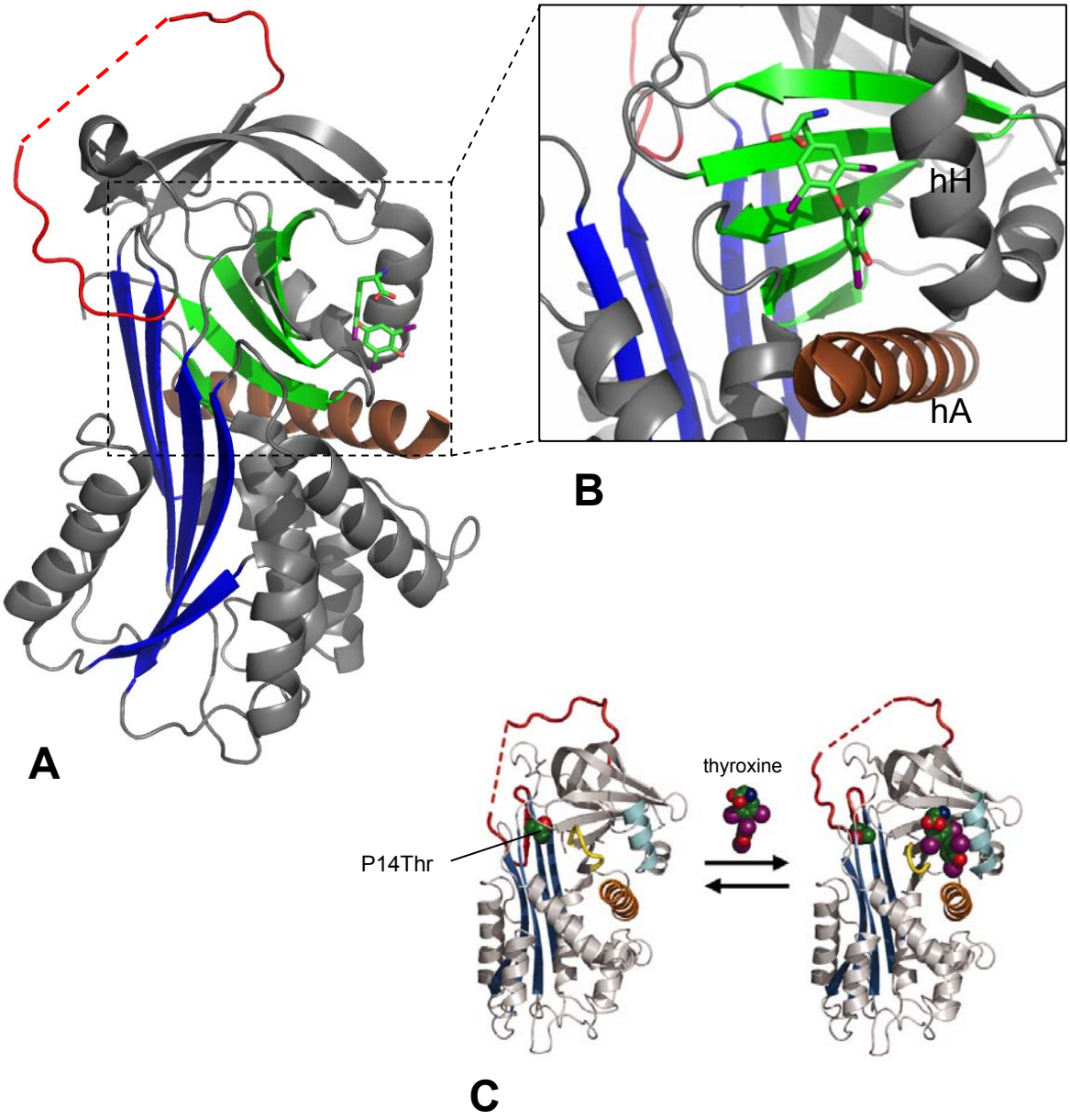


Figure 1.9: The structure of thyroxine-binding globulin (TBG) shown here with bound thyroxine.
A. The structure of TBG, with the main β-sheet A is shown in blue, the β-sheet B that forms the pocket of thyroxine binding site (green) and helix A (hA, brown). The RCL is shown in red, with residues 350-357 of the loop not shown due to the unordered nature of the region. Thyroxine is shown in stick representation. **B**. The binding pocket, enclosed by strands 3-5 of β-sheet B as the base, and helices A and H. **C.** TBG equilibrium between the high-affinity form (right) and the low-affinity form (left). The location for threonine P14 is shown in space-filling representation. All adapted from (Zhou et al., 2006).

1.4.1 HSP47 as a Serpin

HSP47 found in chick embryo fibroblast (CEF) was shown to have the highest homology to the serpin human protein C inhibitor (hPCI, 30% identity) and also a slightly lower sequence identity to the other serpin members (10-30% identity) (Hirayoshi et al., 1991). This is significant as serpins are well known to have a common scaffold of secondary structures. CEF HSP47 is homologous to the serpin family members especially the highly conserved serpin secondary structure, such as part of the serpin RCL region. The serpin characteristics are, some of which are observed in HSP47: (1) small glycoprotein with a single polypeptide chain and several oligosaccharide side chains (though the sugar group was later found not to be crucial for HSP47 function), (2) homologous amino acid sequences, (3) having a common secondary and probably tertiary structure, (4) possessing a conserved reactive centre loop (RCL) and (5) containing conserved flanking sequences around the RCL. The amino acids around the region homologous to serpin RCL of HSP47 resemble more to the sequence found in ovalbumin (a non-inhibitory serpin) and CEF HSP47 also lacks the reactive P_1 bond. This prompted the suggestion that HSP47 is another noninhibitory member of serpin superfamily (Hirayoshi et al., 1991).

In the same year, a HSP47 homologue found in rat skeletal myoblasts (known as gp46 at that time) shows sequence similarities to human α_1-AT and hPCI precursors with about 30% of identical amino acid residues, unsurprisingly similar to that found in CEF. The even distribution of conserved residues suggests similar secondary structure which is syncronous with that observed with CEF HSP47. Rat gp46 shows sequence similarity to α_1-AT and hPCI especially at the residues near the putative RCL cleavage site P_1-P'_1. But unlike the inhibitory serpins, in gp46 residues upstream to the site possess larger side chains, another finding leading to suggestion of HSP47 noninhibitory function as large side chains makes loop insertion into β-sheet A less preferable (Clarke et al., 1991).

It is now well established that HSP47 is a noninhibitory member of the serpin superfamily with a substantial amount of secondary and tertiary structural similarities. This is very helpful due to the lack of HSP47 crystal structure to date. As more serpin crystal structures are obtained, the information could be used to assist the homology modelling of HSP47.

1.4.2 HSP47 Homology Modelling

Three dimensional structures of proteins provide the best information in order to investigate the mechanism and chemistry at atomic resolution. Two methods have been extensively utilised for protein structural determination; X-ray crystallography for crystallisable proteins, and nuclear magnetic resonance (NMR) spectroscopy for non crystallisable proteins and for determination of protein conformation in solution. X-ray is suitable for structure determination of proteins too large to be analysed by NMR. NMR, on the other hand, allows investigation at a condition closer to the native environment of proteins (Mathews et al., 1999). Originally used for structure determination of proteins in aqueous environment, NMR has been actively developed for use for membrane proteins that dissolve in detergent micelles or lipid bilayers (Marassi and Opella, 1998). The use of both techniques to analyse one protein rarely gives conflicting outcome.

Homology modelling has also been used as routine structure 'determination' as the speed at which gene sequences are discovered is too rapid for structural biologists to cope in solving the three dimensional coordinate of the coded proteins (Rodriguez et al., 1998). Additionally, for proteins that do not meet the criteria for, or *resist* X-ray crystallography and NMR spectroscopy, homology modelling is the best (or only) option available to guide biological experiments. These include site-directed mutagenesis and development of protein inhibitor based on structural information. To get a homology model of a considerable accuracy, templates of known protein structures closely-related to the protein of unknown structure must be present. These structures are usually obtained from online databases, namely the RCSB Protein Data Bank governed by the Research Collaboratory for Structural Bioinformatics (http://www.pdb.org) (Berman et al., 2000). As of December 2010, the database holds more than 70,000 of biomolecular structures determined using X-ray crystallography, NMR spectroscopy and cryo-electron microscopy. Then together with the sequence data of the protein to be determined, many programs can be used to construct the model, be it stand-alone or web-based. Among the stand-alone modelling programs are MODELLER (http://www.salilab.org/modeller/), DeepView (Swiss Pdb-Viewer) (http://spdbv.vital-it.ch/) and YASARA (http://www.yasara.org/) while web-based programs are like 3D-JIGSAW (http://bmm.cancerresearchuk.org/~3djigsaw/), SWISS-MODEL (http://swissmodel.expasy.org/) and ESyPred3D

(http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/). DeepView and SWISS-MODEL both use the same server at Expert Protein Analysis System (ExPASy) of the Swiss Institute of Bioinformatics (http://expasy.org/). 3D-JIGSAW has recently been used to model mutant HSP47 protein that has contributed to osteogenesis imperfecta (OI) in Dachshunds (Drögemüller et al., 2009). The mutation was found to be at Leu326 and from nucleotide sequence of the affected individuals, it is substituted with Pro. This could significantly affect the wild-type peptide backbone, thus leading to disruption in the local secondary structural conformation (Drögemüller et al., 2009). The effect of the mutation on the collagen binding and release function of HSP47 however was not discussed in the report.

Due to the nature of HSP47, from our observations, crystallisation has been found to be difficult as it tends to aggregate in the concentration range usually required for crystallisation. Generally, in order to obtain a reliable homology model, an identity between the target protein and the template must first be determined. If the percentage identity falls below 30%, model quality estimation on the basis of sequence identity becomes unreliable (Bordoli et al., 2009). But as HSP47 belongs to the serpin superfamily that have highly conserved secondary structures, this makes molecular modelling an acceptable approach with about 30% identity with its closest serpin relatives. The first attempt to model HSP47 was made in 1995 using mature recombinant mouse (mrm) protein (Davids et al., 1995) using the known structure of hPCI. It was initially predicted to exist in the latent state at physiological pH, as HSP47 failed to inhibit a series of proteases. It is unlikely that latent conformation is adopted as latency is known to increase the thermostability of serpins (Munch et al., 1991), and HSP47 is later shown to be metastable (Dafforn et al., 2001). The work paved the way towards finding a better model to suit HSP47 behaviour. Later, HSP47 was modelled as RCL-exposed, five-stranded β-sheet A, similar to that of α_1-AT (Dafforn et al., 2001) despite the exact position of RCL region relative to overall structure is still being debated. A similar model was also presented to explain the effect of mutation on HSP47 structure found in OI individuals (Drögemüller et al., 2009).

Chapter 2

Results and Discussion: Investigation on the Important Histidine Residues for HSP47 pH-switch Function

CHAPTER 2: INVESTIGATION ON THE IMPORTANT HISTIDINE RESIDUES FOR HSP47 pH-SWITCH FUNCTION

2.1 Introduction

This chapter presents thorough investigation on a series of His-to-Ala (HA) mutants. This first involved computational prediction of all surface residues in order to examine the contributions of histidine residues to HSP47 pH-change behaviour *in vitro*. The prediction of surface His residues was performed on a mouse HSP47 homology model using PropKa program, and the positions were mapped to predict their locations. Each of the residues was subjected to site-directed mutagenesis. A few adjacent residues were mutated simultaneously. Expression and solubility of the mutants were first assessed and the general binding behaviour to gelatin was assayed. Gelatin is the denatured form of collagen, containing both triple helical and single stranded molecules. Soluble and gelatin-binding mutants were then subjected to gelatin column purification in order to investigate their pH release behaviour. It was necessary that the properties of HSP47 WT were established first so comparison with those of HA mutants can be made in order to probe for possible perturbations. Using circular dichroism (CD), the secondary structural content of HSP47 was investigated prior to studying the effects of pH on HSP47 secondary structure. Furthermore, collagen mimetic peptides (CMP) were employed to demonstrate how changes in pH affect the change in HSP47 structure, which in turn triggers the release from collagen.

Similarly, the effects of mutation on HSP47 structure were also investigated by CD. Then the ability of the mutants to arrest collagen fibrillisation was examined. This provided information on the effects of mutation on HSP47-collagen binding interface. Finally, the effects on HSP47 structural transition from alkali to acid state were probed by comparing the change in structure of the constructed mutants to that of WT. Data obtained from these experiments would provide the insights of how His residues participate actively in HSP47 pH-switch mechanism, how mutations perturbs normal pH transition as well as information on collagen binding and release behaviour.

2.2 Homology Model Construction of Mouse HSP47

Section 1.4.2 has described the advantages of homology modelling for proteins in cases where X-ray or NMR structures are absent, mostly due to the protein itself resisting crystallisation. Although there are certain limitations to when working with homology models, templates of known protein structures closely-related to the protein being studied can be used to get a homology model of a considerable accuracy. The main depository is the RCSB Protein Data Bank, where more than 70,000 protein structures have been deposited. A 30% sequence similarity is generally taken as a bench mark in getting a good homology model (Schwede et al., 2008).

Molecular modelling of mouse HSP47 has been reported previously (Davids et al., 1995). It involves alignment of the HSP47 sequence with the one of a known crystal structure of a serpin family member, human protein C inhibitor (hPCI). The study also confirmed the non-protease inhibitory activity of HSP47 despite having high homology to serpin members (around 30% identity). The model also seems slightly inaccurate due to the fact that HSP47 was predicted to exist in latent state at physiological pH values, while a latent serpin is hyperthermostable (with a melting temperature of around 70°C). This behaviour is not observed in mouse HSP47 (Dafforn et al., 2001; Dafforn et al., 2004). HSP47 is then proposed to adopt a conformation similar to the native structure of serpins.

Hence, construction of a more reliable model is needed to guide the mutagenesis experiments. As with most homology modelling, a native serpin scaffold is needed. There have been quite a number of serpin crystal structures reported, such as the inhibitory serpins α_1-antitrypsin (Elliott et al., 1998), antithrombin (Schreuder et al., 1994) and a variant of antichymotrypsin (Wei et al., 1994); as well as the non-inhibitory serpins such as ovalbumin, the first serpin structure found in the native form (Stein et al., 1990) and pigment epithelium-derived factor (PEDF) (Simonovic et al., 2001). Figure 2.1 presents the alignment of the sequence of a mature recombinant mouse HSP47 with some closely-related serpins. The alignment scores are all in the small range of 23-27%.

The first serpin structure to be solved was that of a cleaved α_1-AT (Loebermann et al., 1984). The cleaved structure was reported before the native structure was solved. Research interest on this serpin has been growing dramatically since. It was even selected as a

structural archetype from which other serpins can be modelled. The nomenclature of serpin secondary structures now follows that of α_1-AT (Huber and Carrell, 1989) making this serpin a suitable template for HSP47 homology model construction, with a good sequence similarity (25% pairwise alignment score with mrmHSP47) compared to the other serpins aligned.

The mouse HSP47 homology model was provided by the Nagata group at Kyoto University on a collaboratory basis (Figure 2.1), constructed using Deepview v4.0.1 modelling software (Guex and Peitsch, 1997). α_1-AT was selected as template structure. The model was then used to guide the mutagenesis experiments and all HSP47 figures shown in this chapter and the following chapters are made using this model.

```
1BY7            ----------------------MEDLCVANTLFALNLFKHLAKASPTQNLFLSPWSISST 38
1OVA            ----------------------XGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSA 38
1QLP            MDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNSTNIFFSPVSIATA 60
2CEO            --------------------GSLYKMSSINADFAFNLYRRFTVETPDKNIFFSPVSISAA 40
mrmHSP47        MAEVKKPLEAAAPGTAEKLSSKATTLAERSTGLAFSLYQAMAKDQAVENILLSPLVVASS 60
                                    :      :.:.::: :       *:: .*  : ::

1BY7            MAMVYMGSRGSTEDQMAKVLQFNEVGAAAD----------KIHSSFRSLSSAINASTG-N 87
1OVA            LAMVYLGAKDSTRTQINKVVRFDKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPND-V 97
1QLP            FAMLSLGTKADTHDEILEGLNFNLTEIPEA----------QIHEGFQELLRTLNQPDS-Q 109
2CEO            LVMLSFGACCSTQTEIVETLGFNLTDTPMV----------EIQHGFQHLICSLNFPKK-E 89
mrmHSP47        LGLVSLGGKATTASQAKAVLSAEKLRDEEV----------HTGLGELLRSLSNSTARN 108
                : :: :*    *  :    :  :                   : .:  :   :. .

1BY7            YLLESVNKLFGEKSASFREEYIRLCQKYYSSEPQAVDFLECAEEARKKINSWVKTQTKGK 147
1OVA            YSFSLASRLYAEERYPILPEYLQCVKELYRGGLEPINFQTAADQARELINSWVESQTNGI 157
1QLP            LQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFTVNFGD-TEEAKKQINDYVEKGTQGK 168
2CEO            LELQIGNALFIGKHLKPLAKFLNDVKTLYETEVFSTDFSN-ISAAKQEINSHVEMQTKGK 148
mrmHSP47        VTWKLGSRLYGPSSVSFADDFVRSSKQHYNCEHSKINFRD-KRSALQSINEWASQTTDGK 167
                      . *:  .      .::.  :  *       :*      * : **. ..  *.*

1BY7            IPNLLPEGSVDGDTRMVLVNAVYFKGKWKTPFEKKLNGLYP-FRVNSAQRTPVQMMYLRE 206
1OVA            IRNVLQPSSVDSQTAMVLVNAIVFKGLWEKAFKDEDTQAMP-FRVTEQESKPVQMMYQIG 216
1QLP            IVDLVK--ELDRDTVFALVNYIFFKGKWERPFEVKDTEEED-FHVDQVTTVKVPMMKRLG 225
2CEO            VVGLIQ--DLKPNTIMVLVNYIHFKAQWANPFDPSKTEDSSSFLIDKTTTVQVPMMHQME 206
mrmHSP47        LPEVTK--DVERTDGALLVNAMFFKPHWDEKFHHKMVDNRG-FMVTRSYTVGVTMMHRTG 224
                :  :    .:.      *** : **  *   *. .       * :       * **

1BY7            KLNIGYIEDLKAQILELPYA-GDVSMFLLLPDEIADVSTGLELLESEITYDKLNKWTSKD 265
1OVA            LFRVASMASEKMKILELPFASGTMSMLVLLPDEVS----GLEQLESIINFEKLTEWTSSN 272
1QLP            MFNIQHCKKLSSWVLLMKYL-GNATAIFFLPDEGK-----LQHLENELTHDIITKFLE-- 277
2CEO            QYYHLVDMELNCTVLQMDYS-KNALALFVLPKEGQ-----MESVEAAMSSKTLKKWNR-- 258
mrmHSP47        LYNYYDDEKEKLQMVEMPLAHKLSSLIILMPHHVEP----LERLEKLLTKEQLKAWMG-- 278
                        . .  :: :         :..:*..       :: :*  :. . :. :

1BY7            KMAEDEVEVYIPQFKLEEHYELRSILRSMGMEDAFNKGRANFSGMSERNDLFLSEVFHQA 325
1OVA            VMEERKIKVYLPRMKMEEKYNLTSVLMAMGITDVFSSS-ANLSGISSAESLKISQAVHAA 331
1QLP            NEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNG-ADLSGVTEEAPLKLSKAVHKA 336
2CEO            LLQKGWVDLFVPKFSISATYDLGATLLKMGIQHAYSEN-ADFSGLTEDNGLKLSNAAHKA 317
mrmHSP47        KMQKKAVAISLPKGVVEVTHDLQKHLAGLGLTEAIDKNKADLSRMSGKKDLYLASVFHAT 338
                   .    : :*:  :   ::*   *  :*: .. ... *::* ::    * ::.. * :

1BY7            MVDVNEEGTEAAAGTGGVMTGRTGH--GGPQFVADHPFLFLIMHKITNCILFFGRFSSP- 382
1OVA            HAEINEAGREVVGSAEAGVDAAS----VSEEFRADHPFLFCIKHIATNAVLFFGRCVSP- 386
1QLP            VLTIDEKGTEAAGAMFLEAIPMS----IPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPT 392
2CEO            VLHIGEKGTEAAAVPEVELSDQPENTFLHPIIQIDRSFMLLILERSTRSILFLGKVVNPT 377
mrmHSP47        AFEWDTEGNPFDQDIYGREELRS-----PKLFYADHPFIFLVRDNQSGSLLFIGRLVRPK 393
                    .  *              .        .  ::.*:: : .  : . **:*:   *

1BY7            --------
1OVA            --------
1QLP            QK------ 394
2CEO            EA------ 379
mrmHSP47        GDKMRDEL 401
```

Figure 2.1: Multiple sequence alignment of mature recombinant mouse HSP47 (mrmHSP47) with several closely-related serpins.
Among the serpins selected for alignment is human PAI-1 (PDB ID: 1BY7), ovalbumin (1OVA), α_1-AT (1QLP) and a non-inhibitory serpin thyroxine-binding globulin (2CEO). All amino acid sequences are obtained from Protein Data Bank (http://www.rcsb.org/pdb). "*" means that the residues are identical in all sequences in the alignment; ":" means that conserved substitutions are observed; "." means that semi-conserved substitutions are observed, according to common amino acid side chain grouping (Thompson et al., 1994). The alignment score between mrmHSP47 and PAI-1 is 27%, 26% (ovalbumin), 25% (α_1-AT) and 23% (TBG).

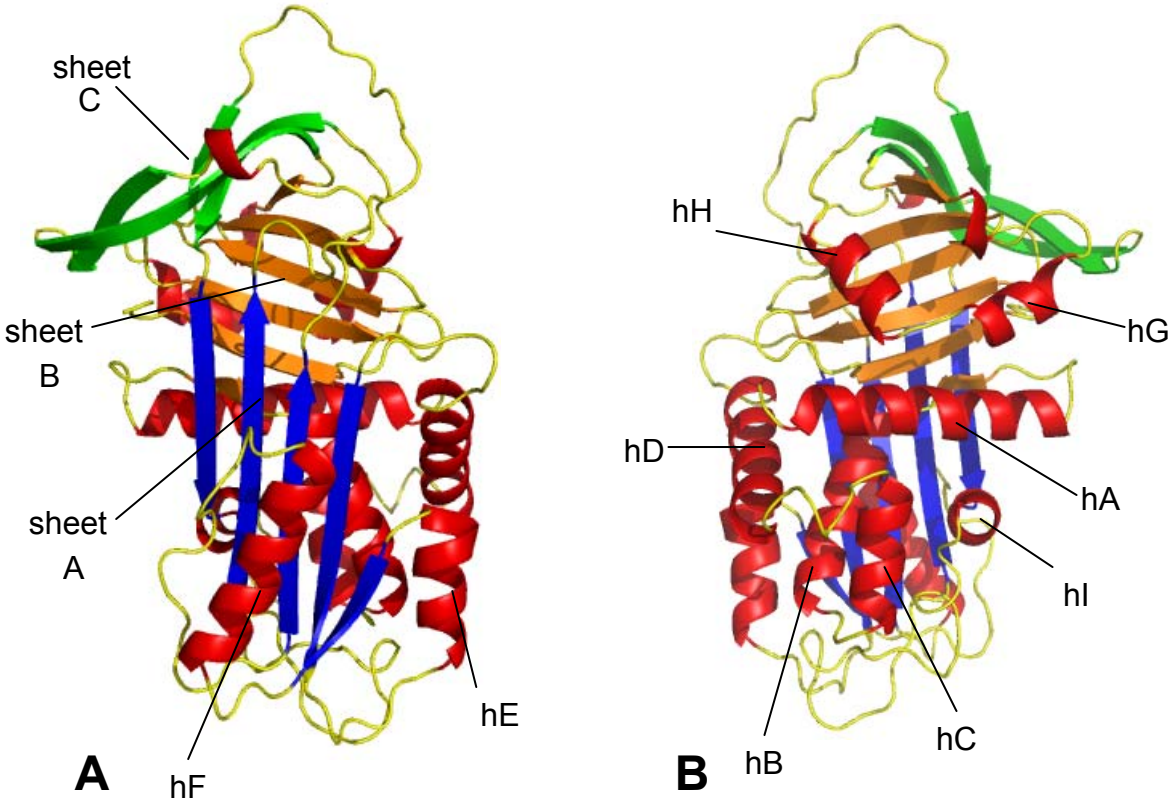


Figure 2.2: Ribbon structure representations of a homology model of HSP47, using α_1-antitrypin structure (see Figure 1.7) as reference.
A. The three main β-sheets are labelled sheet A-C and the α-helices are labelled hA-hI. **B**. 'Back' view of the homology model. These figures and all subsequent molecular figures were produced using PyMol v0.99 (DeLano, 2004) (unless stated otherwise).

2.3 Predicting the Solvent Accessibility and Locations of all Histidine (His, H) Residues

Mouse HSP47 has fourteen His residues altogether (Figure 2.3). Mutating all these fourteen could possibly affect the conformational integrity in case of buried residues. Hence we searched for a good criterion for mutagenesis by covering as much important sites as possible. For this purpose, the location and solvent accessibility of the residues were mapped on the homology model that was constructed earlier. Figure 2.4 shows their predicted locations on the model. To obtain a better prediction on accessibility and ionisation pH values of the residues, we used PropKa software available at http://propka.ki.ku.dk (Li et al., 2005). It is a FORTRAN-based tool that allows calculation of the pK_a values of ionisable groups in proteins based on specific empirical rules.

It has been reported that the predicted pK_a generally falls within 1 pH value to experimentally-determined pK_a in 44 proteins (Li et al., 2005). The values for His residues however must be used with caution, as based on calculations, the relative mean standard deviation (RMSD) of experimental/predicted pK_a for His residues are relatively high compared to other residues (Li et al., 2005).

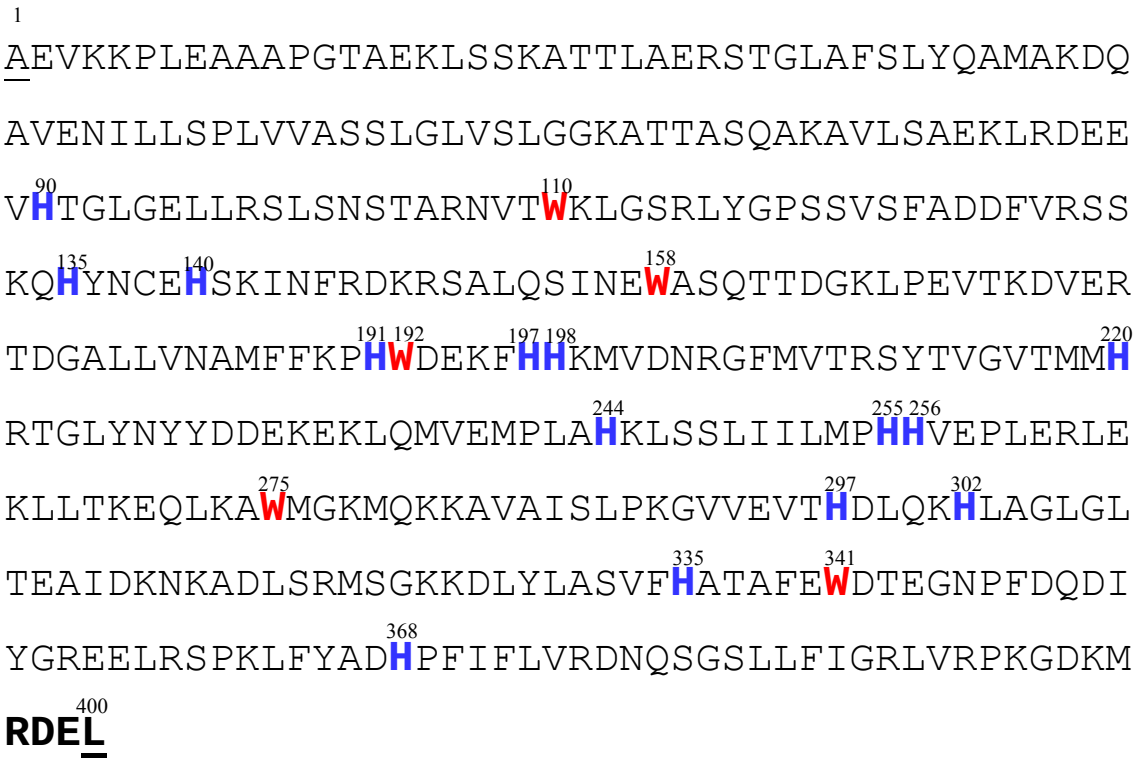


Figure 2.3: Mature mouse HSP47 amino acid sequence (Ala1 to Leu400), shown using one-letter amino acid codes.
The RDEL ER-retention sequence is shown in bold. The His residues are highlighted in blue, and the five Trp residues are highlighted in red.

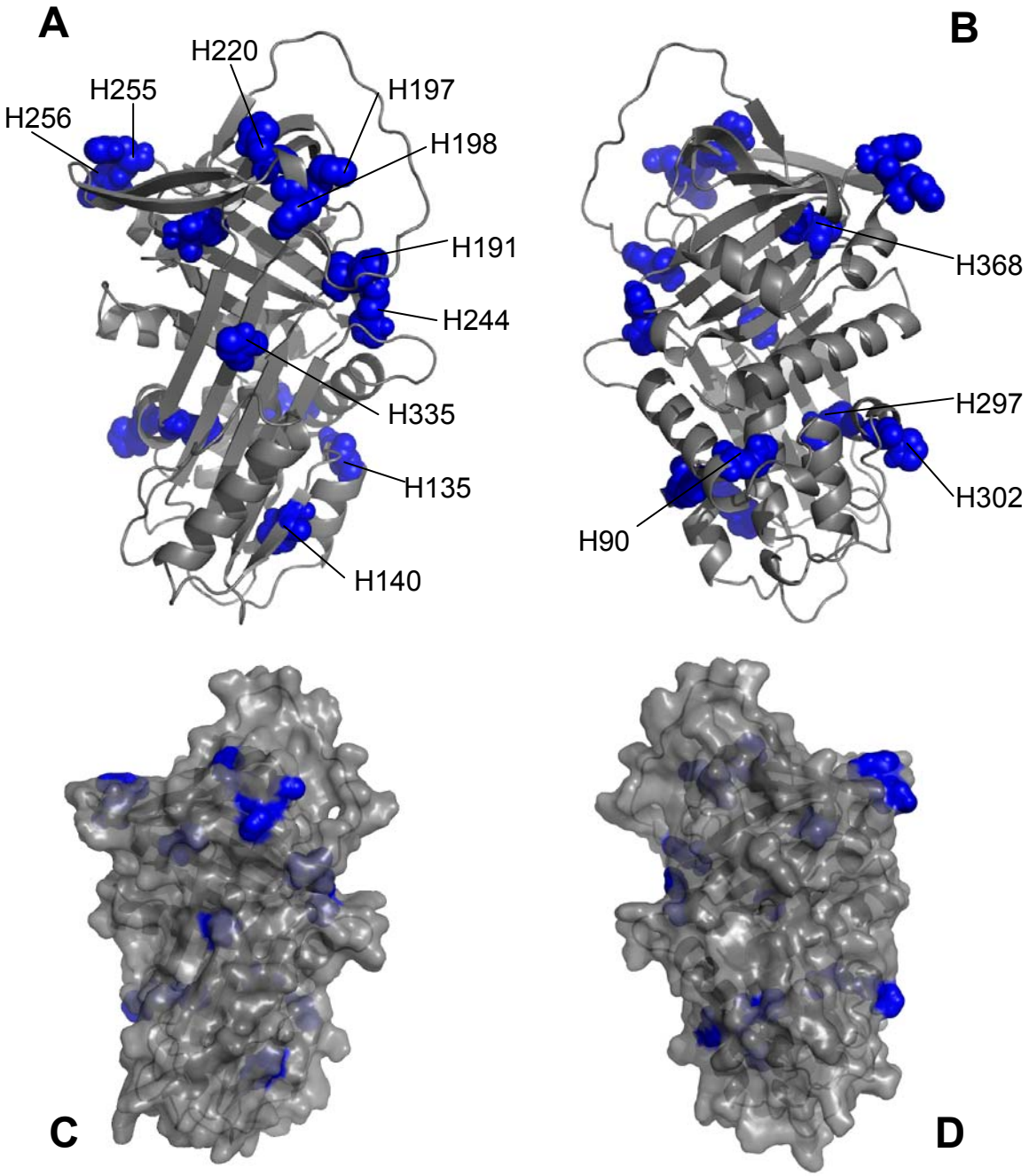


Figure 2.4: The predicted locations of all the His residues (blue) shown here against the overall structure of mouse HSP47 (grey).
A. The 'front' representation of the homology model. **B**. The 'back' representation. One letter amino acid name is used, and the numbers that follow the residue names indicate the amino acid position in HSP47 primary sequence. **C & D.** Surface representation to show the solvent accessibility of all His residues in their respective environments.

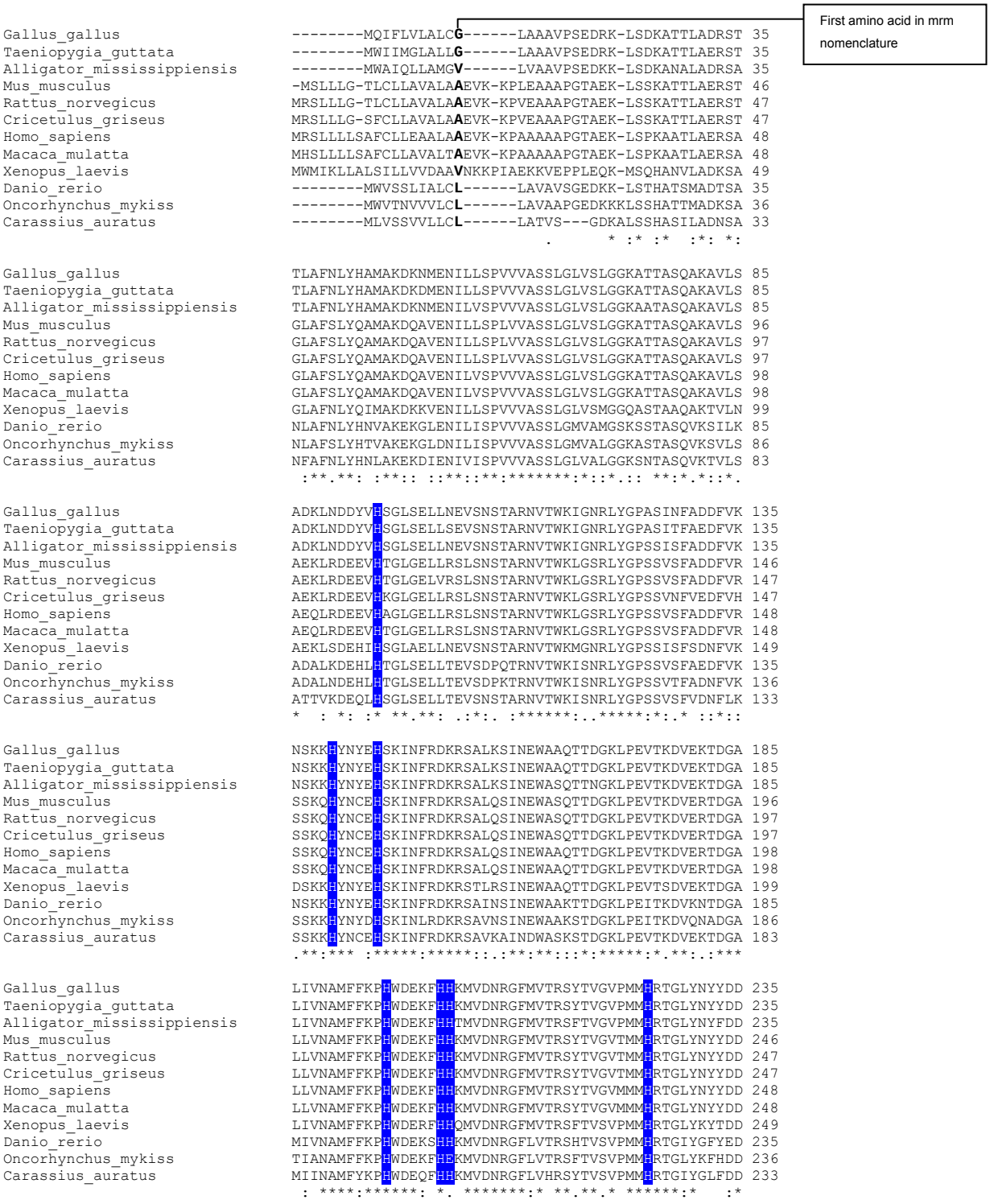
Table 2.1 shows the predicted pK_a values for all His residues, along with the modelled solvent accessibility and the degree of conservation in HSP47 across all reported species (Figure 2.4). There are eight residues altogether that are predicted to be ionised around the HSP47-collagen dissociation pH of 6.3 (Saga et al., 1987). The residues are H135, 191, 197, 198, 244, 255, 256 and 302. They are also highly conserved (more than 90% identity) across all 12 species, excluding H255 which is only 46% identical (Figure 2.4). More detailed observation on the locations of these residues revealed that they can generally be grouped into three clusters. The first cluster, on top of β-sheet A is called the 'breach' cluster as it is close to the breach region of serpin. This region is very critical for serpin function as it is an insertion point for reactive centre loop (RCL) into sheet A following binding to protease substrate (Whisstock et al., 2000). The second cluster is called the 'shutter' cluster, which is found near the centre of sheet A centred at the start of helix B (Stein and Carrell, 1995). This domain is important in facilitating the opening of strands 3 of β-sheet A (s3A) and s5A to accept RCL insertion. The last cluster is the 'gate' cluster, borrowing from the nomenclature of a flexible loop connecting s3C and s4C, which is critical in the conversion of PAI-1 from native to latent structure (Stein and Carrell, 1995). His residues grouped into these clusters are shown in Table 2.2. The locations of these clusters in HSP47 homology model are shown in Figure 2.6.

All eight surface residues were chosen as mutagenesis candidates to investigate the effects of mutation on HSP47 pH-switch function.

Table 2.1: The solvent accessibility of fourteen His residues in mouse HSP47, their predicted pK_as and the degree of conservation across 12 species analysed.
Solvent accessibility and pK_a values were predicted using PropKa* program. His residues with predicted pK_a values close to HSP47-collagen dissociation pH are italicised.

Residue	Solvent accessibility	Predicted pK_a*	Degree of conservation (%)
H90	buried	4.45	100
H135	***surface***	***6.35***	***100***
H140	buried	3.28	100
H191	***surface***	***6.36***	***100***
H197	***surface***	***6.36***	***100***
H198	***surface***	***7.23***	***92***
H220	buried	4.76	100
H244	***surface***	***6.15***	***92***
H255	***surface***	***6.50***	***46***
H256	***surface***	***6.44***	***100***
H297	buried	0.23	100
H302	***surface***	***6.51***	***92***
H335	buried	3.41	100
H368	buried	1.08	100

*error of about 1 pH unit to experimentally-determined proteins (Li et al., 2005).


First amino acid in mrm nomenclature
Gallus_gallus --------MQIFLVLALCG------LAAAVPSEDRK-LSDKATTLADRST 35
Taeniopygia_guttata --------MWIIMGLALLG------LAAAVPSEDRK-LSDKATTLAERST 35
Alligator_mississippiensis --------MWAIQLLAMGV------LVAAVPSEDKK-LSDKANALADRSA 35
Mus_musculus -MSLLLG-TLCLLAVALAAEVK-KPLEAAAPGTAEK-LSSKATTLAERST 46
Rattus_norvegicus MRSLLLG-TLCLLAVALAAEVK-KPVEAAAPGTAEK-LSSKATTLAERST 47
Cricetulus_griseus MRSLLLG-SFCLLAVALAAEVK-KPVEAAAPGTAEK-LSSKATTLAERST 47
Homo_sapiens MRSLLLLSAFCLLEAALAAEVK-KPAAAAAPGTAEK-LSPKAATLAERSA 48
Macaca_mulatta MHSLLLLSAFCLLAVALTAEVK-KPAAAAAPGTAEK-LSPKAATLAERSA 48
Xenopus_laevis MWMIKLLALSILLVVDAAVNKKPIAEKKVEPPLEQK-MSQHANVLADKSA 49
Danio_rerio --------MWVSSLIALCL------LAVAVSGEDKK-LSTHATSMADTSA 35
Oncorhynchus_mykiss --------MWVTNVVVLCL------LAVAAPGEDKKKLSSHATTMADKSA 36
Carassius_auratus --------MLVSSVVLLCL------LATVS---GDKALSSHASILADNSA 33
TLAFNLYHAMAKDKNMENILLSPVVVASSLGLVSLGGKATTASQAKAVLS 85
TLAFNLYHAMAKDKDMENILLSPVVVASSLGLVSLGGKATTASQAKAVLS 85
TLAFNLYHAMAKDKNMENILVSPVVVASSLGLVSLGGKAATASQAKAVLS 85
GLAFSLYQAMAKDQAVENILLSPLVVASSLGLVSLGGKATTASQAKAVLS 96
GLAFSLYQAMAKDQAVENILLSPLVVASSLGLVSLGGKATTASQAKAVLS 97
GLAFSLYQAMAKDQAVENILLSPLVVASSLGLVSLGGKATTASQAKAVLS 97
GLAFSLYQAMAKDQAVENILVSPVVVASSLGLVSLGGKATTASQAKAVLS 98
GLAFSLYQAMAKDQAVENILVSPVVVASSLGLVSLGGKATTASQAKAVLS 98
GLAFNLYQIMAKDKKVENILLSPVVVASSLGLVSMGGQASTAAQAKTVLN 99
NLAFNLYHNVAKEKGLENILISPVVVASSLGMVAMGSKSSTASQVKSILK 85
NLAFSLYHTVAKEKGLDNILISPVVVASSLGMVALGGKASTASQVKSVLS 86
NFAFNLYHNLAKEKDIENIVISPVVVASSLGLVALGGKSNTASQVKTVLS 83
ADKLNDDYVHSGLSELLNEVSNSTARNVTWKIGNRLYGPASINFADDFVK 135
ADKLNDDYVHSGLSELLSEVSNSTARNVTWKIGNRLYGPASITFAEDFVK 135
ADKLNDDYVHSGLSELLNEVSNSTARNVTWKIGNRLYGPSSISFADDFVK 135
AEKLRDEEVHTGLGELLRSLSNSTARNVTWKLGSRLYGPSSVSFADDFVR 146
AEKLRDEEVHTGLGELVRSLSNSTARNVTWKLGSRLYGPSSVSFADDFVR 147
AEKLRDEEVHKGLGELLRSLSNSTARNVTWKLGSRLYGPSSVNFVEDFVH 147
AEQLRDEEVHAGLGELLRSLSNSTARNVTWKLGSRLYGPSSVSFADDFVR 148
AEQLRDEEVHTGLGELLRSLSNSTARNVTWKLGSRLYGPSSVSFADDFVR 148
AEKLSDEHIHSGLAELLNEVSNSTARNVTWKMGNRLYGPSSISFSDNFVK 149
ADALKDEHLHTGLSELLTEVSDPQTRNVTWKISNRLYGPSSVSFAEDFVK 135
ADALNDEHLHTGLSELLTEVSDPKTRNVTWKISNRLYGPSSVTFADNFVK 136
ATTVKDEQLHSGLSELLTEVSNSTARNVTWKISNRLYGPSSVSFVDNFLK 133
NSKKHYNYEHSKINFRDKRSALKSINEWAAQTTDGKLPEVTKDVEKTDGA 185
NSKKHYNYEHSKINFRDKRSALKSINEWAAQTTDGKLPEVTKDVEKTDGA 185
NSKKHYNYEHSKINFRDKRSALKSINEWASQTTNGKLPEVTKDVEKTDGA 185
SSKQHYNCEHSKINFRDKRSALQSINEWASQTTDGKLPEVTKDVERTDGA 196
SSKQHYNCEHSKINFRDKRSALQSINEWASQTTDGKLPEVTKDVERTDGA 197
SSKQHYNCEHSKINFRDKRSALQSINEWASQTTDGKLPEVTKDVERTDGA 197
SSKQHYNCEHSKINFRDKRSALQSINEWAAQTTDGKLPEVTKDVERTDGA 198
SSKQHYNCEHSKINFRDKRSALQSINEWAAQTTDGKLPEVTKDVERTDGA 198
DSKKHYNYEHSKINFRDKRSTLRSINEWAAQTTDGKLPEVTSDVEKTDGA 199
NSKKHYNYEHSKINFRDKRSAINSINEWAAKTTDGKLPEITKDVKNTDGA 185
SSKKHYNYDHSKINLRDKRSAVNSINEWAAKSTDGKLPEITKDVQNADGA 186
SSKKHYNCEHSKINFRDKRSAVKAINDWASKSTDGKLPEVTKDVEKTDGA 183
LIVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVPMMHRTGLYNYYDD 235
LIVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVPMMHRTGLYNYYDD 235
LIVNAMFFKPHWDEKFHHTMVDNRGFMVTRSFTVGVPMMHRTGLYNYFDD 235
LLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTMMHRTGLYNYYDD 246
LLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTMMHRTGLYNYYDD 247
LLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVTMMHRTGLYNYYDD 247
LLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVMMMHRTGLYNYYDD 248
LLVNAMFFKPHWDEKFHHKMVDNRGFMVTRSYTVGVPMMHRTGLYNYYDD 248
LIVNAMFFKPHWDERFHHQMVDNRGFMVTRSFTVSVPMMHRTGLYKYTDD 249
MIVNAMFFKPHWDEKSHHKMVDNRGFLVTRSHTVSVPMMHRTGIYGFYED 235
TIANAMFFKPHWDEKFHEKMVDNRGFLVTRSFTVSVPMMHRTGLYKFHDD 236
MIINAMFYKPHWDEQFHHKMVDNRGFLVHRSYTVSVPMMHRTGIYGLFDD 233

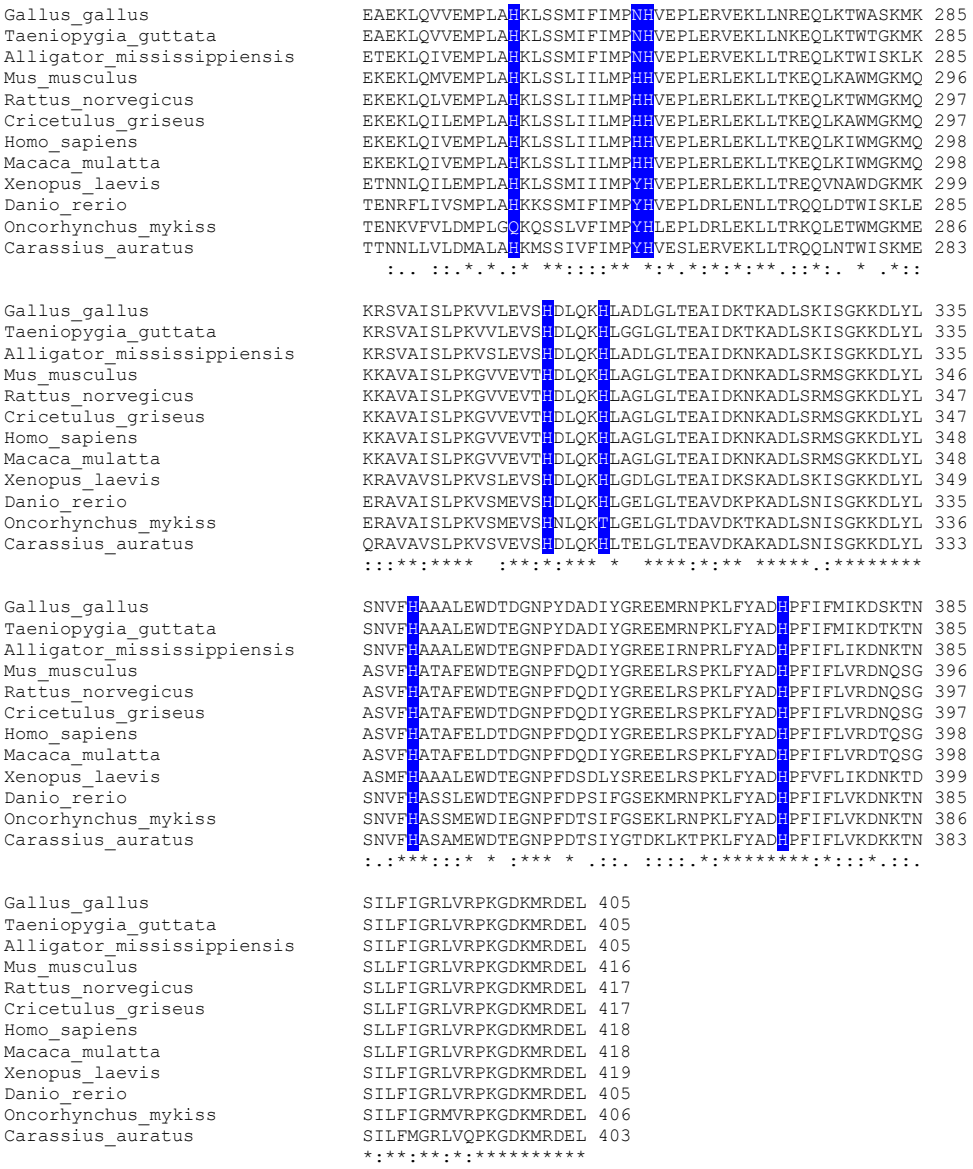


Figure 2.5: Sequence conservation of His residues (highlighted in blue) across 12 species known to carry *hsp47* gene.
All amino acid sequences are obtained from Protein database on NCBI (http://www.ncbi.nlm.nih.gov/protein). "*" means that the residues are identical in all sequences in the alignment; ":" means that conserved substitutions are observed; "." means that semi-conserved substitutions are observed, according to common amino acid side chain grouping (Thompson et al., 1994). The species investigated are *Alligator Mississippiensis* (American alligator), *Carassius auratus* (goldfish), *Cricetulus griseus* (Chinese hamster), *Danio rerio* (zebrafish), *Gallus gallus* (chicken), *Homo sapiens* (human), *Macaca mulatta* (rhesus monkey), *Mus musculus* (mouse), *Oncorhynchus mykiss* (rainbow trout), *Rattus norvegicus* (Norway rat), *Taeniopygia guttata* (zebra finch), and *Xenopus laevis* (African clawed frog).

Table 2.2: HSP47 surface histidine residues grouped into three main clusters.

Clusters	Residues
Breach	H191, H197, H198, H244
Gate	H255, H256
Shutter	H302

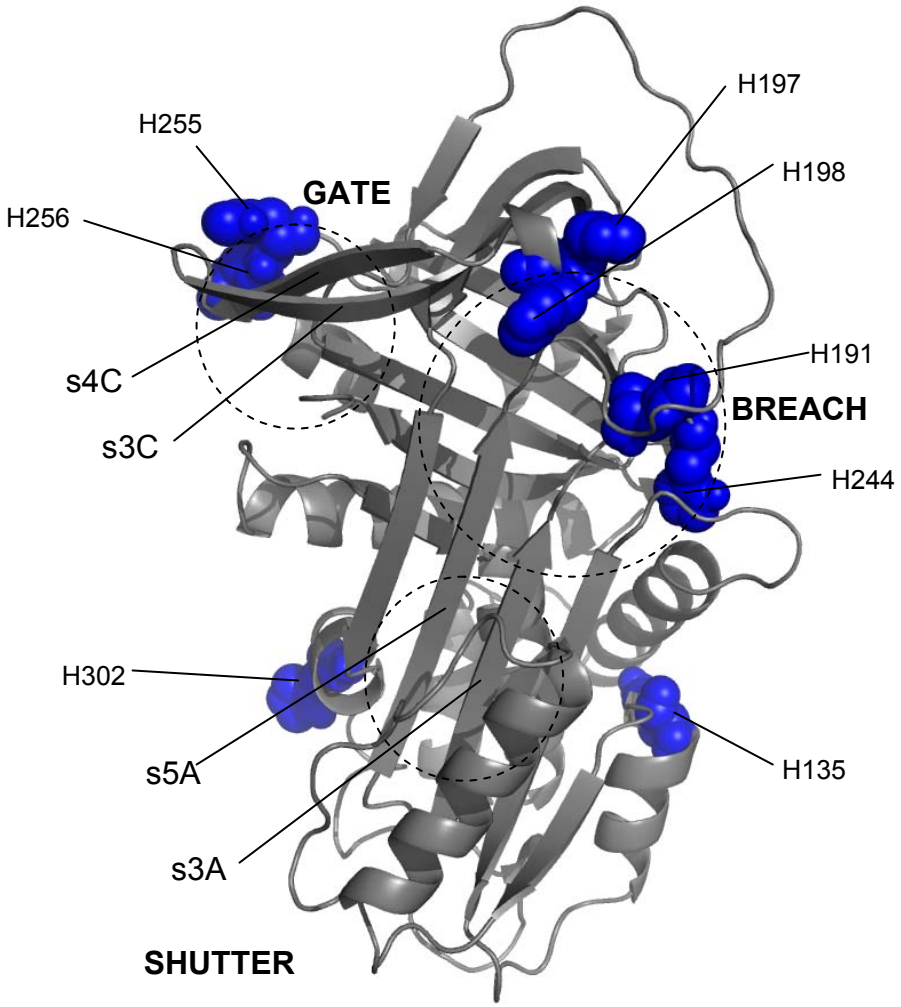


Figure 2.6: The locations of 'breach', 'gate' and 'shutter' cluster in HSP47 homology model.
His residues predicted to be on the surface of the molecule is highlighted in blue and shown in 'spheres' representation. Figure produced with PyMol v0.99.

2.4 Construction of HSP47 Wild-type (WT) Plasmid DNA

A mature recombinant mouse *hsp47* gene was amplified by PCR using a 5′ primer with an *Nde*I restriction site and a 3′ primer with a *Hin*dIII restriction site and cloned into the equivalent restriction sites in the expression vector pET-24b(+). This work was done previously in our laboratory at the Imperial College Genetic Therapies Centre by Dr Dee Olerenshaw and Dr Takayuki Homma. Further details of the recombinant plasmid are given in Chapter 6 Materials and Methods.

2.5 Mutation, Expression and Collagen Binding Ability of HSP47 WT and His Mutants

All His residues of interest, selected based on computational prediction, were mutated to alanine (Ala, A). This allows assessing the effects of mutation on pH-switch mechanism simultaneously with the possible structural contribution of a replaced His residue to the overall structure. His-to-Ala (HA) mutations were introduced by modified site-directed mutagenesis (Sawano and Miyawaki, 2000). Each construct was sequenced to screen the correct mutation and full sequencing was performed to confirm that there are no other mutations in the region between the T7 promoter and terminator. Most of the mutants were found to be equally soluble compared to HSP47 WT when expressed in *E. coli* host strain BL21(DE3)pLysS (Figure 2.6A). HSP47 is known for being a highly insoluble protein when expressed in high concentration under normal *E. coli* growth and expression conditions, from our experience. In order to overcome this, temperature of the main growth period was decreased to 20-25°C (from 37°C which was only used for pre-culture) while IPTG concentration was also lowered to 0.1 mM from the normal 1 mM. This succesfully ensured higher solubility of the expressed proteins and reduced the presence of inclusion bodies, as can be seen in Figure 2.7A.

Batch binding experiment showed that the binding behaviour of the mutants to gelatin beads is not significantly impaired (Figure 2.7B). This then allows the use of gelatin column to observe the collagen release behaviour and subsequent purification of the mutants for further biophysical experiments.

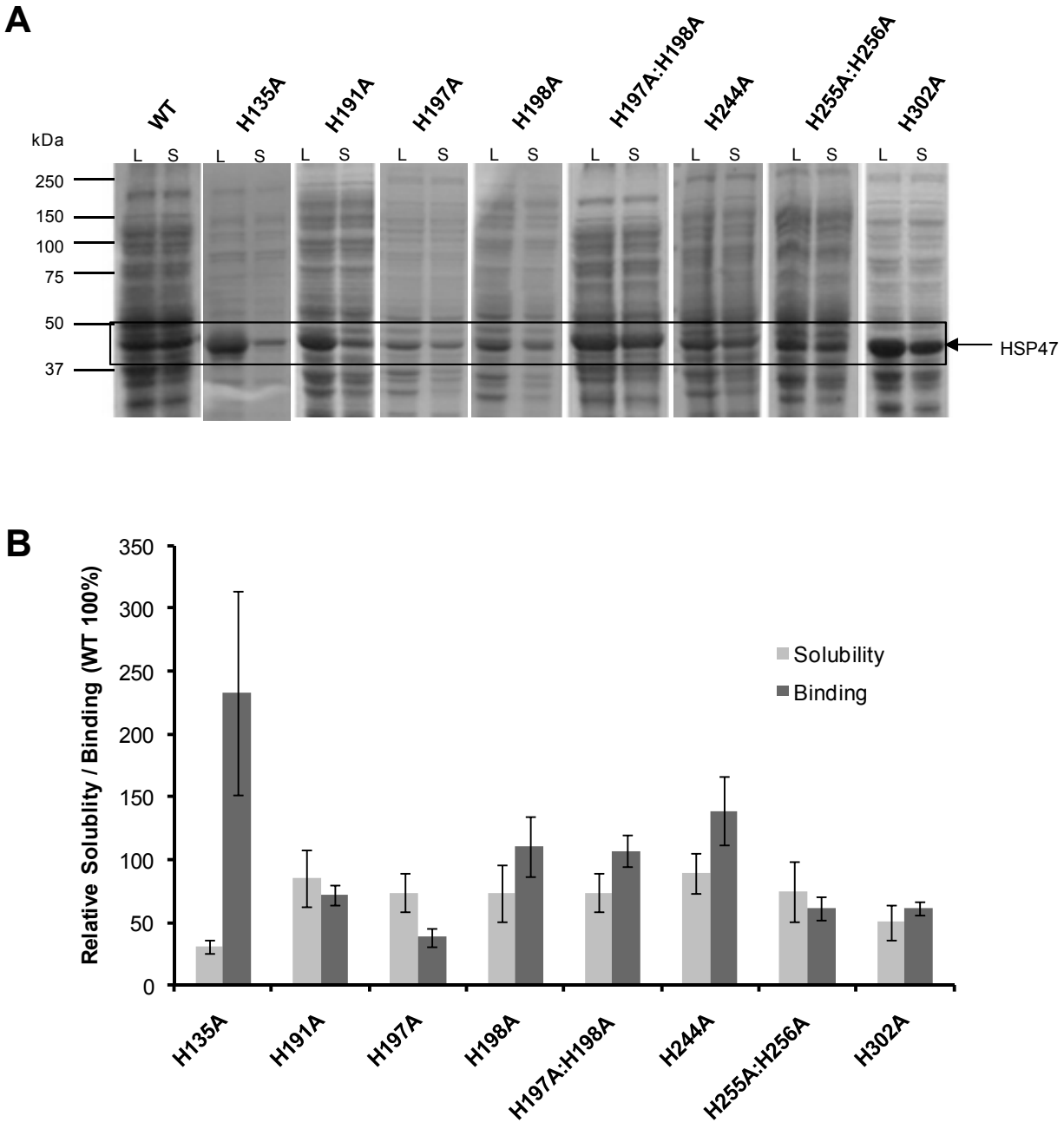


Figure 2.7: Expression, solubility and binding ability of all HA mutants constructed, relative to WT.
A. Expression level of HSP47 WT and HA mutants. Separation was done on 12% Tris-glycine SDS-PAGE. L: Cell lysate, S: Centrifuged lysate (supernatant). **B.** Binding ability of the mutants, relative to WT (100%, n=3) determined using gelatin pull down assay.

2.6 Purification of HSP47 WT and HA Mutants

Gelatin-agarose was used as column packing matrix since mouse HSP47 WT and HA mutants were shown to be able to bind to gelatin from the batch gelatin binding assay (Figure 2.7B). The column was connected to an FPLC system for quick, single step purification. Collagen/gelatin affinity techniques have been widely used for HSP47 purification (Dafforn et al., 2001; El-Thaher et al., 1996; Nishikawa et al., 2010) due to the aforementioned advantage. Purification of HSP47 WT gave a distinctive peak during pH gradient elution with sodium phosphate buffer from pH 8.0 to 4.1 in about 15 column volumes (Figure 2.8). Fractions were collected in the presence of 1 M of sodium phosphate pH 8.0 to immediately recover the alkaline pH at which HSP47 is in its native form. SDS-PAGE analysis with CBB detection indicates that the eluted protein is more than 95% pure and suitable for direct biophysical analysis without further purification (Figure 2.9). Typical yields are between 0.3-0.5 mg/mL.

Mutants H135A, H191A, H244A, H255A:H256A and H302A also had a similar peak to WT during elution (Figure 2.10). Although the concentration of peak fractions varies due to the slight difference in expression level and solubility, all the mutants are above 95% in purity after purification as estimated from SDS-PAGE analysis. A more thorough investigation also revealed that the location of H191A elution peak is shifted to slightly higher pH value. The measured pH value of the peak fraction is also slightly higher than that of WT and the other mutants (*i.e.* pH 5.9-6.1), instead of around 5.7. This suggests that the release mechanism of HSP47 from gelatin is affected when His residue at position 191 is replaced with Ala, even though the column binding ability seems unchanged.

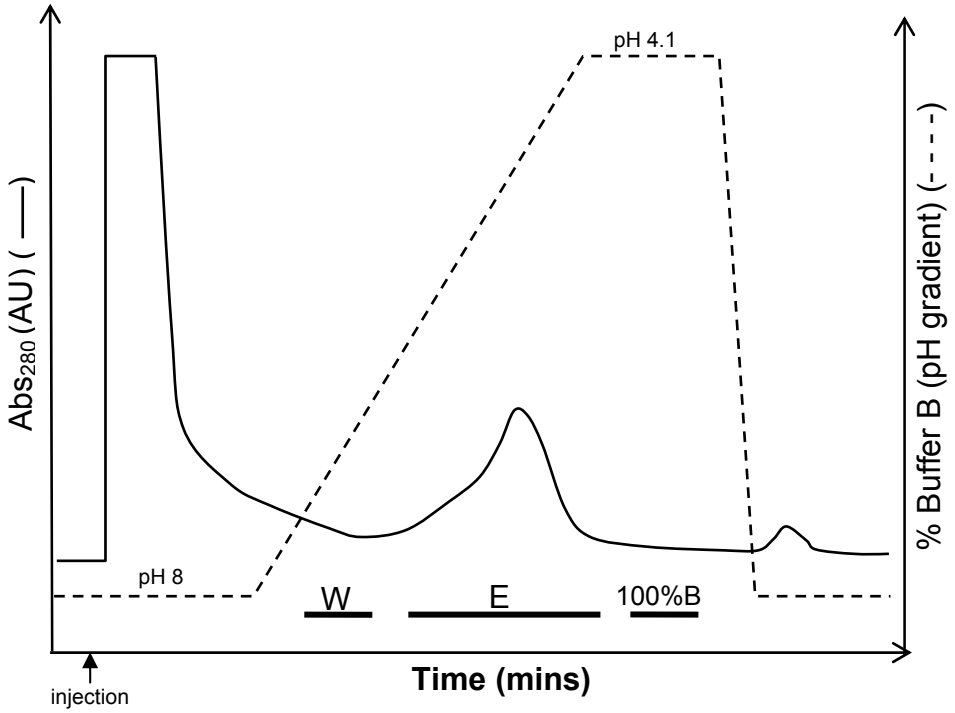


Figure 2.8: FPLC purification trace (reproduction) of HSP47 WT using a gelatin-agarose affinity column.
Dotted line represents pH gradient from pH 8.0 to 4.1. Fractions at different stages of purification (black bars) were analysed with Coomassie Brilliant Blue (CBB) staining after separation on 12% Tris-glycine SDS-PAGE.

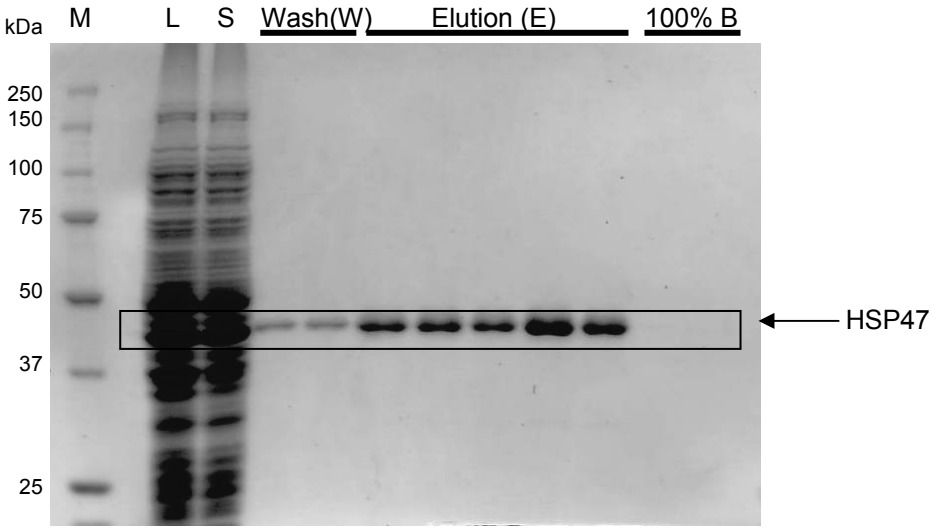


Figure 2.9: 12% Tris-glycine SDS-PAGE analysis of HSP47 WT.
Lane M: Precision Plus Protein Standard (Bio-rad). L: Cell lysate, S: Soluble fraction (supernatant).

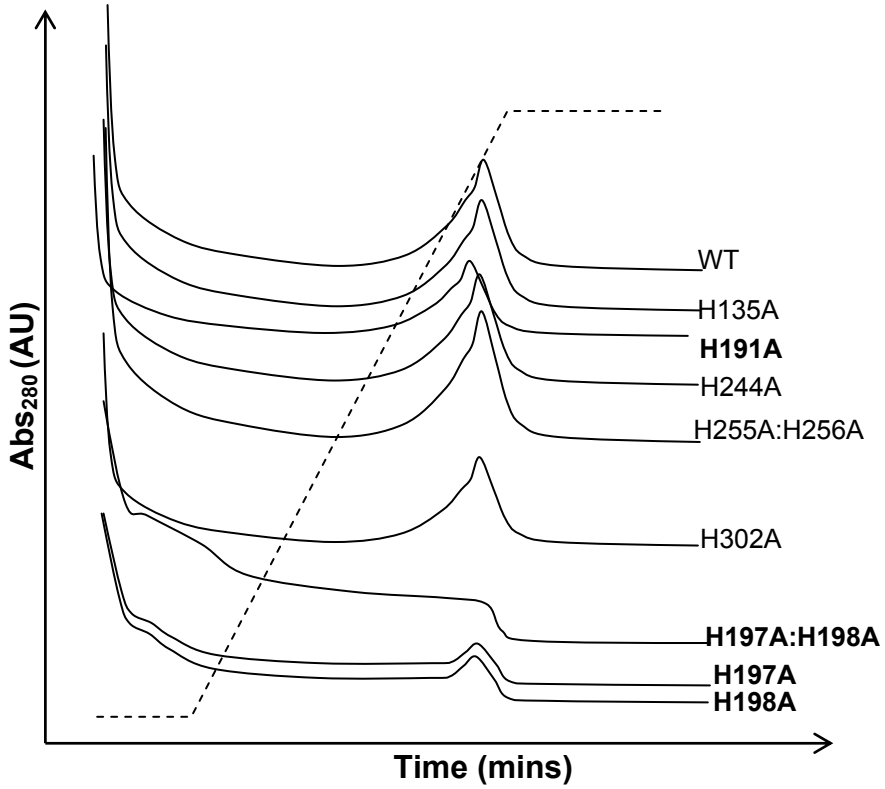


Figure 2.10: Reproduction of the FPLC traces of all HA mutants in comparison to HSP47 WT.
Mutants of interest are highlighted in bold. Dotted line represents buffer pH gradient from 8.0 to 4.1.

On the other hand, mutants H197A and H198A showed an interesting behaviour. A significantly reduced elution peak was recorded during gelatin-agarose purification (Figure 2.11), despite good expression and solubility. In addition, during washing and elution, the mutants seem to be slowly eluted off the column. Replacing both residues (H197A:H198A double mutant) amplifies this effect. SDS-PAGE analysis of the eluted fractions (tail fractions) of the double mutant shows the presence of high concentration of HSP47 (Figure 2.12), at high purity. Unlike H191A mutation that perturbs HSP47 release behaviour, mutations of His at positions 197 and/or 198 instead alter the column binding behaviour, weakening the ability of the mutants to withstand extensive washing. Further downstream, complete elution occurs when the elution pH value reaches the region of normal dissociation value. These suggest that there are other His residues that are ionised and triggers the complete release from gelatin column.

These observations suggest that replacement of His residues at these positions perturbs the gelatin binding behaviour in such a way that the HSP47-collagen interactions (investigated here using gelatin) are no longer strong and specific, especially in the column experiment. It is possible that minor structural alteration causes disruption in the strong binding affinity. The effect of His to Ala mutation on other proteins has been reported before, for example on the receptor-associated protein (RAP) to probe its pH-sensitive binding behaviour toward the low density lipoprotein receptor (LDLR) family (Lee et al., 2006). RAP is an ER-resident protein, with a mode of action similar to that of HSP47 chaperone during collagen synthesis and transport to the Golgi. Another example is the substitution of His 369 with Ala in bleomycin hydrolase (BH), which changes the hydrogen bonding network and causes a complete loss in catalytic activity (O'Farrell and Joshua-tor, 2007). This demonstrates that HA mutation could affect the binding specificity between a protein and its ligand. In HSP47, this is possible especially if H197 and H198 are located at the HSP47-collagen binding interface.

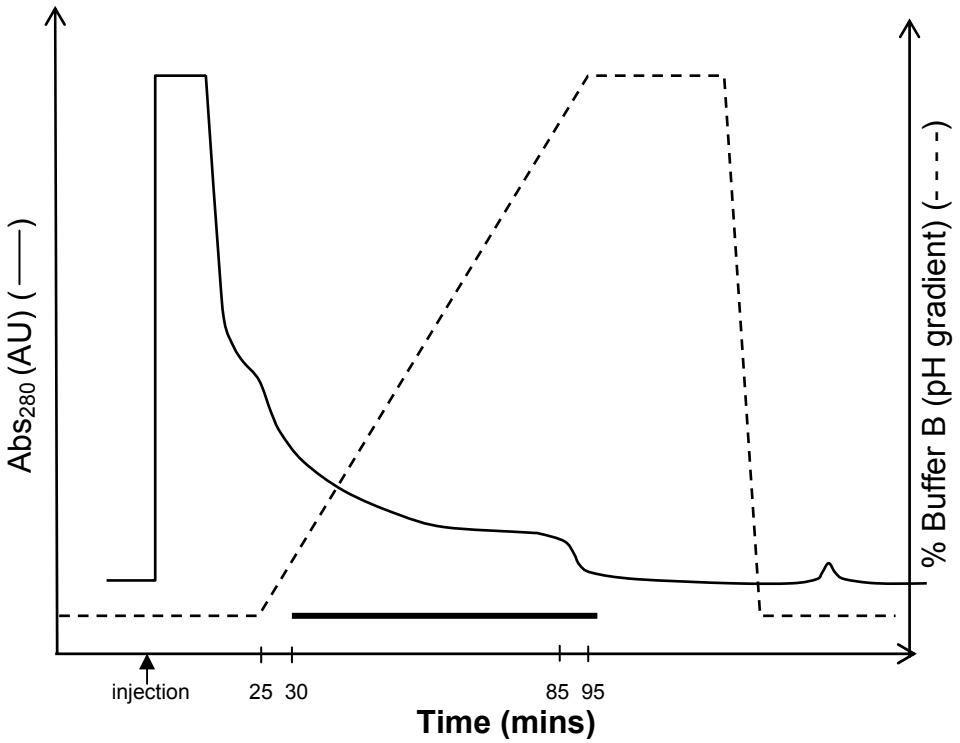


Figure 2.11: Reproduction of the FPLC trace of H197A:H198A through gelatin-agarose column.
Black bar indicates interval where fractions were collected for SDS-PAGE analysis. Dotted line indicates buffer B gradient. Fractions numbers given are approximate.

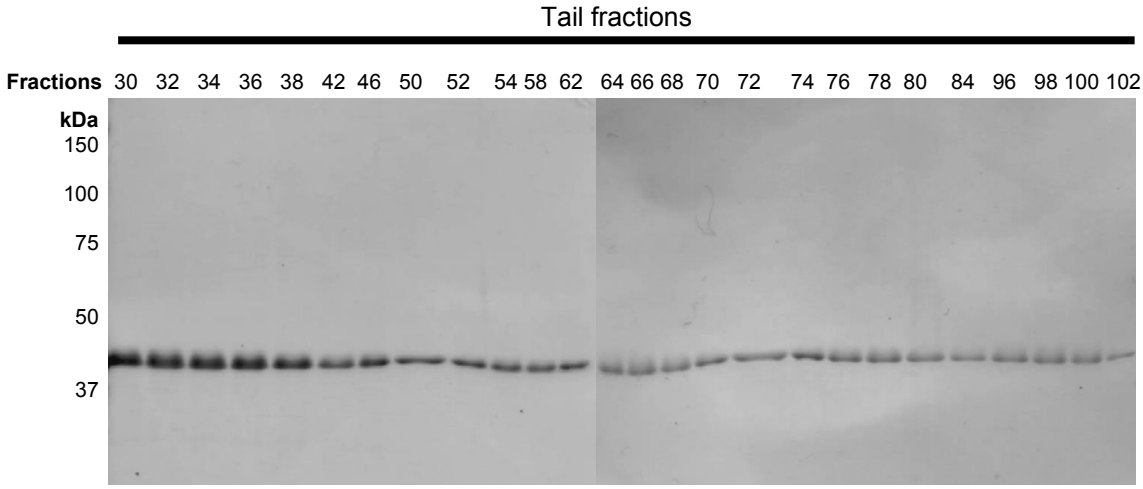


Figure 2.12: 12% Tris-glycine SDS-PAGE analysis of H197A:H198A through gelatin-agarose column.
Approximate fraction numbers are also given for comparison with the FPLC trace.

2.7 Investigation of the Biophysical Properties of HSP47 WT

In this section, the biophysical properties of HSP47 WT are first established. Characterisations of HSP47 have been reported previously using biophysical techniques such as circular dichroism and fluorescence spectroscopy. So to enable comparison with the constructed mutants, the properties of the recombinant mouse HSP47 expressed in pET-24b(+) vector were analysed and are initially presented. Then the results for HSP47 HA mutants are discussed, by comparing with the properties of that of WT.

Circular Dichroism Structural Analysis

CD was utilised to assess the conformation of the mutants expressed in comparison to that of WT (presented in Section 2.10). It would highlight any potential distortion in the protein structure due to the abolition of salt bridge between the respective residues and their neighbours. In human prolactin (hPRL), three His residues held together by aromatic stacking have been previously suggested to contribute to the biological function (Tettamanzi et al., 2008). hPRL is a 23 kDa endocrine hormone and a member of cytokine family. It was found that mutation of one His residue seems to perturb the pK_a of the other, as observed with nuclear magnetic resonance (NMR) spectroscopy (Tettamanzi et al., 2008). After NMR, CD is the next best option to obtain indirect information about possible interactions between the residues.

CD spectrum of HSP47 WT revealed that it is a typical α/β-protein (Figure 2.13). The stability of the purified protein was then assessed using temperature titration by monitoring the change in CD signal at 222 nm ($\Delta\Delta A_{222}$) as a function of temperature. That particular wavelength was chosen as it provides information on the status of the α-helical portions of the protein during the melt and also a longer wavelength that would eliminate the effect caused by changes in light scatter that can occur during thermal denaturation. As the temperature increases, there will be a loss in secondary structure of the protein (Dafforn et al., 2004). Other than detecting stability, the melt profile can also be used in serpin studies to suggest potential transition from native to latent structure. Although this has never been reported before for HSP47, as a member of serpin family with highly conserved secondary

structure, this behaviour could not be ruled out completely. HSP47 WT shows a transitional midpoint (T_m) of around 59°C (± 0.5), estimated using the 'tangent method' and using the equation described in Section 6.15 (Persikov et al., 2004). In 'tangent method', the point at which transition starts to occur and the end-point are traced and estimated from the fully folded and fully unfolded state. In cases where these are difficult to trace (especially the end-point due to poorer readings at high temperature), an assumption has to be made where an end-point is reached when the changes in CD signal is not more than 500 MRE unit. There is a possibility of introducing large errors into the midpoint determination as this is not a standard method, but this could be minimised with many repeat experiments, or by overlaying several curves together for direct comparison.

The presence of such transition at lower temperature are attributed to the melting of a typical native form of serpin family members, of which the β-sheet A is composed of only five strands with the sixth strand exposed, forming the serpin loop (Dafforn et al., 2001). Latent form melts at a much higher temperature (Dafforn et al., 2004).

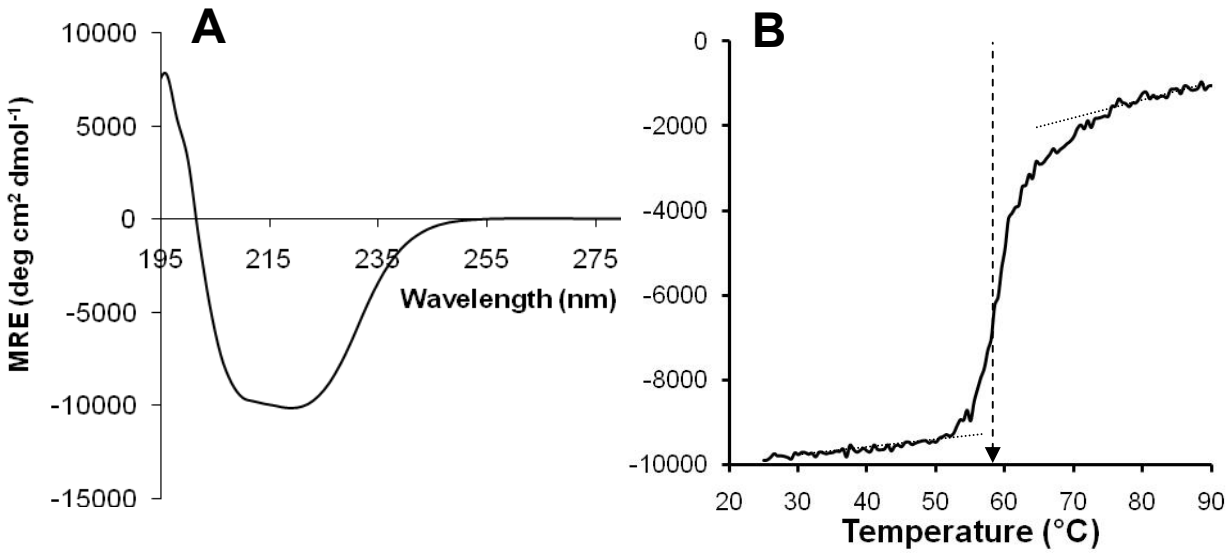


Figure 2.13: Far-UV CD spectrum (A) and thermal denaturation profile (B) of HSP47 WT.

Concentration was fixed to 3 μM and measured in 50 mM sodium phosphate buffer pH 8.0 with 100 mM NaCl. Dashed arrow in (B) represents the predicted melting point. Thermal denaturation is a non-reversible process as aggregation could be observed in the CD cuvette after titration. Dotted lines represent the estimated start and end of transition, used in the 'tangent' method of melting point estimation.

Collagen Anti-fibrillisation Property

Collagen triple helices are known to self-interact at optimum condition (neutral pH, temperature of around 30°C) to form a higher order structure called the collagen fibril. This process, like the self-assembly in proteins, is an entropy-driven process which involves the loss of solvent molecules from the surface of protein molecules (Kadler et al., 1996). It is regarded as involving two processes, the first of which is fibril nucleation, whereby soluble collagen molecules aggregate forming nuclei. Then it is followed by the growth of the nuclei into fibrils by accumulation of further soluble collagen particles (Wood, 1960). As a result, the molecules form fibrils with circular cross-section, which has minimum surface area by volume ratio. *In vivo,* cleavage of the globular propeptides at the amino- and carboxy- termini by aminoprotease and carboxyprotease triggers the fibrillogenesis process before being stabilised by intra- and intermolecular chemical cross-linking (Section 1.1.2). The most abundant collagen types in vertebrates (type I, II, III, V and XI) form fibrils with *D*-period (the distance between cross striations of collagen triple helical molecule in staggered arrangement) between 65-67 nm, which gives collagen fibrils their quintessential axial periodicity (Kadler et al., 1996) defined by the 'hole' and 'overlap' zones as illustrated in Figure 2.14.

In vitro fibrillisation experiment is usually condicted using acid soluble collagen, as mild acid solution typically yields more collagen during preparation *via* extraction from collagenous tissues. The fibrillisation process involves a lag phase followed by a propagation phase and finally an equilibrium phase where the fibrils are in equilibrium with the monomers (Kadler et al., 1988; Thomson and Ananthanarayanan, 2000). Collagen type I fibrillisation is typically studied at 34°C because the rate of fibril propagation decreases above 37°C in the presence of micro-unfolding monomers. These monomers undergo reversible unfolding and become a larger fraction of the monomer and fibril system (Kadler et al., 1988). Fibril formation is observed by monitoring turbidity spectrophotometrically at 313 nm, which is proportional to the fibrillar material formed (Williams et al., 1978).

The turbidity plotted as a function of time for type I collagen solution shows an expected fibrillisation curve with three apparent phases (Figure 2.15). Fibrillisation was initiated by bringing the pH of the solution from acidic (pH 3) to near neutral (pH 7.4), at

which fibrillisation occurred immediately and monitoring was recorded at 10-minute intervals until equilibrium phase was reached. The lag phase observed was about 10 to 20 minutes, double from that reported previously (Thomson and Ananthanarayanan, 2000). This may be due to the collagen used in this case; pepsin treated collagen type I is known to form shorter fibrils over a longer period of time than untreated collagen (Gelman et al., 1979). Pepsin is usually used in the manufacturing process to increase collagen solubility.

HSP47 is known to have the ability to arrest collagen fibril formation *in vitro* at neutral pH. This activity is abolished at HSP47-collagen dissociation pH, reportedly around pH 6.2-6.3 (Thomson and Ananthanarayanan, 2000). As the efficiency of HSP47 to arrest fibrillisation relates to the functional integrity of HSP47 collagen binding interface, the ability of HSP47 WT to arrest fibrillisation was first investigated. The finding could then be used to observe the effect of HA mutation of HSP47 ability to prevent fibril formation.

When HSP47 WT was included at an equal molar ratio to collagen, fibrillisation was completely inhibited with no turbidity observed over the duration of the assay (Figure 2.15). Inhibition at 1:2 (HSP47:collagen) molar ratio was incomplete owing to insufficient capture of collagen by HSP47 WT and fibrillisation was observed to proceed at a slower rate, and hence slight turbidity was recorded near minute 110. This inhibition is specifically due to the prevention of the interaction between collagen triple helix molecules by HSP47. A similar experiment conducted using heat-inactivated HSP47 did not show complete inhibition of fibril formation, unlike with active chaperone molecules (data not shown).

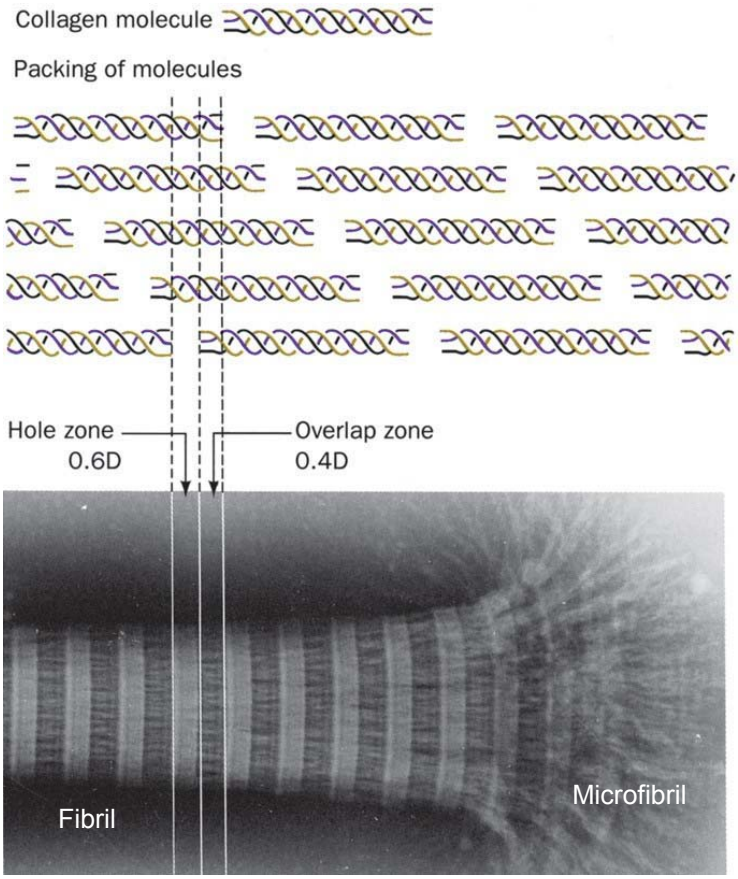


Figure 2.14: Schematic representation of the staggered structure of a collagen fibril, to illustrate the *D*-spacing.
The *D*-spacing is the distance between cross striations ('hole' and 'overlap'), with each collagen molecule is about 300 nm long (4.4 *D*-units). The 'hole' is formed at the end of one triple helix and the beginning of the next (Prockop et al., 1979). Figure adapted from (Voet and Voet, 1995).

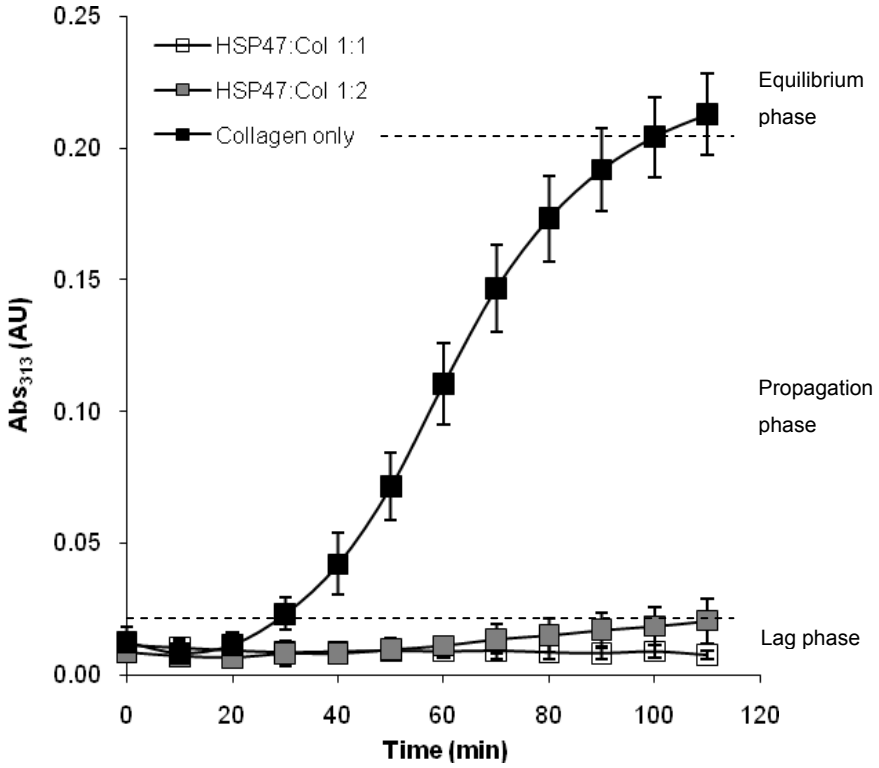


Figure 2.15: The fibrillisation curve of collagen solution upon incubation at 34°C, pH 7.3-7.4 at 2 µM concentration (■).
Between 0-20 mins, the solution is said to be in lag phase, 20-100 mins (propagation phase) and 100- mins (equilibrium phase). Upon addition of HSP47 at the same molar ratio, fibril formation is completely inhibited (□) at 1:1 (HSP47:collagen). Incomplete inhibition is observed in the presence of lower HSP47 concentration (▣), at 1:2 (HSP47:collagen).

Investigation of Secondary Structural Changes during pH Titration

CD spectroscopy has been used previously to demonstrate that HSP47 structurally undergoes a two-step conversion from high to low pH states *via* an intermediate state (El-Thaher et al., 1996). This transition was suggested to involve His protonation but which His residue or His residue cluster responsible for this change is unclear. Substituting a particular His residue with Ala would eliminate the imidazole side chain ionisation. If the residues are critical for the structural change, the mutant constructed would exhibit aberrant collagen binding/release behaviour and so may follow a different structural transitional pathway from its binding to non-binding states. This can be obtained indirectly by examining the spectral transition upon gradual change in pH of the protein solution. pH was sequentially decreased by the addition of 1-2 µL aliquots of phosphoric acid. Phosphoric acid was used instead of hydrochloric acid to avoid the presence of chloride (Cl^-) ions, which is known to interfere with CD signal (Dafforn et al., 2004).

Figure 2.16 shows the effect of pH on the ellipticity of HSP47 WT in the far-UV region. It is apparent that there is a significant change in the overall structure of the protein during the process. Curve fitting procedure with k2D deconvolution program (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (Whitmore and Wallace, 2004; Whitmore and Wallace, 2008) suggests that decrease in pH induces the decrease in α-helical character and increase in β-sheet structure of the protein (Table 2.3). Even though the use of curve fitting for CD spectra is debatable when discussing the dynamics of protein structural change, a similar observation has been reported previously (Thomson and Ananthanarayanan, 2000) where the β-structure was shown to increase to nearly 40% at acidic pH. Hence this can be used to estimate of the extent of HSP47 conformational change with pH.

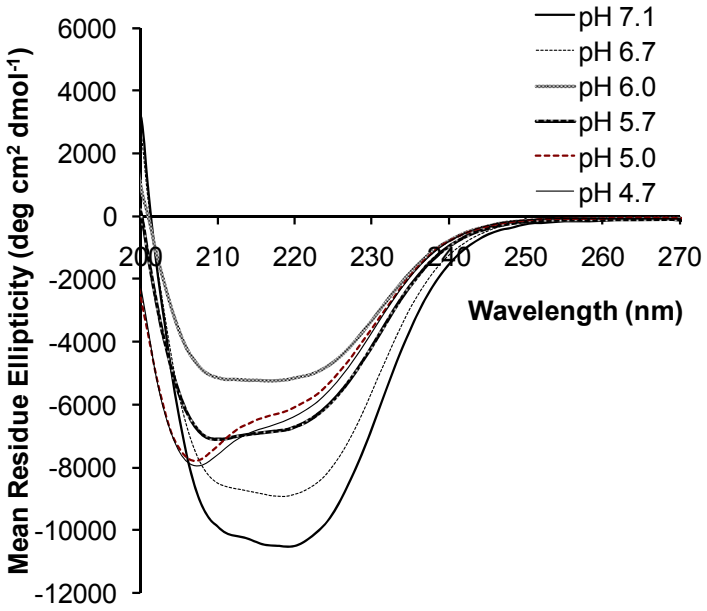


Figure 2.16: Effect of pH on the ellipticity of HSP47 WT.
Protein concentration is 3 μM, in 50 mM sodium phosphate buffer pH 8.0 with 100 mM NaCl, at 25°C.

Table 2.3: pH-induced changes in the secondary structural content of HSP47 WT.
Structural contents were estimated with k2D software on Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml).

pH	α-helix	β-sheet	Random coil	Error
7.1	0.32	0.17	0.51	0.085
6.7	0.31	0.13	0.56	0.087
6.3	0.31	0.13	0.57	0.087
6.0	0.21	0.26	0.52	> 0.22
5.7	0.25	0.20	0.55	> 0.22
5.3	0.14	0.33	0.53	> 0.22
5.0	0.15	0.30	0.55	0.182
4.7	0.18	0.28	0.55	0.182
4.3	0.14	0.36	0.51	> 0.22

By plotting the change in CD signal at 222 nm ($\Delta\Delta A_{222}$) as a function of pH, the resulting curve would show the secondary structural rearrangement, particularly α-helical content, upon change in the pH of the environment. Reduction in ellipticity at 222 nm is generally considered as a loss in α-helical content of proteins (Hirst and Brooks, 1994). For HSP47 WT, the observation was done independent of the spectra recording and titration was automated as pH was progressively monitored using an ultraslim electrode. For HSP47 WT, an abrupt change in structure is seen between pH 6.7 and 5.7 in the downward titration (Figure 2.17). This fits well with the results reported previously in a study that used transformed data from individual spectral scans at different pH values (El-Thaher et al., 1996). Analysis of the data obtained in this set of measurements revealed that the middle point for this transition (inflection point) is at pH 5.70 (± 0.10) (the procedure and assumptions are as described in the determination of melting temperature section). This value deviates slightly from the reported HSP47/collagen dissociation value of 6.3 (Saga et al., 1987), that can be attributed to the different environment in which the protein is studied (Chen et al., 2002; Garrett et al., 1998) such as the composition of the collagen type used in the study and buffer components. This is still in the typical pK_a range for surface histidine residues, as in human prolactin (hPRL) the range is reported to be between 5.8 to 6.6 (Tettamanzi et al., 2008).

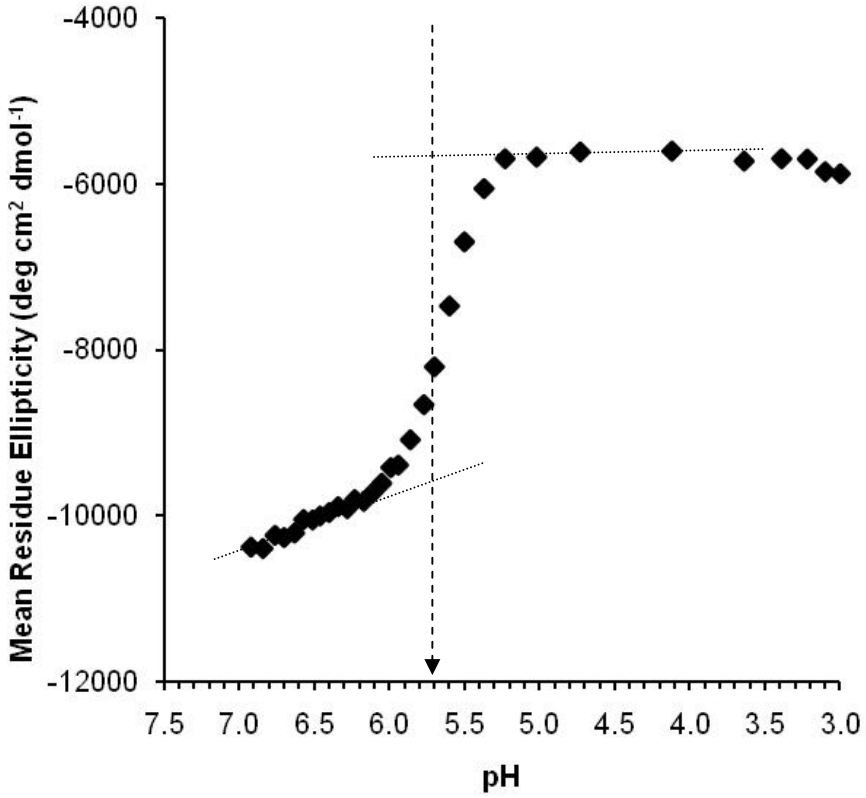


Figure 2.17: Change in CD signal upon titration of HSP47 WT, monitored at 222 nm. Data can be obtained either by plotting ellipticity at 222 nm from individual spectral scans at different pH values or automated titration measurement module at fixed wavelength available with the instrument. Both techniques gave comparable results. Broken arrow shows the transitional midpoint, calculated using the formula explained in Chapter 6 (Materials and Method). Dotted lines represent the start and end values of transition, estimated based on the 'tangent' method.

2.8 Characterisation of CMPs

In order to investigate the significance of the structural change to HSP47/collagen binding, a synthetic collagen mimetic peptide (CMP) was utilised as collagen substitute. CMPs have extensively been used in the study of HSP47-collagen interaction (Asada et al., 1999; Dafforn et al., 2001; Koide et al., 1999; Koide et al., 2006a; Nishikawa et al., 2010; Okano-Kosugi et al., 2009).

Among the mostly-used CMPs in HSP47-related binding experiments are the 30-amino acid long (Pro-Pro-Gly)$_{10}$ (Dafforn et al., 2001) and (Pro-Hyp-Gly)$_{10}$. The (POG)$_{10}$ peptide mimics collagen well as the basic collagen amino acid repeat is (X-Y-Gly)$_{10}$ where X is usually Pro and Y is either Pro or Hyp. Hyp has been known to stabilise triple helical peptide *via* hydration effect (Bella et al., 1995) or direct intermolecular hydrogen bonding interactions (Kramer et al., 2001). Therefore, it creates a stable triple helical peptide, which is suitable for various physicochemical studies (X-ray crystallography, NMR) and has been shown to assemble into higher order structures under defined conditions (neutral pH, temperature higher than 35°C, 1 mM) (Kar et al., 2006). (POG)$_{10}$ would be the control peptide in the identification of HSP47 binding partner candidate. Additionally, incorporation of Arg at Y position is known to enhance HSP47 binding affinity, compared to other amino acids. This is probably due the formation of a hydrophobic interface by the Arg side chain that directly interacts with HSP47 hydrophobic cleft (Koide et al., 2002; Nishikawa et al., 2010). For this reason, a CMP with one Arg residue was also included. The sequences of the two peptides are shown below:

(POG)$_{10}$, POG : POG POG POG POG POG POG POG POG POG POG

(POG)$_4$PRG(POG)$_5$, PRG : POG POG POG POG P**R**G POG POG POG POG POG

The ability of the peptides to form triple helices was investigated by CD. CMPs tend to form polyproline type II (P_{II}) conformation upon folding due to the rigidity of the imino acid backbone. This conformation gives rise to a distinctive positive peak at around 215-220 nm and a negative peak at around 195-197 nm (Goodman and Kwak, 1999; Tifany and Krimm, 1972). This was observed here with POG and PRG peptides after incubation at 4°C overnight (Figure 2.18). The extent of triple helical formation can be estimated from the ratio of positive over negative peak, R_{pn}. The calculated R_{pn} values for POG and PRG are 0.09 and 0.07, respectively, compared to a previously reported R_{pn} value for the native collagen of approximately 0.13 (Feng et al., 1996). The R_{pn} value shows that POG and PRG peptides have the propensity to form triple helical molecules.

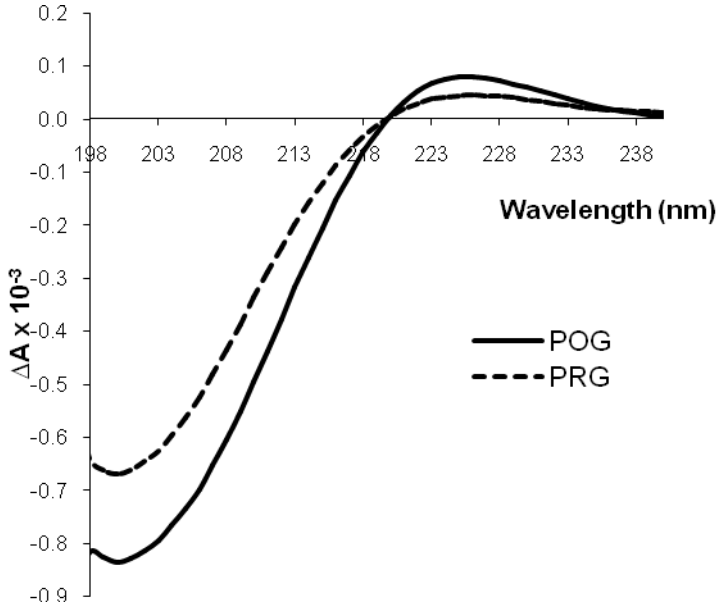


Figure 2.18: CD spectra of POG and PRG peptides.
Measurements were done at 20 μM in sodium phosphate buffer pH 8.0 with 100 mM NaCl, at 25°C. Ellipticity is expressed in Δ*A* instead of MRE. The appearance of maximum at 225 nm and minimum at 200 nm indicates that the peptides adopt triple helical conformation.

To investigate the binding affinity of HSP47 to the peptides, a pull-down assay was designed with the peptides immobilised onto cyanogen bromide (CNBr)-activated Sepharose beads. CNBr reacts with hydroxyl (–OH) groups on Sepharose to form reactive cyanate ester groups. The peptides can be coupled to CNBr-activated Sepharose, under mild conditions, *via* primary amino ($–NH_2$) groups. The activated groups on the beads react with primary $–NH_2$ groups on the peptide to form isourea linkages as shown in Figure 2.19 (Kohn and Wilchek, 1982). Monitoring of the coupling efficiency was performed using HPLC system coupled to particle detector unit. Figure 2.20 shows the chromatogram of the peptides before and after coupling, in which more than 99% of the peptides were successfully coupled to the Sepharose beads. The unreacted active groups were also blocked with ethanolamine to prevent unspecific binding in the next experiments.

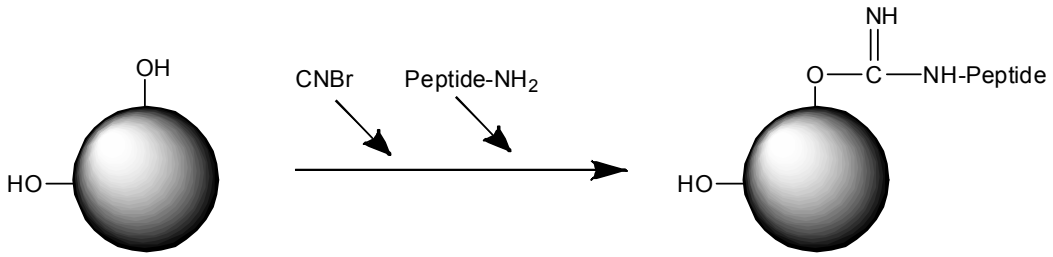


Figure 2.19: Sepharose beads activation by CNBr and peptide coupling to the activated matrix.
Figure adapted from Affinity Chromatography Handbook (GE Healthcare).

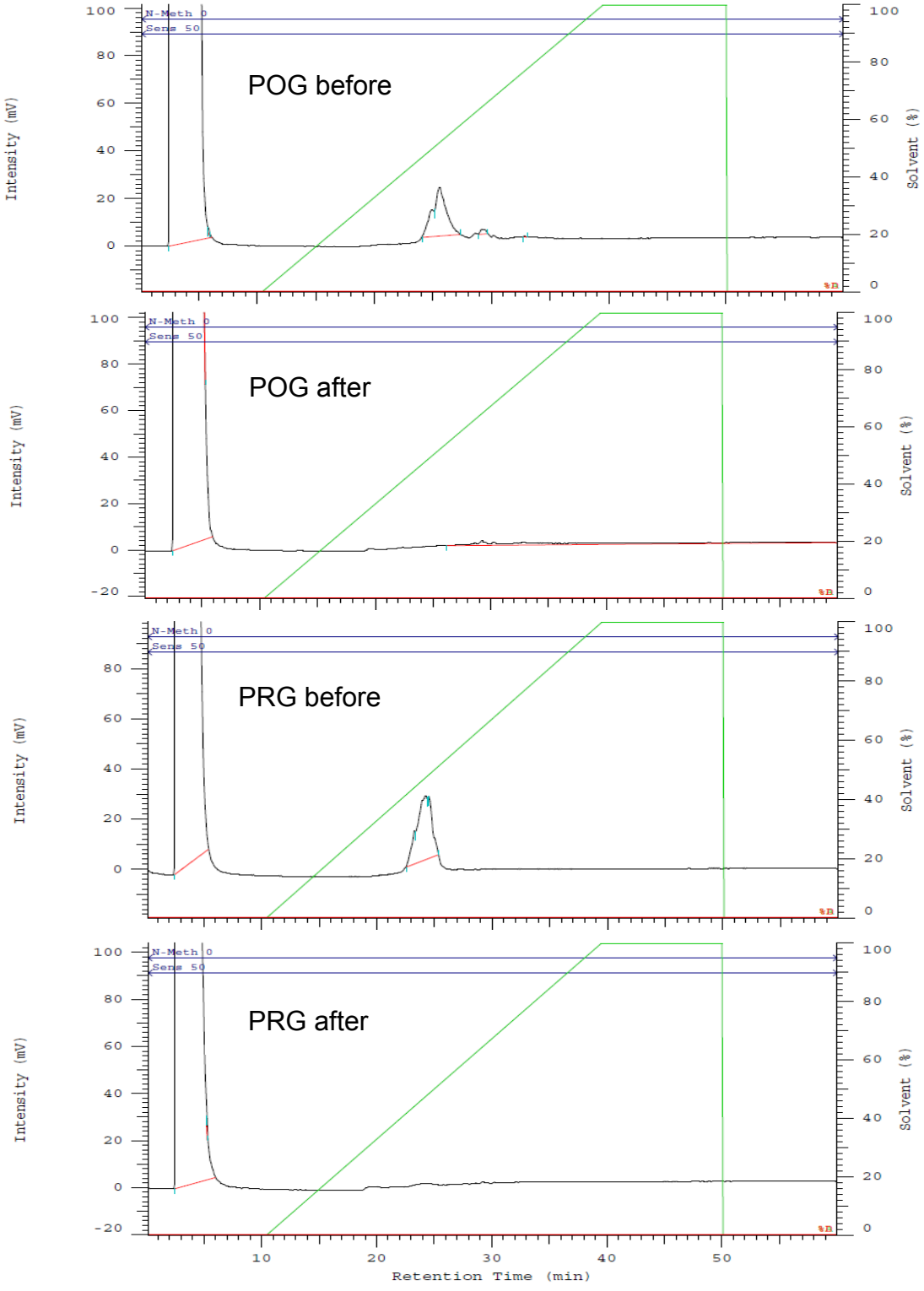


Figure 2.20: Chromatograms of POG and PRG peptide (0.75 mg/mL) before and after coupling.
Peptide samples were in 50 mM sodium phosphate pH 8.0 with 100 mM NaCl.

HSP47 binding assays were then performed on the beads with the immobilised peptides, using the same protocol as in the pull down assay with gelatin agarose beads. The obtained gel image revealed that PRG beads retain much more HSP47 than POG (Figure 2.21) suggesting a significantly stronger binding of HSP47 to PRG than to POG. Though understandably HSP47 shows lower binding affinity than it does to the much larger molecules of gelatin (lane labelled 'GA'), HSP47 binding to PRG beads is evidence of successful coupling and confirm previous observations (Koide et al., 2002). This suggests that PRG peptide synthesised could be a suitable HSP47 binding partner.

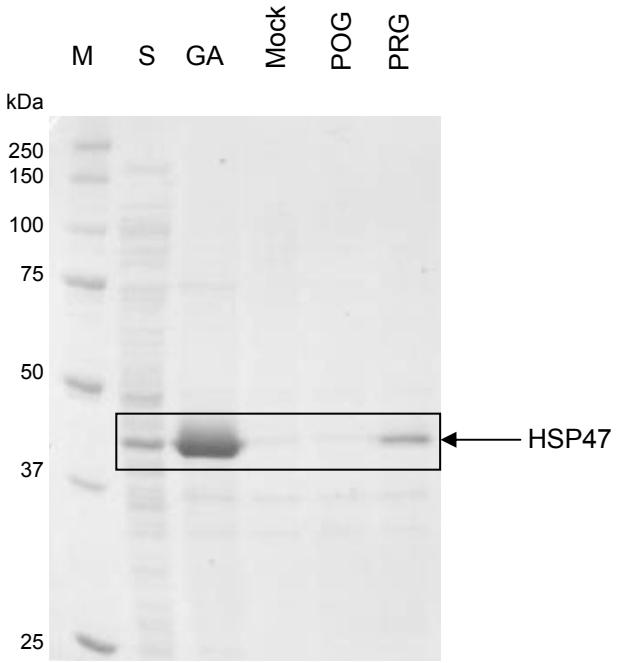


Figure 2.21: Affinity of HSP47 to immobilised peptides, visualised with CBB staining after separation on 12% SDS-PAGE under reducing condition.
M: Precision Plus Protein Standard (Bio-rad). S: cell supernatant (total soluble protein content). GA: Gelatin-agarose beads. Mock: uncoupled Sepharose beads.

2.9 Binding Affinity of HSP47 WT to PRG Peptide

CD was used to probe the binding of HSP47 to PRG peptide in response to pH change. The binding was investigated by monitoring the changes in CD signal upon addition of the peptide, with a blank titration for background subtraction. The change in signal at 216 nm (ΔA_{216}) was then plotted as a function of PRG concentration giving a saturation binding isotherm that can be fitted to a single site binding model (Figure 2.22). The wavelength at 216 nm was chosen instead of at 225 nm due to the change in CD signal is larger at that wavelength, giving a higher quality titration data (Dafforn et al., 2001).

It was found that HSP47 still binds PRG peptide at pH 6.5. Then as the pH reaches 5.5, complete dissociation occurs based on the non-converging curve fit. This suggest that structural transition that occurs at around that pH is accompanied by the release of HSP47 binding to peptide, which can be related to full length collagen as well. An attempt to narrow down the pH range of HSP47-peptide dissociation by observing CD spectral patterns of smaller pH intervals was unsuccessful due to the detection limit of the CD spectropolarimeter and the noise that accompanies every titration. Despite this, the information obtained is useful in explaining the nature of HSP47 pH-triggerable release. Combined with HSP47 WT gelatin column release behaviour, it is now established that the dissociation from collagen is accompanied (or triggered) by a large change in its secondary structure. As Arg residue forms an important part of the binding interface, it is possible that the structural change triggers the change in the hydrophobic interface of HSP47, which in turn results in dissociation.

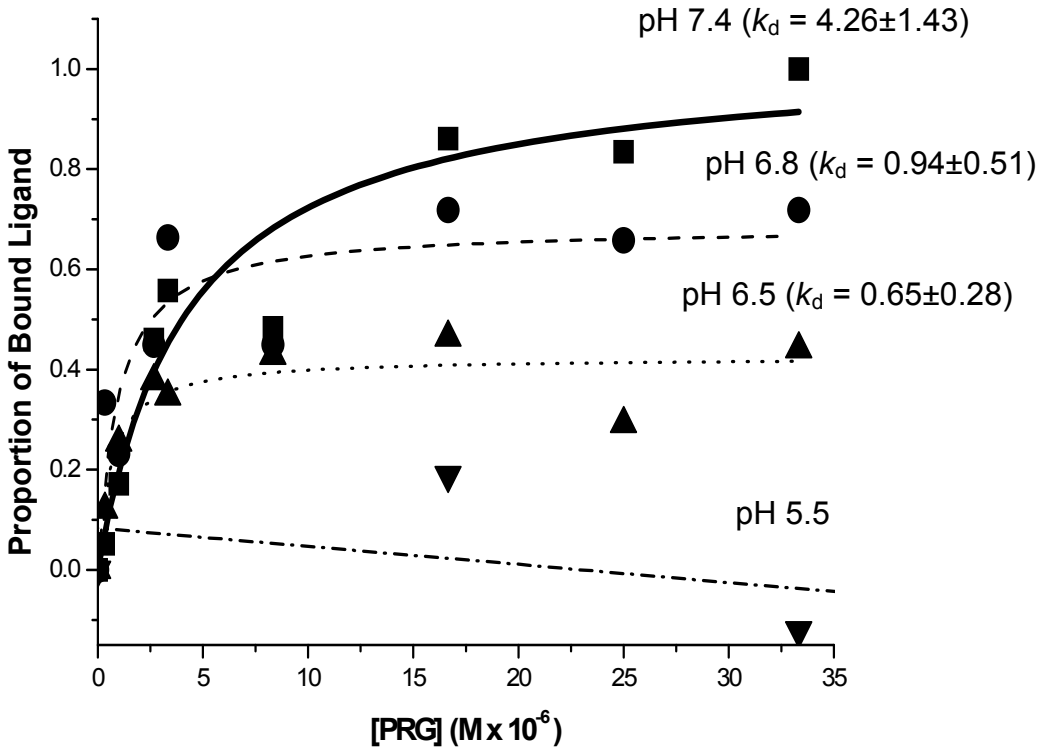


Figure 2.22: Binding isotherms showing the effect of pH on the binding of HSP47 WT (0.8 µM) to PRG peptide.
Isotherms were determined at pH 7.4, 6.8, 6.5, and 5.5. All isotherms were plotted as percentage of saturation for comparison, except for at pH 5.5 where data was fitted to linear function for easier visualisation. Curve fitting and dissociation constant (k_d) determination was performed using OriginPro v8.5.

To discuss this process further, it is important to gather all the information from previous studies that involve HSP47-collagen binding interface. To date, there has been no specific collagen binding site proposed for HSP47 at atomic resolution due to the absence of the three-dimensional structure of the chaperone. An interesting suggestion was made based on work with synthetic peptides on the orientation of HSP47 binding interface (Koide et al., 2006a), that revolves around the crucial Arg residue. Using peptides that incorporate photoreactive *p*-benzoyl-$_{L}$-phenylalanine (Bpa) residue at various positions along the peptide chain, a binding pattern of HSP47 to peptide upon UV-cross linking was obtained. It was found that Bpa residues located at positions from -10 to 8 to the central Pro-Arg-Gly (known as the HSP47 recognition sequence, with Arg at position 0) are recognised by HSP47 (Koide et al., 2002). The length of this recognition region is about 5.3 nm, leading to suggestion that HSP47 binds collagen in such a way that the longer axis is parallel to the common axis of the collagen triple helix (Figure 2.23A).

Alternatively, the hydrophobic region outlined by sheet B, helix A (hA), helix G (hG) and helix H (hH) in the HSP47 homology model has also been speculated to be a collagen binding site (Dafforn et al., 2001; Davids et al., 1995) (Figure 2.23B). This was suggested based on the possibility of the modelled groove to accommodate a collagen triple helical peptide *via* hydrophobic interactions. Deletion of hA and strands 4 of sheet B (s4B) and s5B has been shown to abolish the binding (Davids et al., 1995), which suggests an affected structural conformation. Upon detailed observation of the homology model that we constructed in this study, it was found that many residues located in sheet B, hA, hG and hH may contribute to the hydrophobicity of the groove (*e.g.* Lys, Pro, Ile). This makes the proposal worthy of further investigation.

In both proposals, hydrophobic interaction seems to play a major part. For this reason, it is very possible that release of HSP47 from collagen is triggered by structural rearrangement that abolishes hydrophobic interface in contact with a collagen chain. This rearrangement could be a consequence of the ionisation of His imidazole side chain. Subsequent study with HA mutants should elucidate this even further.

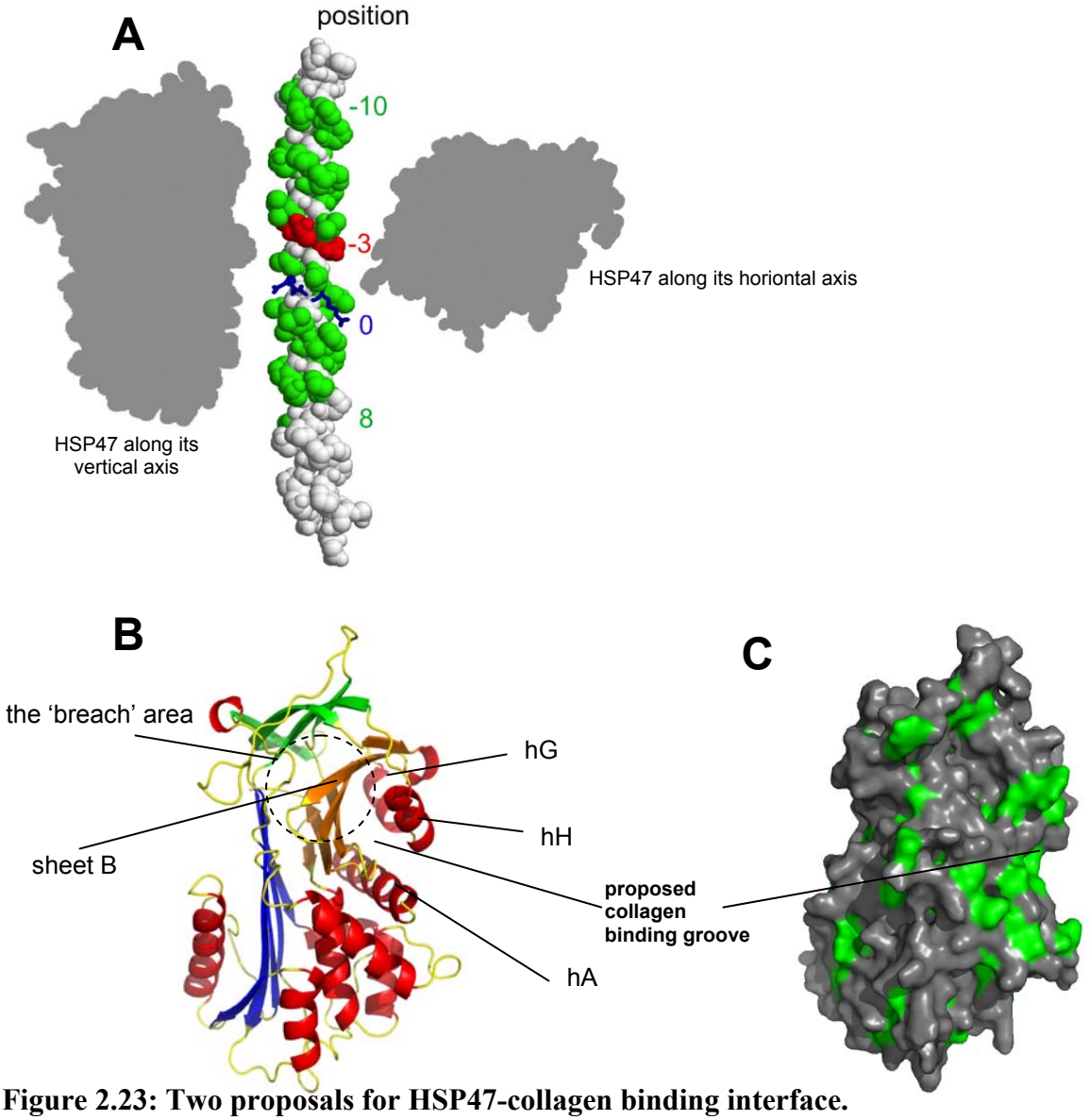


Figure 2.23: Two proposals for HSP47-collagen binding interface.
A. HSP47-collagen complex based on UV cross-linking of photoreactive residues to HSP47, with the critical Arg residue as position '0'. Residue highlighted in red could not be replaced with Bpa due to the detrimental effect of substitution to HSP47 binding. In the model HSP47 is proposed to bind collagen parallel to the common axis of the triple helix. From (Koide et al., 2006a). **B**. Hydrophobic collagen binding groove suggested from earlier HSP47 homology modelling work. β-sheet B serves as the 'base' of the groove, with helices A, H and G help to create a closed hydrophobic environment. From (Dafforn et al., 2001; Davids et al., 1995) shown on current homology model of HSP47. **C. '**Surface' representation of **B** to show the hydrophobic regions, especially near the proposed collagen binding groove. Hydrophobicity classifications of amino acid residues are adapted from http://www.sigmaaldrich.com/life-science/metabolomics/learning-center/amino-acid-reference-chart.html based on (Monera et al., 1995).

2.10 Characterisations of HSP47 HA Mutants

CD Structural Analysis

CD spectra of the mutants were first examined to assess conformational changes of the expressed proteins and probe any distortion in the protein structure due to the substitution of hydrophobic environment of the respective residues and the neighbouring residues. Overlaying spectra of WT, H135A, H191A, H244A, H255A:H256A and H302A (H197A:H198A could not be analysed as this mutant does not bind well to gelatin-agarose column) suggests that the mutants have secondary structural content similar to that of WT as all of the mutants give almost identical spectra (Figure 2.24A). This means that the content of β-sheets and α-helices and consequently the secondary structure are not significantly altered upon mutation of these His residues, at alkaline pH.

As shown in Figure 2.24B, the melting temperatures (T_m) of these mutants narrowly varies between 57-58°C (WT 59.0 ± 0.5; H135A 57.0 ± 0.5; H191A 57.0 ± 0.5; H244A 57.5 ± 0.5; H255A:H256A 58.0 ± 0.5; H302A 58.0 ± 0.5°C) which is not markedly different to WT (59°C). Again, this suggests that the respective mutations do not induce HSP47 to adopt a different conformation, *e.g.* the typical hyperthermostable latent form of serpins, as the latent form normally has a very high melting temperature, more than 70°C (Dafforn et al., 2004).

Combination of spectral and melting data presented here suggests that mutations of His residues at positions 135, 191, 244, 255, 256 and 302 present no noticeable alterations in the secondary structure of HSP47. This is also consistent with their predicted locations at the surface of the protein.

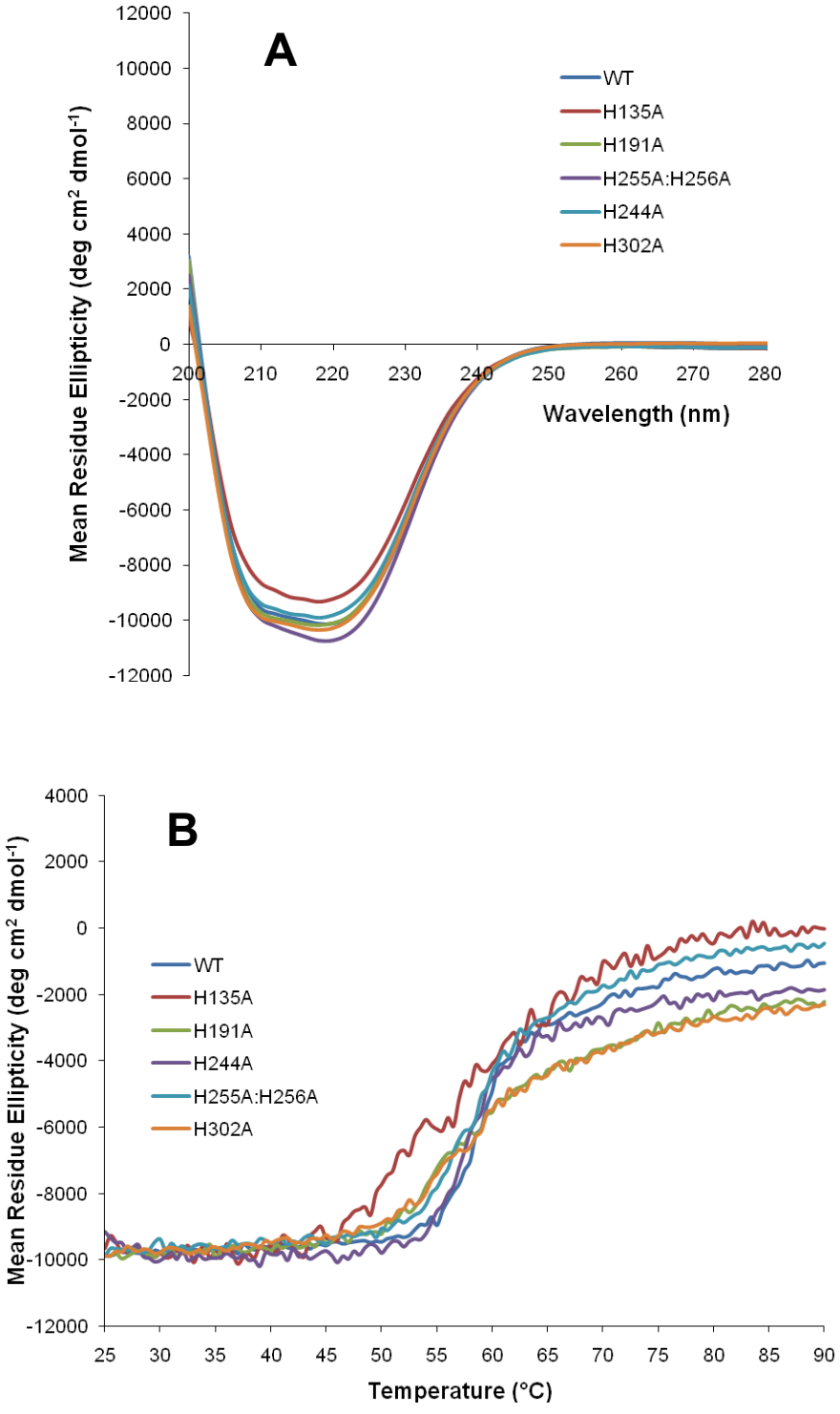


Figure 2.24: Far-UV CD spectra (A) and thermal denaturation profiles (B) of HSP47 WT and HA mutants.
Concentrations were between 2.8-3.0 µM and measured in 50 mM sodium phosphate buffer pH 8.0 with 100 mM NaCl.

Collagen Anti-fibrillisation Property

It has been demonstrated that mouse HSP47 WT has the ability to prevent collagen fibrillisation at equal molar ratio of HSP47 to collagen and so do all of these HA mutants as shown in Figure 2.25. This means that the collagen fibrillisation property is not lost upon mutations suggesting the HA mutations do not substantially alter the collagen binding region nor the binding interface. This is also in agreement with the CD data that the native folding of HSP47 is not altered upon introduction of the mutation. Consequently all HA mutants in the 'gate' and 'shutter' cluster are able to prevent fibrillisation effectively in the same way and to the same extent as HSP47 WT. In addition, the mutant carrying 'breach' cluster mutation (using H191A as model), is also able to fully arrest fibrillisation. This confirms our observations from gelatin-agarose pull-down binding assay (Figure 2.7B). It also underlines that the differences between 'breach' HA mutants (His residues at positions 197 and 198) and HSP47 WT in terms of gelatin column elution behaviour (Figure 2.10), are probably not associated with concomitant functional defects in the initial collagen recognition of the mutants. Whether or not there is another mechanism involves after the initial docking is unknown, but it is now getting clearer that the His residues actively participating in the pH-switch behaviour localises in the 'breach' cluster of the chaperone. It is also tempting to investigate further the effect of mutations on the structural transitions of HSP47 induced by pH-change.

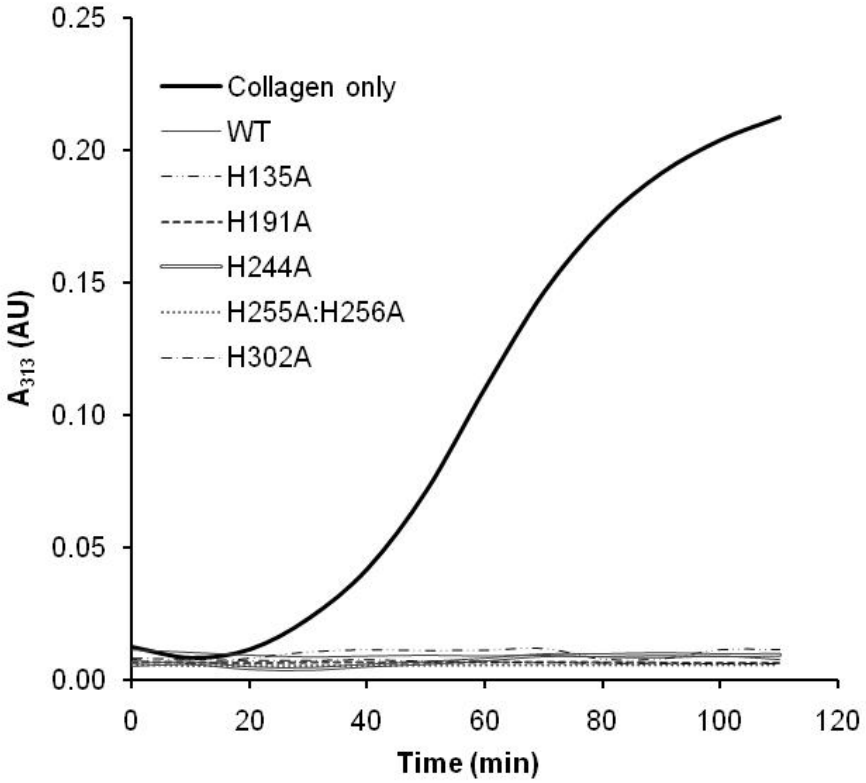
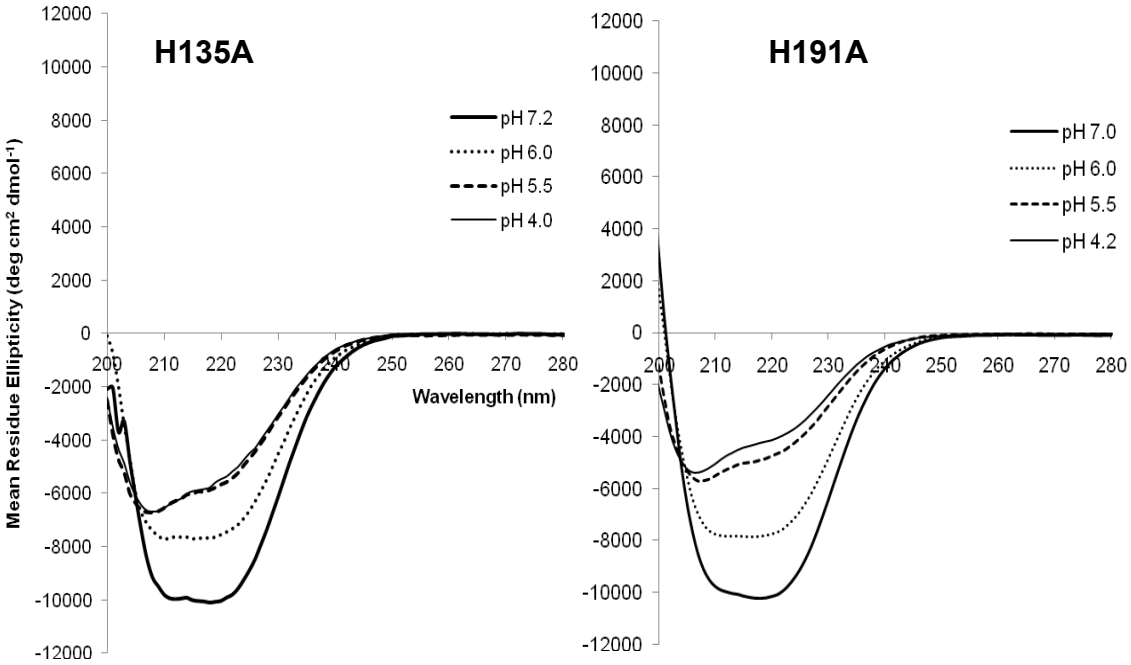


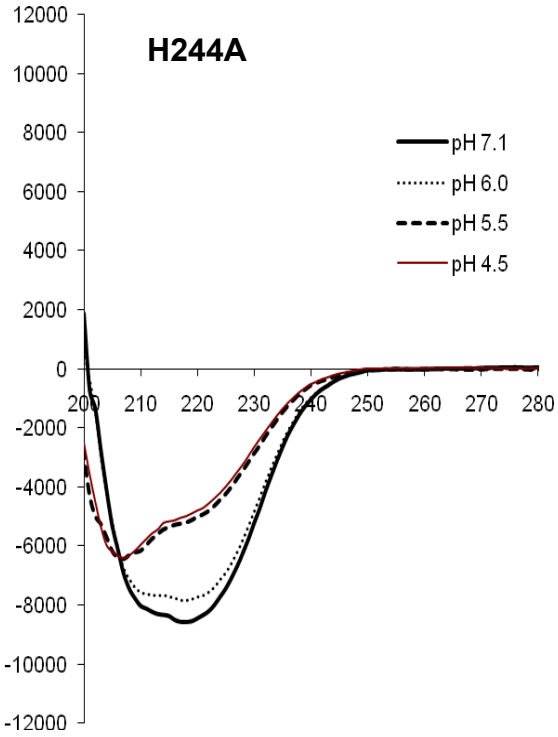
Figure 2.25: The fibrillisation curve of collagen solution in the presence of HSP47 WT and HA mutants upon incubation at 34°C.
pH of the reaction mix is between 7.3-7.4, with 2 μM concentration of collagen and HSP47. Assays were performed in triplicate.

Secondary Structural Changes during pH Titration

Towards Predicting the Structural States of Mutants HSP47 during Titration

As the HA mutants were constructed to probe the critical residues participating in the HSP47 pH-switch mechanism that triggers its release to collagen, pH titration would be a useful approach to probe the structural changes occurring during this transition. The overall structural change is first investigated by observing CD spectra at certain pH values. As seen in Figure 2.26, the transitions of the mutants showed similar behaviour to WT. From the spectra (representing the secondary structures possessed), it can be concluded that all mutants (apart of those at positions 197 and 198, in which the effect is unknown as they could not be purified using gelatin-agarose column) undergo a similar secondary structural transitional pathway from alkaline to acid states. It can also be concluded that the initial and final states are not significantly changed from the common shape of the CD spectra. These suggest similar secondary structural content. This information is useful in determining that the introduced mutations alter the pH transition of the mutants but not their initial alkali and final acid states.


H135A
Mean Residue Ellipticity (deg cm2 dmol-1)
Wavelength (nm)
pH 7.2
pH 6.0
pH 5.5
pH 4.0
H191A
pH 7.0
pH 6.0
pH 5.5
pH 4.2


H244A
pH 7.1
pH 6.0
pH 5.5
pH 4.5

cont....

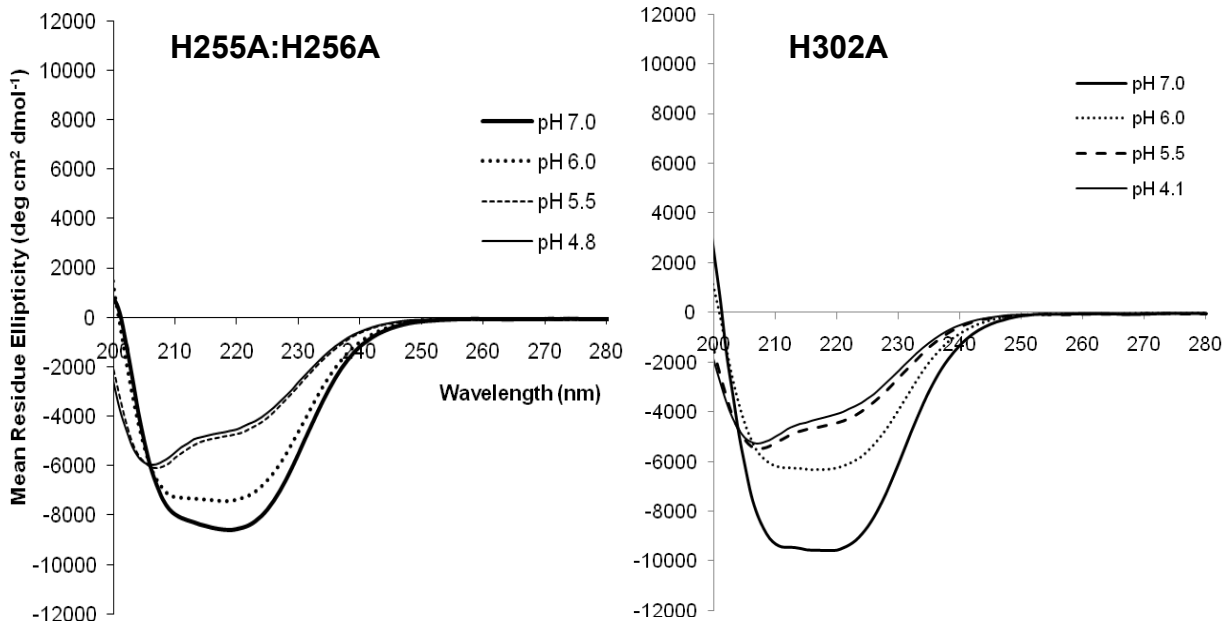


Figure 2.26: Effect of pH on the ellipticity of HSP47 HA mutants.
Protein concentrations were between 2.8-3.0 μM measured in 50 mM sodium phosphate buffer pH 8.0 with 100 mM NaCl, at 25°C.

Towards Finding the Collagen-release pH of the HSP47 Mutants

A more specific structural transition can be observed by looking at the change in ellipticity at 222 nm ($\Delta\Delta A_{222}$), which mostly corresponds to the change in α-helical character of the protein. As shown in Figure 2.27, mutants H135A, H244A, H255A:H256A and H302A have similar transitional midpoints to that of WT (WT at pH 5.70 ± 0.10 (mean ± SEM); H135A at pH 5.75 ± 0.05; H244A at pH 5.65 ± 0.05; H255:H256A at pH 5.75 ± 0.10; H302 at pH 5.75 ± 0.05). These values correspond to the dissociation pH of the mutants from the gelatin column (as demonstrated for WT), and thus can aptly be regarded as the collagen dissociation pH for each mutant. On the other hand, H191A was found to have a significantly higher midpoint compared to WT, *i.e.* pH 5.90 (± 0.05) against pH 5.70 for WT. These agree well with earlier FPLC chromatogram showing that this particular mutant elutes at a slightly higher pH value, and therefore at higher collagen dissociation pH. It can be concluded that the absence of H191 residue from HSP47 causes the chaperone to release its binding from collagen prematurely. It is very likely that this residue acts as key residue, locking the binding of HSP47 to collagen and only allows its release once the correct pH is reached. Data obtained from HA mutant experiments are summarised in Table 2.4.

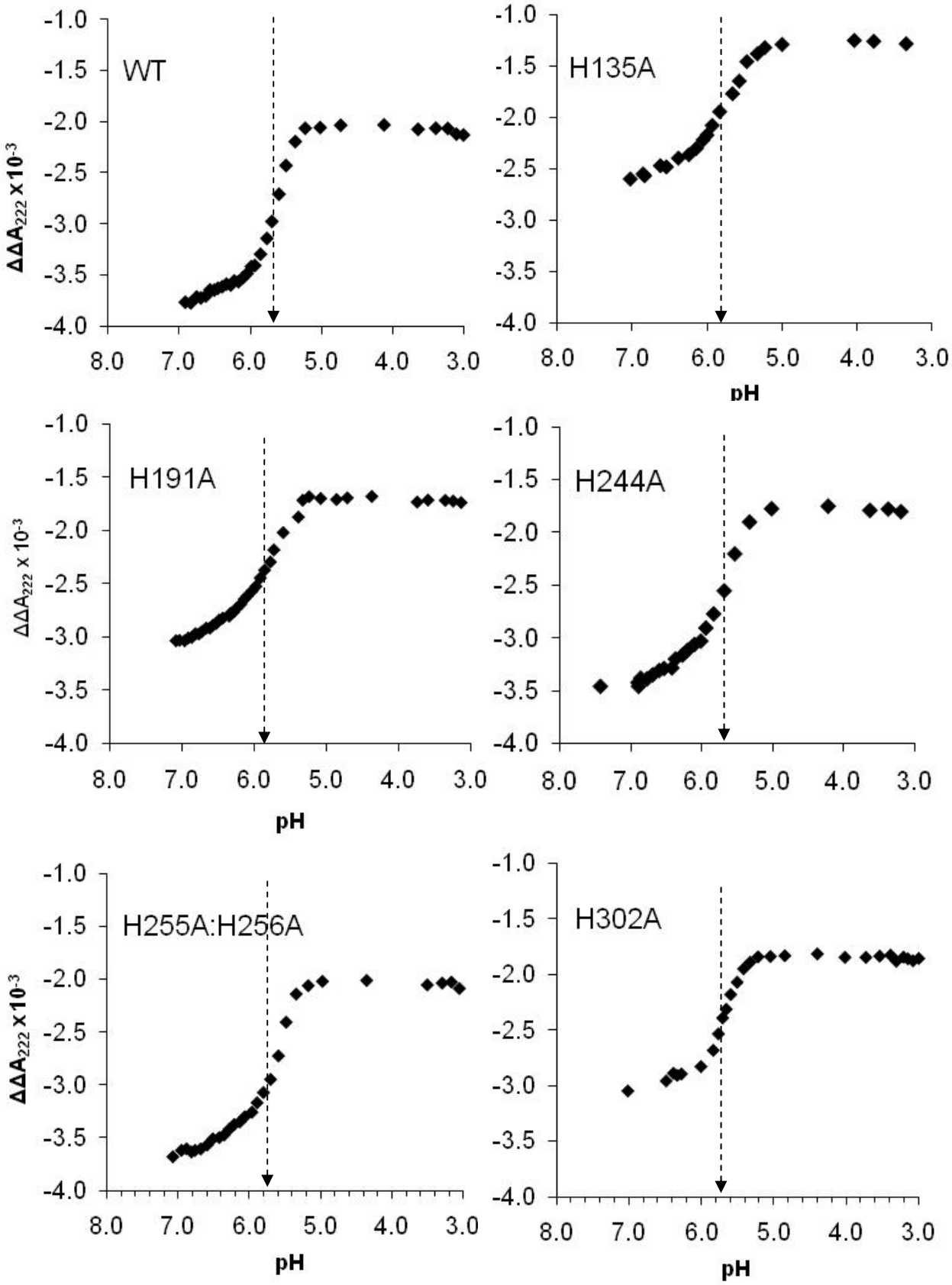


Figure 2.27: Change in CD signal upon titration of HSP47 WT and HA mutants, monitored at 222 nm.

Table 2.4: Summary of data on biophysical properties of HSP47 WT and mutants. Data shown includes thermal stability, inflection points observed under pH titration study, degree of conservation in HSP47 family and the secondary structural locations of the mutations introduced based on HSP47 homology model.

Mutations	Locations (Huber & Carrell nomenclature) (Huber and Carrell, 1989)	Conserved in HSP47 family?	T_m (°C) ± 0.5	Collagen anti-fibrillisation property	pH titration inflection points
WT	-	-	59.0	-	5.70 ± 0.10
H135A	hE	100%	57.0	Yes	5.75 ± 0.05
H191A	s3A-s4C, breach	100%	57.0	Yes	5.90 ± 0.05
H197A: H198A	s3A-s4C, breach	100% 92%	*	*	*
H244A	s2B-s3B, breach	92%	57.5	Yes	5.65 ± 0.05
H255A: H256A	s3B-hG, gate	46% 100%	58.0	Yes	5.75 ± 0.10
H302A	hI, shutter	92%	58.0	Yes	5.75 ± 0.05

*The properties of H197A:H198A could not be determined as it could not be eluted normally from gelatin-agarose column.

Determination of H191A Collagen-release pH using Fluorescence Anisotropy

Results presented in this chapter suggest that one mutant, H191A, exhibits notable behaviour especially its release from gelatin/collagen. This is based on several observations: First, the pH of the peak fraction of H191A in gelatin column elution experiment was found to be slightly higher than that of HSP47 WT; second, using CD, its pH induced trans-conformational change from alkali to acid state was also found have a midpoint at higher pH value compared to WT. This section is aimed at further investigating the active participation of H191A in the 'breach' cluster in triggering HSP47 release from collagen. Other than native collagen and gelatin, HSP47 is known to be able to bind CMP, as described in Section 2.9. Gelatin was used in the column elution experiment (Section 2.6), while collagen was used in the fibrillisation assay (Section 2.10). Native collagen has been used to demonstrate the different release of HSP47 WT and H191A mutant *via* fibrillisation assays at different pH conditions. Unfortunately, collagen fibril formation does not occur at acidic condition, even just below pH 7 (data not shown). For the same purpose, gelatin-agarose beads have also been used *via* modified pull down assay by washing the bound HSP47 with buffers at different pH values. This was also unsuccessful probably due to aggregation of gelatin-bound HSP47 when washed with low pH buffer, as observed on Western blot membrane probed with HSP47-specific antibody (data not shown). Hence, CMP is the next best candidate, used as substrate in fluorescence anisotropy experiment. Compared to the other two techniques described above, anisotropy spectroscopy measurement does not require high concentration of HSP47 and substrate due to its high sensitivity. This should eliminate aggregation problem at low pH. Data obtained from this technique should also come only from HSP47/CMP direct binding and release effect.

Anisotropy (r) is defined as the extent of polarisation of fluorescence emission when a sample is excited with polarised light. Anisotropy arises from the presence of transition moments for absorption and emission that lie along specific directions within the fluorophore structure (Lakowicz, 2006). The instrumental setup is similar to the normal fluorescence spectroscopy measurement, except that excitation and emission polarisers are positioned in the light-path for polarisation (Figure 2.28).

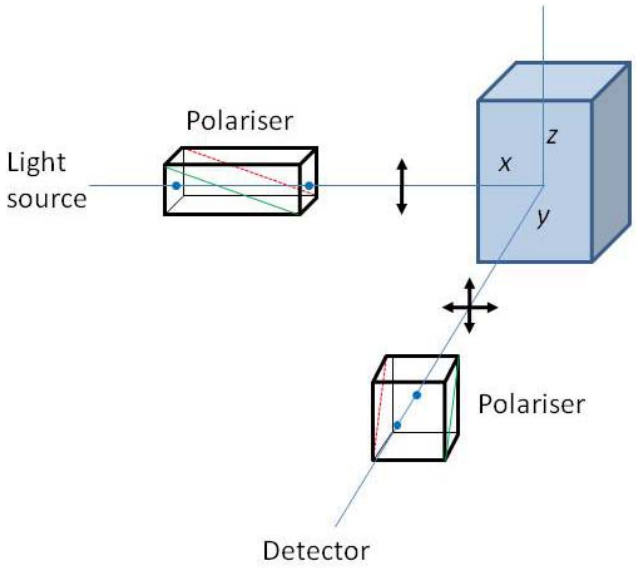


Figure 2.28: Measurement of fluorescence anisotropies shown in schematic diagram. Adapted from (Lakowicz, 2006).

The sample is usually excited with vertically polarised light, with the electric vector parallel to the *z*-axis. The emission intensity is then measured through a polariser, where when the emission polariser is oriented parallel to excitation polarisation the intensity is named $I_{\parallel}$. And when the polariser is perpendicular to the excitation polarisation the fluorescence intensity is called $I_{\perp}$. Anisotropy is then calculated according to the following equation:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \qquad \text{(Eq. 2.1)}$$

Anisotropy does not have a unit and it is independent of the total intensity of the sample and concentration of fluorophore. Rather, it is an intensity ratiometric measurement. Sometimes, the term polarisation is also used to describe this phenomenon though the use of anisotropy is more preferable. Polarisation is defined by:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \qquad \text{(Eq. 2.2)}$$

These two terms are related by these equations (Lakowicz, 2006):

$$P = \frac{3r}{2+r} \qquad \text{(Eq. 2.3)}$$

$$r = \frac{2P}{3-P} \qquad \text{(Eq. 2.4)}$$

When small fluorophores, such as fluorescent dyes, are excited with polarised light, the emitted light is largely depolarised because free molecules tumble rapidly in solution during its fluorescence lifetime. The fluorescent lifetime (τ) is the time between excitation and emission. But if the fluorophore binds to a larger molecule, its effective molecular volume increases. The rotation is slowed so that the emitted light is in the same plane as the excitation light. Because of this, bound fluorophore will have high intrinsic polarisation value, therefore high anisotropy, while free molecules have low value (Invitrogen, 2006) (Figure 2.29).

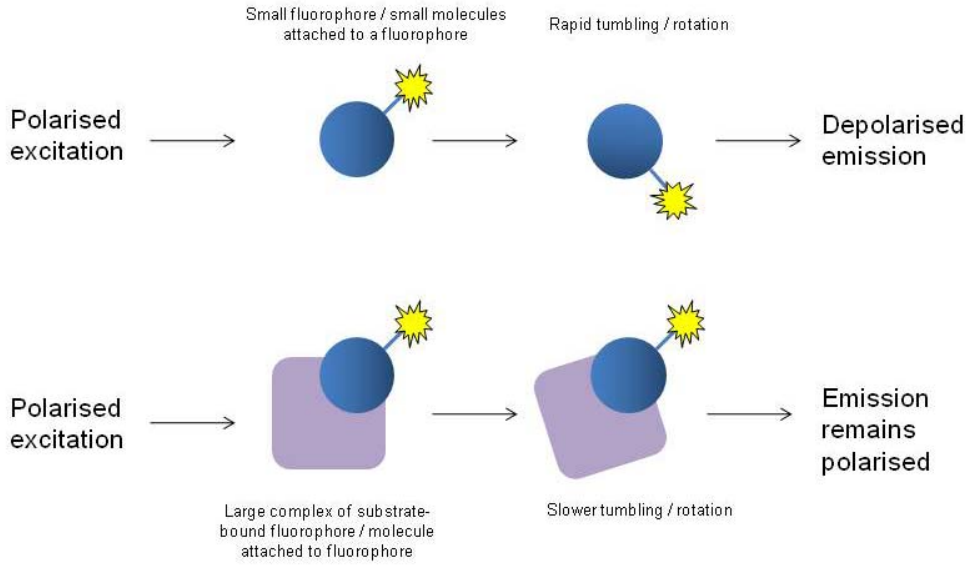


Figure 2.29: The origin of anisotropy difference between small, free molecule and large complexes.
Adapted from (Invitrogen, 2006).

A pH-stable fluorophore, with relatively high fluorescence lifetime is required to be coupled to CMP, as it does not possess intrinsic fluorescence. Among many amine reactive dyes, BODIPY (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-*s*-indacene-3-propionic acid) (Figure 2.30) was found to be suitable, particularly due to its stability across a wide pH range due to the absence of ionic charge, and relatively high fluorescence lifetime, τ~6 ns. This dye also has a high extinction coefficient and a very high quantum yield, approaching 1. Even though it has a small Stokes shift (10-20 nm), excitation at suboptimal wavelength can be used to avoid interference from light scattering during fluorescence detection (Hermanson, 1996). As an amine-reactive dye, BODIPY was coupled to PRG peptide (Figure 2.31) and conjugation was found to increase the hydrophobicity of PRG peptide due to the presence of BODIPY molecule at the *N*-terminus, as observed in the HPLC chromatogram of the purified conjugate in Figure 2.32. Thus it can be estimated that the conjugation reaction is almost 100% efficient due to absence of unconjugated PRG peptide.

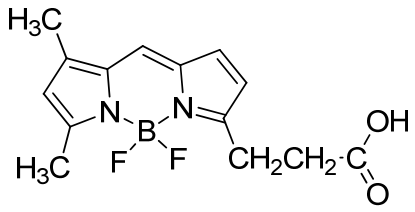


Figure 2.30: Molecular structure of BODIPY.

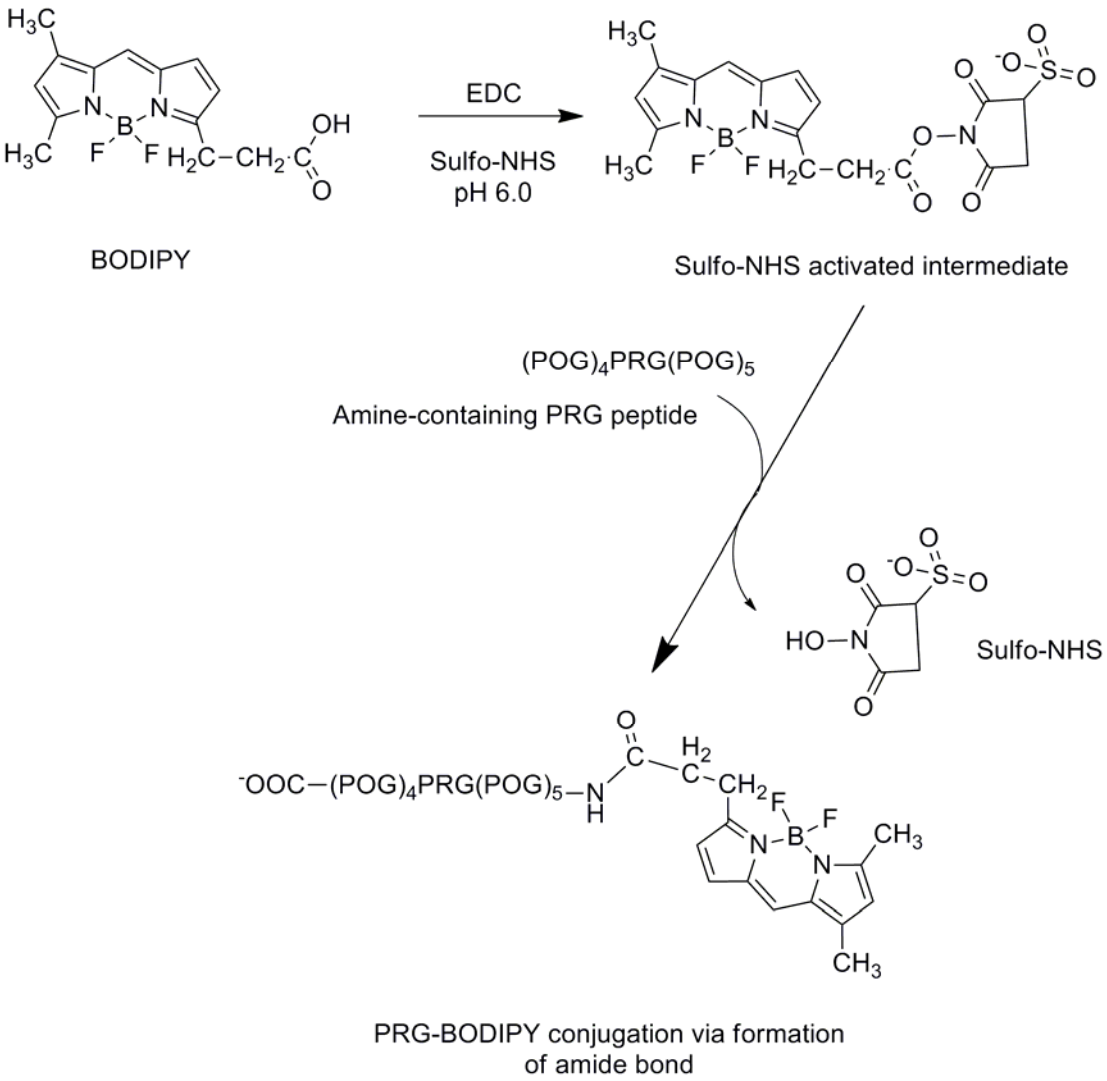


Figure 2.31: Coupling strategy for the conjugation of BODIPY to PRG peptide, adapted from (Hermanson, 1996).

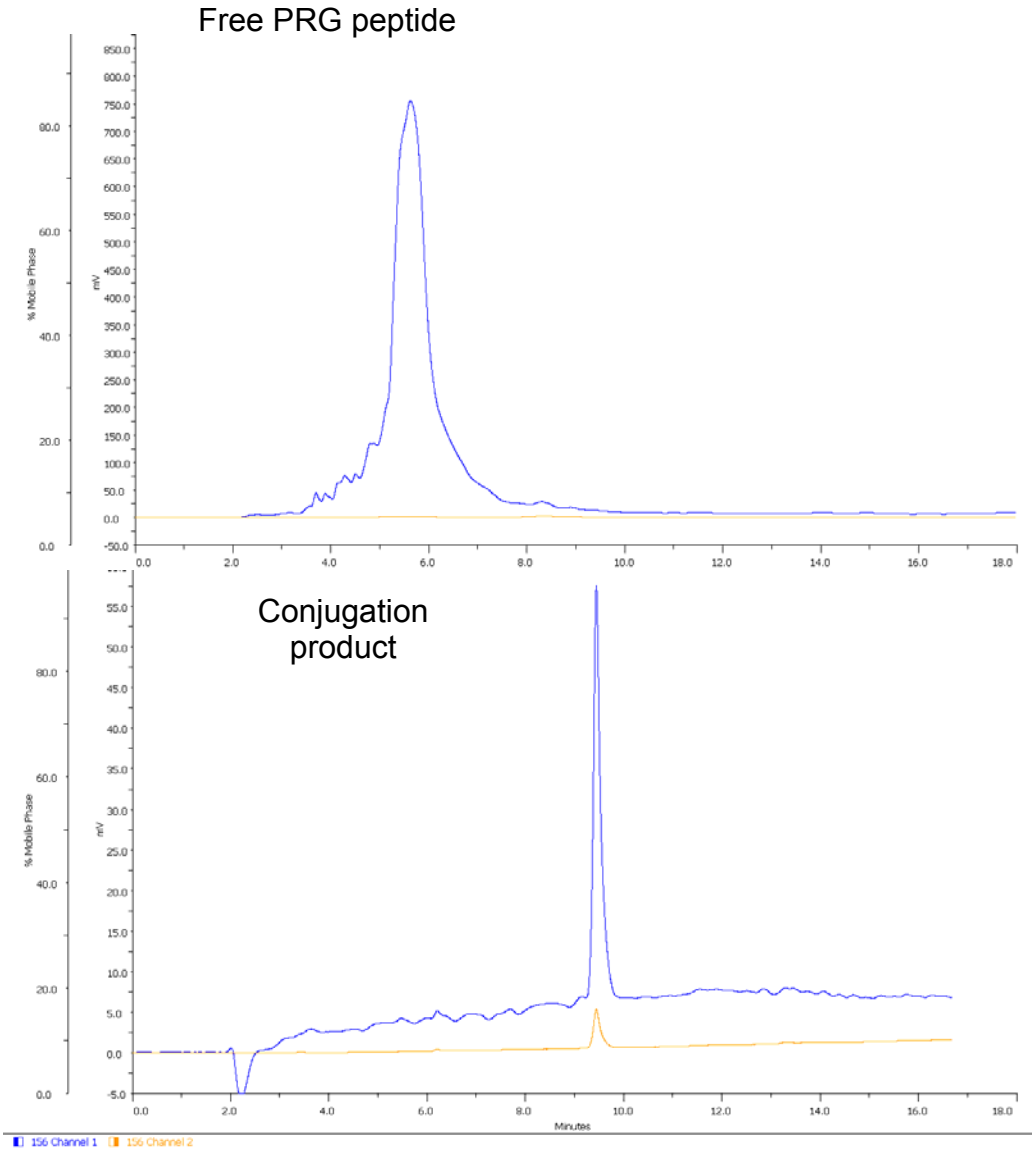


Figure 2.32: Reversed-phase HPLC analysis of conjugation reaction of PRG peptide to BODIPY dye, detected at 214 nm (blue) for peptide backbone absorption.

The conjugate was then used for anisotropy measurement. It was found that the anisotropy of the conjugate (0.0112 ± 0.0012) does not vary significantly from that of free BODIPY in water (0.0123 ± 0.0005), even though conjugation was predicted to increase the anisotropy of BODIPY due to decrease in molecular tumbling. Titration of the conjugate with increasing concentration of HSP47 (up to three-fold molar excess) also does not show any significant change in anisotropy (Figure 2.33).

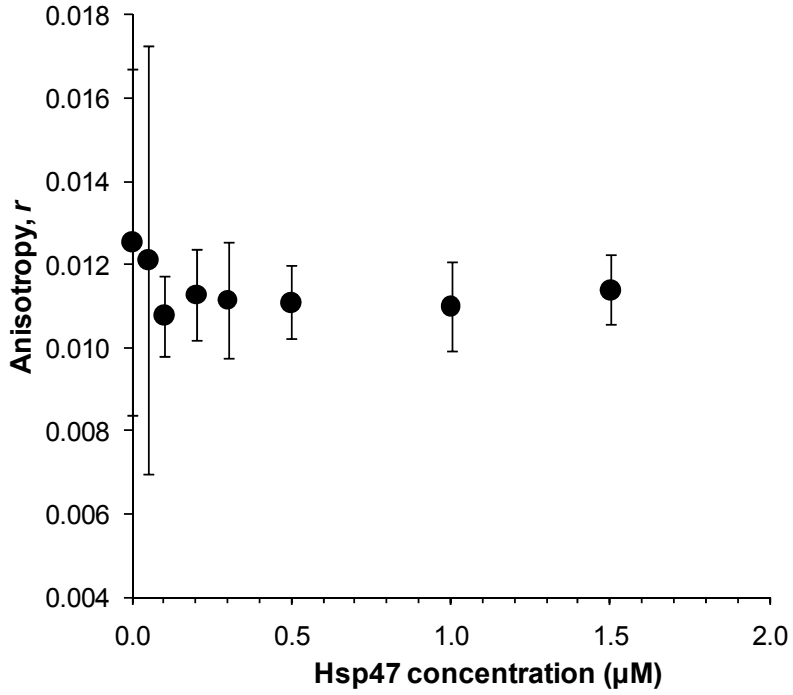


Figure 2.33: The plot of anisotropy change of PRG-BODIPY conjugate upon titration with increasing concentration of HSP47.
The concentration of conjugate was fixed at 0.5 μM. Titration did proceed beyond 3-fold molar excess of HSP47 to conjugate as HSP47 at high concentration tends to form insoluble aggregates. The experiment was performed in 50 mM sodium phosphate pH 8.0 with 100 mM NaCl, at 25°C.

To investigate this, Perrin equation and Stokes-Einstein-Debye equation were employed to calculate the molecular volume of the fluorescence species, so the hydrodynamic radius (Stokes radius, R_h) can be predicted for comparison with the values reported in the literatures. Perrin equation states that anisotropy depends on rotational diffusion of the molecule, expressed as rotational correlation time, Φ, based on Eq. 2.5 below. Φ depends on the size and shape of the molecule (González Flecha and Levi, 2003).

$$r = \frac{r_o}{1 + \tau / \phi} \qquad \text{(Eq. 2.5)}$$

Where r_o is the anisotropy in the absence of diffusion, τ is the fluorescence lifetime of the probe.

Most anisotropy experiments with macromolecules assume that it is a spherical particle rotating with small angular velocities. With this assumption, the rotational correlation time can be expressed by the Stokes-Einstein-Debye equation below:

$$\phi = \frac{\eta . V}{k . T} \qquad \text{(Eq. 2.6)}$$

Where η is the solvent viscosity coefficient, V is the molecular volume of the particle, k is the Boltzmann constant and T is temperature. Equations 2.5 and 2.6 can then be combined to give Eq. 2.8:

$$r = \frac{r_o}{1 + \frac{\tau . k . T}{\eta . V}} = \frac{\eta . r_o}{\eta + c} \qquad \text{(Eq. 2.7)}$$

From this relationship, the coefficient c and r_o can be obtained by plotting the change in anisotropy against different solvent viscosity (η). Figure 2.34 shows the plot constructed using the conjugate in the presence of increasing concentration of glycerol. From the plot, r_o was predicted as 0.0413 ± 0.0047 and c was 2.7756 ± 0.5234 mPa.s. Then using the average BODIPY fluorescence lifetime (τ) of 6 ns based on previous reports on BODIPY and its derivatives (Karolin et al., 1994; Mikhalyov et al., 2002) , k as 1.38×10^{-23} J/K, T = 293.15 K, the molecular volume of the conjugate was then calculated as 8.74 ± 1.64 nm^3. This gives a Stokes radius (hydrodynamic radius, R_h) of 1.27 ± 0.46 nm. The value obtained is slightly smaller than those reported for triple helical structure for collagen mimetic peptide which is

in the range of 1.6-2.5 nm (Kar et al., 2006; Kotch and Raines, 2006; Pires and Chmielewski, 2009) for 8-10 repeats of (POG) peptide. This computed value of R_h is also prone to error due to the assumption that the particle being considered is spherical, which is acceptable for proteins but should be exercised with care for non-spherical peptide (Edward, 1970).

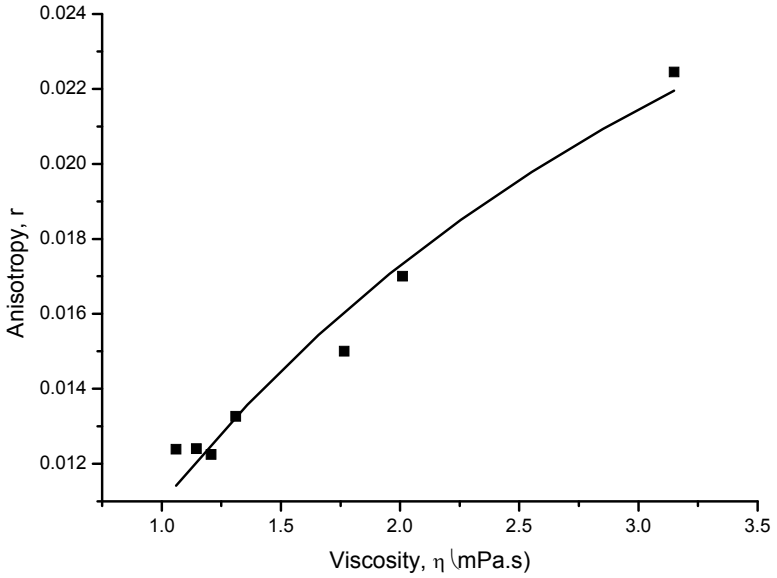


Figure 2.34: Dependence of conjugate anisotropy on the viscosity of solution.
The line represents the best fit of the date based on Eq. 2.7.

Furthermore, in order to determine the presence of any helical character after conjugation of BODIPY to PRG peptide, CD spectra of the coupled and uncoupled peptides were collected at equal concentration. From Figure 2.35, it can be concluded from the spectra that no triple helical character is present in the conjugate solution, unlike those observed with free peptide. Thus these suggest that conjugation prevents the formation of triple helical conformation of the collagen mimetic peptide PRG, although the mechanism is still unclear. It is possible that the steric effect at the *N*-terminus of the peptide chain increases after conjugation with BODIPY. This could then lead to the unchanged anisotropy of PRG-BODIPY conjugate upon titration with HSP47, as previous report suggests that HSP47 binds well to triple helical peptide but significantly lower to single stranded ones (Koide et al., 2006a). Despite this, fluorescence anisotropy is a very useful technique to study the interaction of HSP47 with its substrate, provided that the barrier to triple helical formation after conjugation could be overcome.

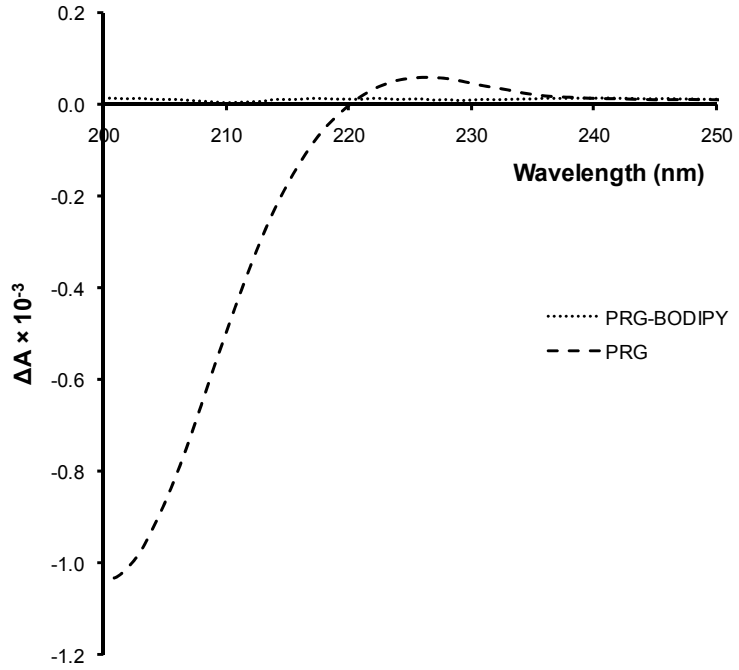


Figure 2.35: CD spectra of free PRG peptide and PRG conjugated with BODIPY. Measurements were performed at a concentration of 20 μM in water, at 25°C.

2.11 Conclusion

The investigation on HA mutants proved to be useful as the location of His residues that play the most important roles can now be narrowed down to the 'breach' cluster. Mutating H197 and/or 198 to Ala affects the typical collagen binding specificity while mutant H191A shows an affected release characteristic. It is therefore very likely that 'breach' cluster is the most important cluster in the HSP47 pH-change, allowing HSP47 to bind collagen at high pH and triggering the release at lower pH. Hence, study on the dynamics of these structural transitions was recommended and in the next chapter this was tackled by employing intrinsic tryptophan fluorescence spectroscopy. Overall, data presented in this chapter account for more than half of the His residues of HSP47. It is not at all implausible to propose that the residues actively involved in the pH-switch mechanism is highly localised to a certain region of the chaperone. It should also be highlighted that these suggestions are made based on the homology model of HSP47, depending on the accurate location of the residues mentioned. Otherwise, the high secondary structure similarity between serpin family members may support these proposals. The findings reported in this chapter were used in proposing the model for HSP47 pH-switch mechanism which would be presented in Chapter 4.

Chapter 3

Results and Discussion: HSP47 pH-induced Dynamics Probed using Intrinsic Tryptophan Fluorescence Spectroscopy

CHAPTER 3: HSP47 pH-INDUCED DYNAMICS PROBED USING INTRINSIC TRYPTOPHAN FLUORESCENCE SPECTROSCOPY

3.1 Introduction

Fluorescence spectroscopy has been used extensively to study protein conformation and interaction. Aromatic amino acids, such as tyrosine (Tyr, Y) and tryptophan (Trp, W), are able to act as intrinsic chromophores and therefore able to produce fluorescence signals (Figure 3.1). The nature of this process depends on the micro-environment of these chromophores, making it useful in studying the behaviours of proteins. Thus, intrinsic protein fluorescence can be used to monitor proteins of interest in response to heat, pressure, pH change and denaturants. When coupled with other biophysical techniques such as CD and NMR, conformational changes during these processes can be monitored in even greater detail. As fluorescence spectroscopy is very sensitive to both chemical and physical interference, high purity compounds are desirable (Pain, 2004).

The aim of this study is to locate any Trp residue(s) that may be in a hydrophobic/hydrophilic environment at HSP47 alkaline state and exposed to a different environment upon transition to acid state. Each Trp was mutated to phenylalanine (Phe, F) that has similar structural bulkiness compared to Trp, but do not contribute to fluorescence emission when excited at 295 nm. This could suggest an important dynamic region during pH transition if the fluorescence is quenched upon mutation. Direct monitoring of a single Trp residue is most desirable as this eliminates possible interference from other residues while multiple Trp-to-Phe (WF) mutations are not advisable as it has been shown in previous studies that such mutations pose additive instability effects on the protein (Hasselbacher et al., 1995; Meagher et al., 1998). This supports the approach to construct single WF mutant to observe if there is any deletion effect on WT fluorescence behaviour during pH titration. For this purpose, the fluorescence behaviour of HSP47 WT was initially established.

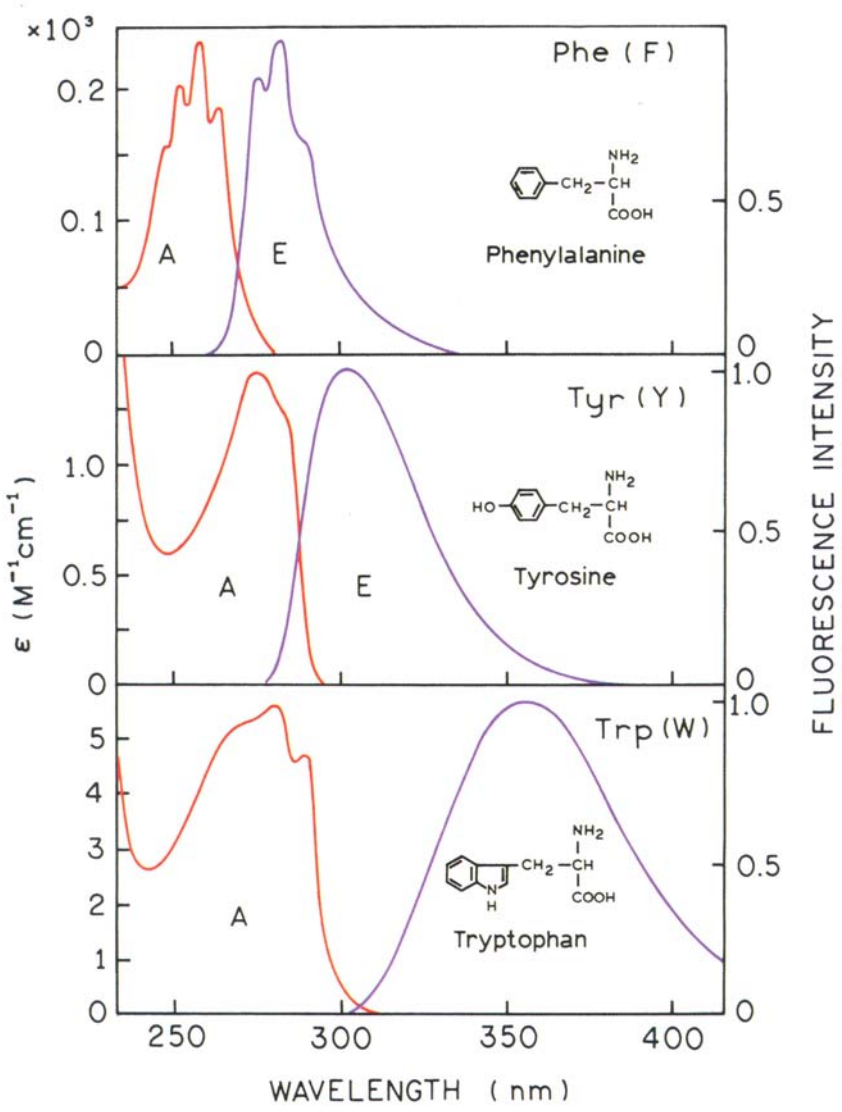


Figure 3.1: The absorption (A) and emission (E) spectra of Phe, Tyr and Trp in aqueous solution pH 7 (Lakowicz, 2006).

In the previous chapter, the 'breach' cluster of HSP47 has been established to contain the His residues participating actively in the pH-triggerable release of binding to collagen. There should also be a significant regional structural transition that takes place during the process. In the absence of a 3-dimensional structure, the extent of the structural rearrangement is unknown. Although it is likely that whole protein experiences this change, it is even more likely that there are certain regions that significantly undergo this process. This is in light of the localisation of the important His residues in the 'breach' cluster of HSP47.

Intrinsic fluorescence spectroscopy is among the first methods used to study the structural transition of HSP47 from alkaline to acidic pH (El-Thaher et al., 1996) and ever since has been one of the methods of choice among HSP47 molecular researchers. These include studies on its structure (Dafforn et al., 2001; Thomson and Ananthanarayanan, 2000) and interactions with collagen and collagen-related peptides (Dafforn et al., 2001; Koide et al., 2006a; Koide et al., 2002; Macdonald and Bächinger, 2001).

In serpins, intrinsic Trp fluorescence has extensively been used previously to study the interaction between serpins and their cognate proteases. In the same year as the discovery of the crystal structure of the RCL of α_1-antitrypsin (α_1-AT, α_1-proteinase inhibitor) (Elliott et al., 1996), tryptophan fluorescence was used to distinguish between the various forms of the serpin (Koloczek et al., 1996). This includes the latent form, the A-sheet dimer ('head to tail' polymerisation), C-sheet dimer ('head to head' polymerisation) and the A-sheet dimer propagating from the C-sheet dimer. Wild-type α_1-AT does not readily adopt these forms, but only after a cycle of denaturation and renaturation. α_1-AT has two Trp residues, one in strand 3 of β-sheet A (W194) and the other in strand 2 of β-sheet A (W238). In the latent form, which is characterised by RCL insertion into the A-sheet, W238 is more exposed to solvent than W194 based on the shifting of emission maximum (λ_{em} of W238 is red-shifted by 4 nm compared to W194). This was determined using fluorescence quenching with potassium iodide (KI) as the quencher. W238 is assumed as quenchable component as it is exposed to the environment and *vice-versa* for W194. With fluorescence quenched resolved spectra (FRQS) and red-edge excitation techniques, the two forms of polymers can be distinguished based on dipole relaxation of W238 and W194: A-sheet polymerisation and C-sheet polymerisation. The former is characterised by the insertion of RCL into β-sheet A, and the latter is formed by interactions in the C-sheet. W238 is at least 50% more exposed in A-sheet polymers compared to C-sheet polymers (Koloczek et al., 1996).

The 'breach' residue W194 is then shown to be crucial in controlling the rate of polymerisation of antichymotrypsin (ACT) (Pearce et al., 2007). Again, the location of the residue at the *N*-terminus of RCL is probably the reason for the residue to adopt this role. In its native form, replacement of Trp with Phe (W194F) does not significantly alter the inhibitory property. The polymerisation process was found to be up to 14 times faster with the mutant than WT upon heating at 48°C, which induces polymerisation. Meanwhile, the rate of polymer intermediate species is accelerated 35 times when monitored using the

intensity of external reporter bis-8-anilinonaphthalene-1-sulfonate, also known as 4,4′-bis(1-anilinonaphtalene-8-sulfonate) or bis-ANS (Pearce et al., 2007). This compound was first shown to bind tightly to the hydrophobic pocket of bovine serum albumin (BSA) accompanied by substantial enhancement in fluorescence of the reporter (Rosen and Weber, 1969). Generally bis-ANS fluoresce brightly in hydrophobic environment and is virtually non-fluorescent in polar environment, *i.e*. water (Takashi et al., 1977). This particular property makes it suitable for probing hydrophobic patches on protein surfaces, protein unfolding process, and in the study using ACT, for monitoring of the progress of its polymerisation. Data obtained using WF mutant of ACT demonstrate that some Trp residues play an active role in structural integrity and stability in serpins. Bis-ANS has also been used in studying α_1-AT polymerisation mechanism (James and Bottomley, 1998), which effectively is a continuation to another study using intrinsic Trp fluorescence reported two years earlier (Koloczek et al., 1996). As explained above, bis-ANS binds hydrophobic patches on the protein surface. The fluorescence of bis-ANS enhanced significantly upon binding to polymerised α_1-AT, suggesting the formation of more hydrophobic regions compared to the native protein. The first step of polymerisation, which is the opening of β-sheet A, is characterised by the increase in hydrophobicity in the area in vicinity to it. This then enhances the fluorescence of bis-ANS. In the second phase which is the addition of the monomer to the initiator, it is characterised by a gradual decrease in bis-ANS fluorescence signal. This phenomenon is related to the insertion of RCL of a donor molecule into the initiator, resulting in displacement of bis-ANS molecules bound to the 'opened' β-sheet A (James and Bottomley, 1998).

In a separate study, interaction of antithrombin (AT) and its heparin substrate has also been probed using Trp-to-Phe mutants (Meagher et al., 1998). Compared to HSP47 which has five Trp residues, AT has four Trp residues instead (here called AT-W49, 189, 225 and 307). The locations of two of the residues are analogous to the Trp residues in HSP47 (called HSP47-W). They are AT-W189, located on helix F (similar to HSP47-W158), and AT-W225 which corresponds to HSP47-W192 at the end of strand 3 of β-sheet A. The other Trp residues of AT are located at the beginning of helix-A (AT-W49) and finally AT-W307 is on helix H. The four WF mutants were used to monitor fluorescence changes upon binding to its substrate, heparin. The possibility of energy transfer between the residues is ruled out as one-third of the sum of fluorescence spectra of the mutants resembles that of WT. This is

consistent with calculation as distances between the residues are known to be between 15-35 Å. There is no significant energy transfer can be observed between residues separated more than 2 × R_o (Moens et al., 2004). R_o refers to *Förster distance* which is the distance at which resonance energy transfer (RET) is 50% efficient. The critical distance between Trp-Trp for RET is generally reported to be between 5-12 Å (Steinberg, 1971; Zhong et al., 2002). The mutants are then used to explain the structural rearrangement undergone by AT upon heparin binding in terms of fluorescence enhancement of the Trp residues. Steady-state fluorescence was also used to show the effect of heparin binding, most of which are experienced by two residues only (W225 and W307) based on the enhancement of fluorescence emission. The environment around W189, which resides at the furthest distance from heparin binding site and RCL, on the other hand, is less sensitive to the binding. Overall, Trp residues located in close proximity to the RCL and nearby regions involved in the *stressed* to *relaxed* transition experience the largest change in intensity and wavelength shift of fluorescence emission (Meagher et al., 1998).

These extensive studies on members of serpin family demonstrate the potential of fluorescence spectroscopy in elucidating protein folding processes, especially with respect to their secondary structural transition induced by different factors.

3.2 Mapping the Locations of Trp Residues in HSP47 Homology Model

HSP47 has five Trp residues at positions 110, 158, 192, 275 and 341 (Figure 3.2). In the model, W110 is located at the end of sheet 2 of β-sheet A (s2A) close to the loop between s2A and helix D (hD), W158 at the hF, W192 in the loop between s3A and sheet 4 of β-sheet C (s4C), W275 at the hH and finally W341 is located at the loop after s5A near the 'hinge' region, the region at which the RCL turns and joins β-sheets A to become s4A in latent or cleaved form of serpin (Hopkins et al., 1993). These Trp residues produce a strong intrinsic fluorescence emission when excited at 295 nm. All residues are predicted to be solvent accessible, and the accessibility ranges from almost 90% (W275) to 3% (W192). W110 and W341 are predicted to be about 30% exposed, while W158 is about 40%. Calculations were done using ASA-View program on the online server at http://gibk21.bse.kyutech.ac.jp/netasa/asaview/ (Ahmad et al., 2004). This is one of the structural classification and analysis tools that can be found on the RCSB Protein Data Bank http://nist.rcsb.org/pdb/home/home.do.

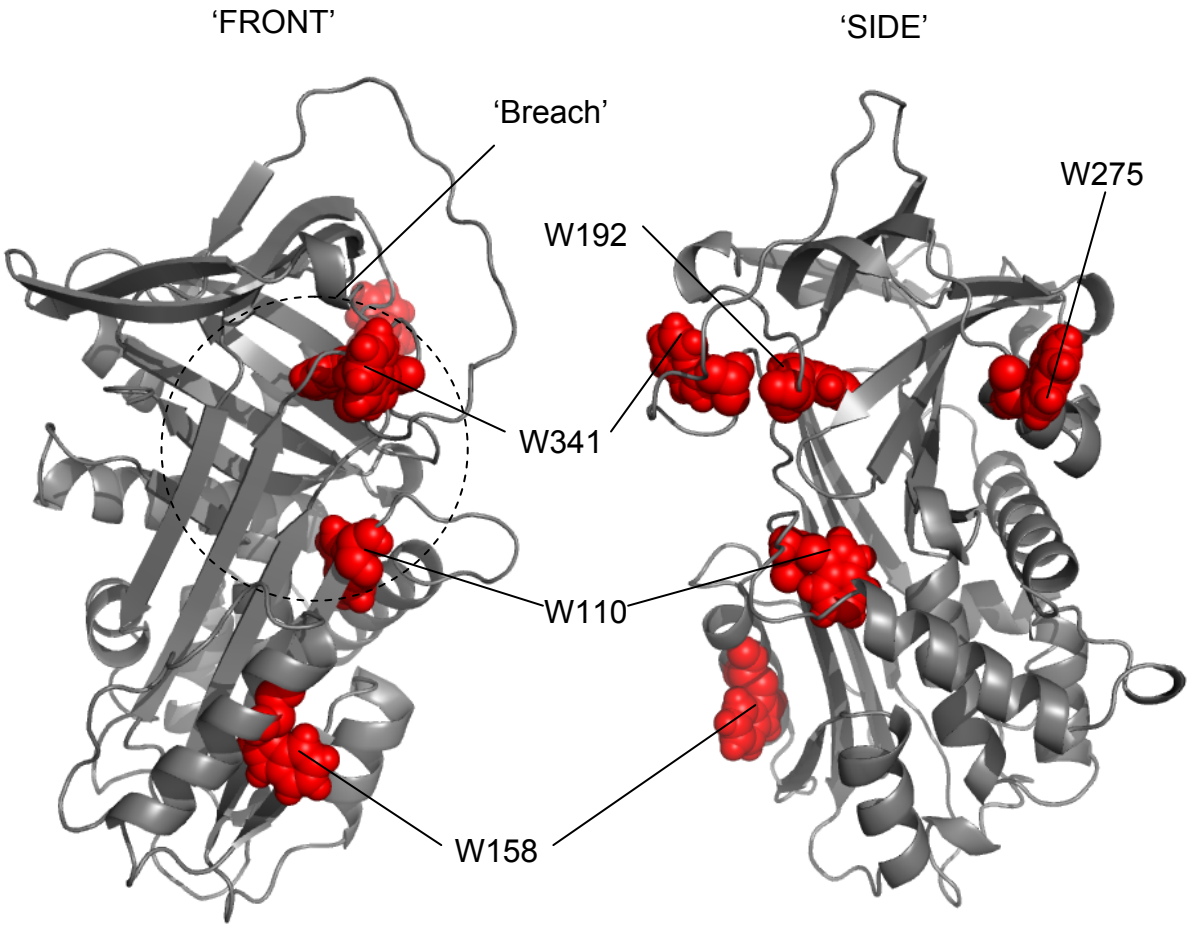


Figure 3.2: HSP47 homology model showing the locations of five Trp residues.
Trp residues are shown in red in 'spheres' representation.

3.3 Construction, Expression and Purification of Trp-to-Phe (WF) Mutants

WF mutations were introduced as in the HA mutant construction and expressed in *E. coli* BL21(DE3). Their expression and solubility was first assessed to investigate possible destabilising effect of the mutation to the protein. Expression of all mutants were found to be comparable to WT when analysed on 12% Tris-glycine SDS-PAGE gel stained with CBB (Figure 3.3A). Figure 3.3B shows solubility of the mutants compared to that of WT. Replacement of Trp residues with Phe seems to affect HSP47 solubility greatly, in contrast to HA mutants. The most dramatic effect is observed with W192 mutation, when the protein expressed has very low solubility and not detectable on the gel. W110F has an acceptable solubility for purification, while W158F also has a low solubility, but detectable in the soluble fraction of *E. coli* lysate. The variations in solubility upon mutation probably relates to the degree of solvent exposure. W192 is the least exposed, which explains the very low solubility of the Phe variant. W158 and W275 are probably less exposed than predicted, based on the adverse effect of mutations on their solubilities.

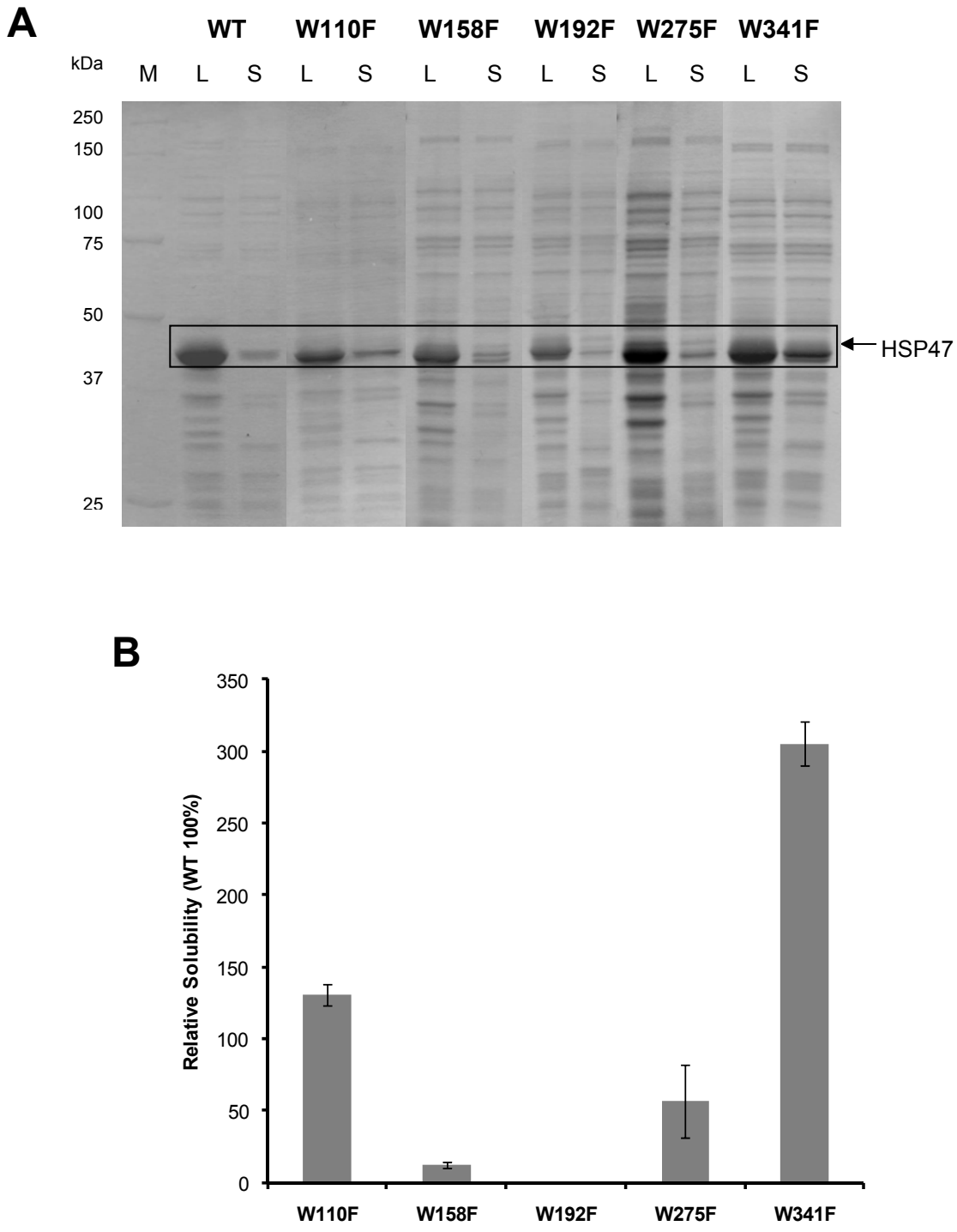


Figure 3.3: Expression and solubility of HSP47 WT and WF mutants.
A. Total protein content (L) and soluble protein fraction (S) WT and mutants expressed in *E. coli* BL21(DE3) separated on 12% SDS-PAGE and stained with CBB. HSP47 bands are highlighted in black square. M: Precision Plus Protein Standards (Bio-rad). **B**. Solubility of mutants relative to WT (100%, n=3). W192F is hardly soluble in the same buffer conditions as used with the other mutants. Quantification was done using NIH ImageJ software after immunodetection with anti-HSP47 antibody.

Glycerol has widely been reported to increase protein solubility when added at a certain concentration, especially in protein crystallisation when a high concentration is often required as it reduces buffer polarity thus stabilises the hydrophobic regions (Gerlsma, 1968; Sedgwick et al., 2007; Simpson and Kauzmann, 1953; Tanford et al., 1962). It has been suggested that the presence of glycerol causes the alteration in the effective protein size because of a small expansion of its radius, or by the enhancement of a layer of solvent limiting the distance of closest approach between the protein molecules (Farnum and Zukoski, 1999). For this reason, glycerol was added during cell lysis to facilitate solubilisation of mutants W158F, W192F and W275F. Effect of glycerol addition to the solubility of the mutants is shown in Figure 3.4. Addition of glycerol at 15%(v/v) has different effects on the solubility of these mutants. W192F shows a two-fold increase to that of WT, while W158F only shows slight increment of its solubility. A reverse effect is observed with W275F, where it was found to be less soluble. Hence, HSP47 WT, W110F, W275F, and W341F were purified without glycerol in the lysis buffer, while W158F and W192F were lysed with 15%(v/v) glycerol in the buffer. The mutants were purified using HisTrap Ni-NTA column as the mutation is not expected to perturb its binding and release property. This is supported by the gelatin-agarose binding assay (Figure 3.5) where all soluble mutants possess more than 50% binding ability compared to that of WT. Replacement of gelatin-agarose column with Ni-NTA column offers a much faster purification at higher flow rate with a higher protein binding capacity. The yields of W275F and W158F after purifications were found to be very low, so these mutants were not investigated further. Otherwise, a typical yield was between 0.18 to 0.20 mg/mL after imidazole removal.

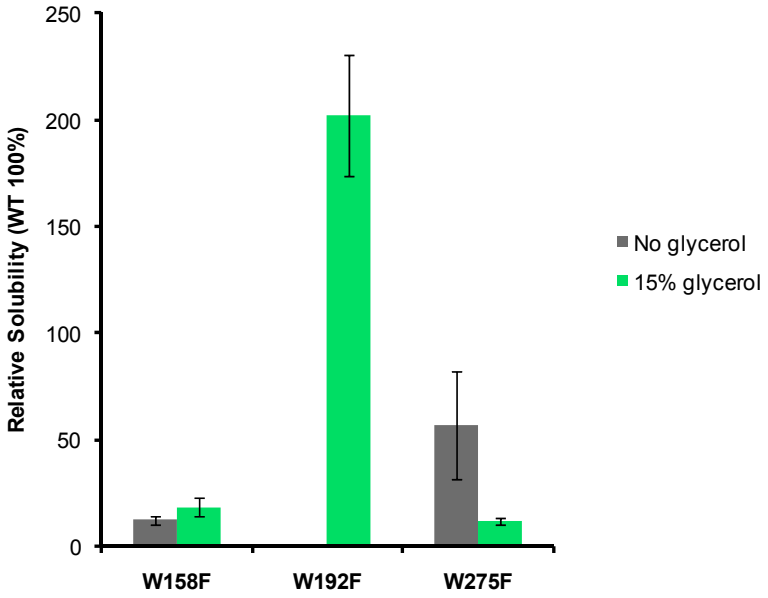


Figure 3.4: Effect of 15%(v/v) glycerol on HSP47 WF mutants solubility.
Solubility is relative to WT (100%, n=3). Quantification was done using NIH ImageJ software after immunodetection with anti-HSP47 antibody. Glycerol does not seem to change W158F solubility, increases significantly the solubility of W192F, but has an adverse effect on the solubility of W275F.

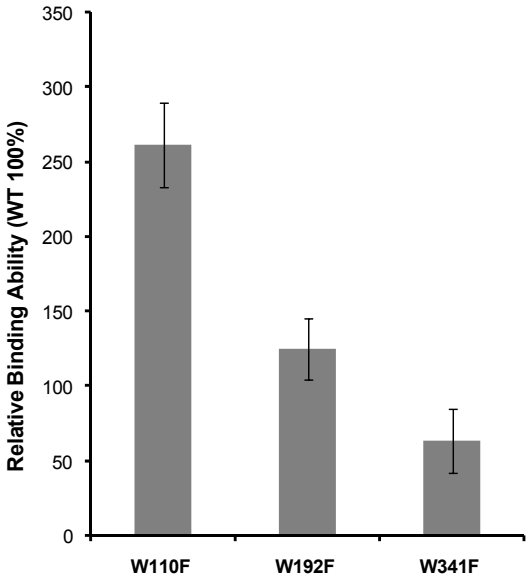


Figure 3.5: The binding ability of soluble WF mutants constructed to gelatin immobilised on agarose beads.
Binding ability is relative to WT (100%, n=3). Binding assay was performed in lysis buffer containing 15% glycerol. Quantification was done using NIH ImageJ software after immunodetection with anti-HSP47 antibody.

3.4 Structural Analysis of Purified Mutants

Effects of WF mutation on the secondary structure of HSP47 are first investigated using far-UV circular dichroism (CD). Comparison of the spectra of all purified mutants with that of WT revealed that there is no significant difference or perturbation in the shape of the observed spectra (Figure 3.6A), except that of W110F. There is also a slight variation in signal intensity of the spectra that belong to mutants W192F and W341F, which suggests that Trp residues in HSP47, to some extent, contribute to the overall CD signal. This has also been noted elsewhere, for example in the mutational work performed on Trp residues in barnase. It reveals the significant difference in the far-UV CD spectra of the mutants compared to WT (Vuilleumier et al., 1993). Similar effects are also seen when other aromatic amino acids are mutated, i.e. Tyr and Phe. In the case of dihydrofolate reductase (DHFR), all but one of Trp-to-Phe mutants constructed also show slightly different far-UV CD spectra compared to WT (Ohmae et al., 2001). One reason proposed for decreased CD signal characteristic of mutant DHFR relates to the formation of exciton coupling between two adjacent Trp residues that contribute to CD signal. Exciton coupling is simply defined as interaction between two electronic transition in different groups, or also called μ-μ coupling (Strickland and Beychok, 1974). Mutation of either residue in the pair diminishes the signal. Aromatic side chains have also been suggested to contribute to CD signal *via* interactions with other side chains such as amides, carboxylate, guanidino and even with peptide bonds of the main protein chain. Among the factors that determine these interactions include the dipole strength and electronic transition moment of the functional groups, and also the distance between the groups (Strickland and Beychok, 1974). The distance referred to for such interactions in barnase is between 6-8 Å (Vuilleumier et al., 1993). How exactly the excitonic pair is formed by Trp residues in HSP47 is difficult to predict based on the homology model alone. Oservation of the model suggests that there are plenty of possible interactions between aromatic side chains of HSP47 which makes the theory plausible (Figure 3.7). In case of W110F, there is a possibility of slight structural perturbations, but it does not affect its overall activity as this mutant is still capable of binding to gelatin as good, as WT (Figure 3.5).

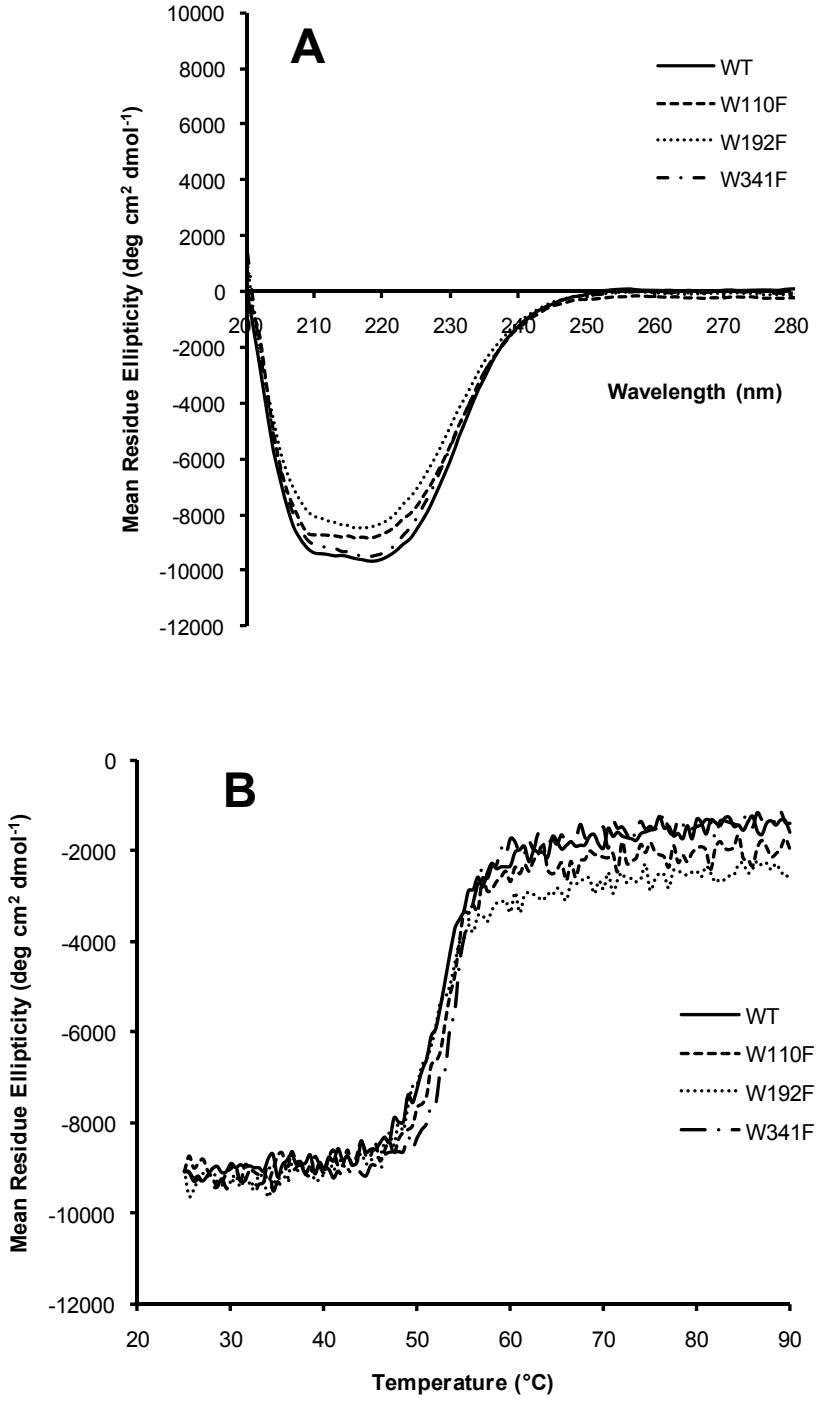


Figure 3.6: Far-UV CD spectra (A) and thermal denaturation profiles (B) of HSP47 WT and WF mutants.
Concentration was fixed to 2 μM and measured in 50 mM sodium phosphate buffer pH 8.0 with 100 mM NaCl.

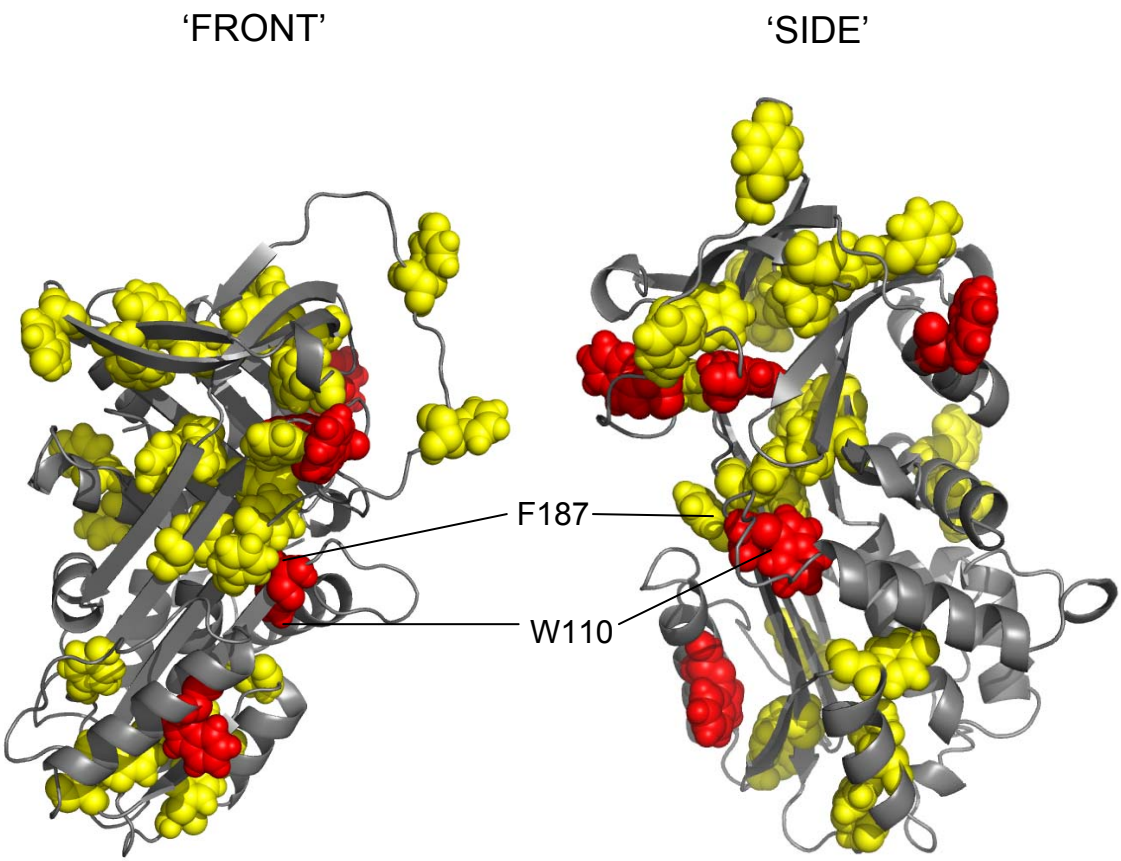


Figure 3.7: 'Front' and 'Side' visualisations of HSP47 homology model showing all five Trp residues (red) and all Tyr and Phe residues (yellow).
The predicted closest distance between the side chains of W110 and F187 is 8.5 Å. Aromatic residues are shown in 'spheres' representation. Atomic distance computed using Deepview Swiss-pdbViewer v4.0.1.

In order to investigate whether the slight spectral difference is accompanied by perturbed structure and hence HSP47 stability, CD temperature titration experiment was conducted. Based on the melting profile, the stability of WF mutants is not significantly perturbed by the mutations. Compared to HSP47 WT, which has a transitional midpoint (T_m) of 52 ± 0.5°C, the T_ms for the mutants differ only in the range of 2°C (W110F 53 ± 0.5; W192F 52 ± 0.5; W341F 54 ± 0.5°C) as shown in Figure 3.6B. This indicates that there is no major perturbation in structure or destabilisation due to the mutation occurs.

3.5 Intrinsic Fluorescence Titration of HSP47 WT

Intrinsic protein fluorescence is observed when a chromophore-containing amino acid (such as Trp and Tyr) absorbs light and becomes excited from ground electronic state (S_0) to higher electronic state vibrational level (S_1). The molecules then relax rapidly to the lowest vibrational level of S_1, a process called internal conversion (generally occurs at 10^{-12} s or less) with the energy is lost to the surroundings as thermal energy. From the lowest energy vibrational state of S_1, the molecules then return to the higher excited vibrational ground state level before reaching thermal equilibrium. This process results in the emission of radiation or lower energy and longer wavelengths, called fluorescence. Some energy also dissipates in non-radiative ways (Lakowicz, 2006; Pain, 2004).

Tyr and Trp absorb light around 280 nm with quantum yields (Q) of 0.1 and 0.2 respectively. Q is simply defined as the number of emitted photons relative to the number of absorbed photons. The Q value for Phe is very low at only 0.04, and it also has the shortest absorption and emission wavelengths. This is why protein fluorescence study usually involves only Tyr and Trp residues. Meanwhile, the fluorescence of Tyr occurs at around 303 nm when excited at 280 nm, and it is relatively insensitive to solvent polarity. Trp, on the other hand, emits at around 350 nm (in water) and highly sensitive to solvent polarity and its local environment. Trp residues in protein can also be specifically excited at 295 nm to minimise absorption by Tyr residues (Lakowicz, 2006).

Fluorescence spectra of mouse HSP47 WT at different pHs are shown in Figure 3.8. Spectra recorded at nearly neutral pH show an intensity maximum at 328 ± 1 nm (λ_{em}), indicating that the Trp residues are, on average, packed in a less polar environment,

compared to fully exposed residues (Pain, 2004). As pH decreases down to 6.5, the fluorescence intensity (I_f) is relatively unchanged. But as pH is lowered to 6.0 or further, a red-shift starts to occur accompanied by an increase in intensity. Hence, it is possible that at this pH, a structural rearrangement is triggered that changes the local environment around which the Trp residues are located. This exposes them to a more polar environment as can be seen from the shift of λ_{em} to higher wavelength. The trend in shifting continues as titration progresses to pH 5.7, when the intensity reaches its peak. At this pH the λ_{em} has shifted to near 335 nm. Further addition of acid decreases the intensity. As the fluorescence decay of tryptophan is known to be stable in the pH range of 8-4 (De Lauder and Wahl, 1970), the quenching is probably caused by neighbouring side chains. It is probable that the ionised His imidazole side chain quenches proximal Trp residues that come to close contact due to the structural rearrangement (Czabotar et al., 2004; Vos and Engelborghs, 1994; Willaert and Engelborghs, 1991). Aggregation was not observed in this case and the concentration before and after titration was not significantly changed. The changes in λ_{em} and I_f can then be plotted against pH to observe the behaviour more clearly (Figure 3.9).

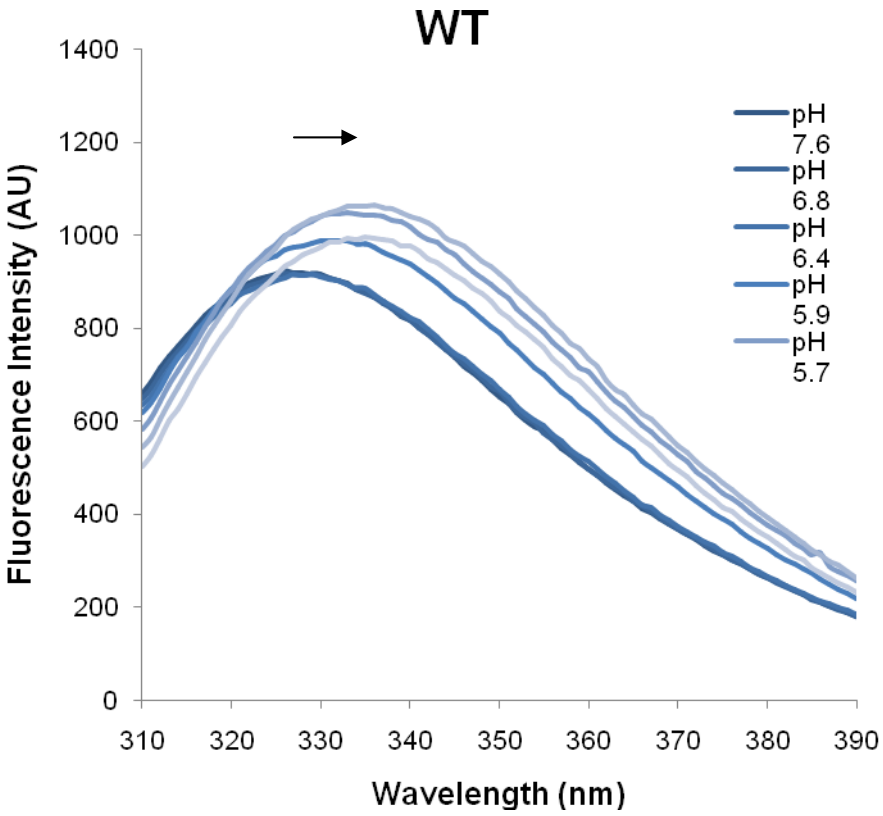


Figure 3.8: Fluorescence emission spectra of HSP47 WT recorded with an excitation wavelength of 295 nm.
Arrow indicates the shift in maximum wavelength during titration. Concentration was fixed at 1 µM in sodium phosphate buffer pH 8.0 with 100 mM NaCl, at 25°C.

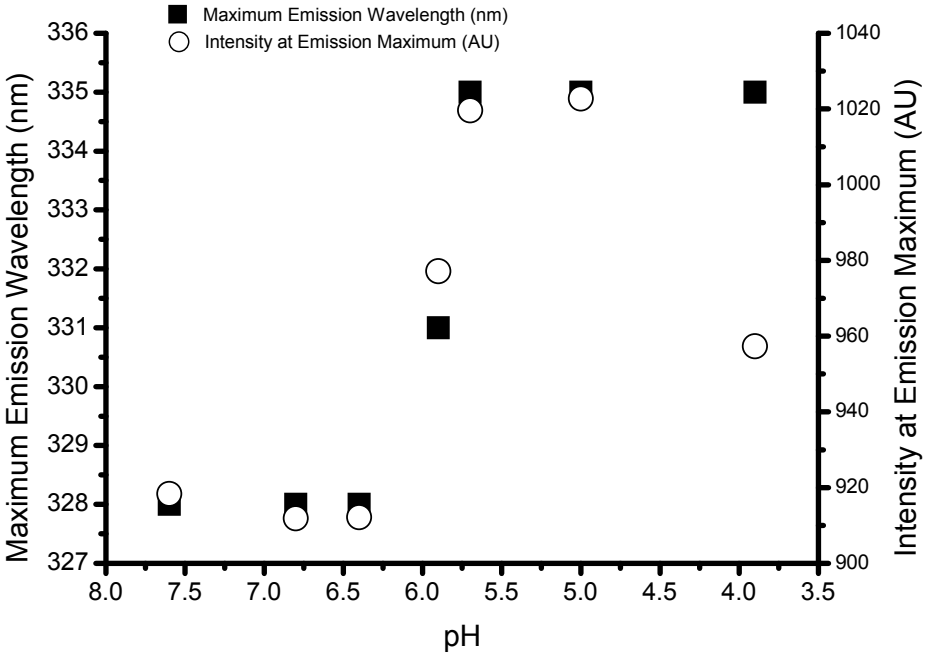


Figure 3.9: The plot of the change in maximum emission wavelength (λ_{em}) and fluorescence intensity (I_f, at the initial λ_{em}) against change in pH of HSP47 WT.
Values were taken from individual spectra recorded during pH titration. Initial λ_{em} was taken as it gives much better data for plotting and visualisation purposes. λ_{em} values include ±0.5 nm instrumental errors.

The trend of the change in λ_{em} is comparable to HSP47 structural change during pH titration observed with CD as described in the previous chapter, though the actual mechanism that gives rise to this might not be similar. But the most significant change can be seen between pH 6.3 and 5.5, which is also in the His imidazole ionisation range. Based on the change in λ_{em} value, three apparent phases can be observed: (Phase 1) Minimal changes in maximum emission wavelength between pH 8.0 to approximately 6.3. (2) A sudden shift in maximum wavelength between pH 6.3 and 5.5 to longer wavelength indicating a more exposed environment of the intrinsic Trp residues. (3) A constant maximum emission wavelength as titration progresses further. Similarly while monitoring the maximum fluorescence emission intensity at the initial λ_{em}, three distinctive phases can also be identified: (Phase 1) The gradual decrease in intensity as the pH decreases from 8.0 to 6.3. (2) The sudden increase in intensity as pH reaches 6.3, until about pH 5.7. (3) The gradual decrease as titration continues below pH 5.7 probably due to quenching as explained above. The relationship between the change in λ_{em} and the change in secondary structure monitored using CD is difficult to predict due to the complex nature of Trp fluorescent emission. But overall it can be suggested that the main rearrangement of HSP47 during pH transition can be monitored using intrinsic Trp fluorescence spectroscopy. It is from the so called *alkaline* (non-protonated) state (>pH 6.3) to the *acid* (protonated) state (<pH 5.5), which occurs in the His imidazole pK_a window. This rearrangement is accompanied by the release of binding to collagen (as demonstrated using collagen mimetic peptide in Chapter 2).

Reverse titration was also performed on HSP47 WT from acidic to alkaline pH with the addition of NaOH in small aliquots to investigate the reversibility in the change of λ_{em} and intensity. The change in λ_{em} is totally reversible suggesting the reversibility of HSP47 pH transition (Figure 3.10A). In contrast, maximum emission intensity progressively decreases during reverse titration probably due to irreversible quenching effect by ionising side chains (Figure 3.10B). Based on these observations, the construction of Trp to Phe mutants seemed appropriate to indirectly probe a more specific locality at which this transition occurs based on the change in λ_{em}. It is concluded in Chapter 3 that from alanine mutagenesis of surface His residues, the important residues participating in pH-switch mechanism are probably localised at the top of β-sheet A, the so called 'breach' cluster. Given this finding, it is also possible that the change in structure only centred on that particular cluster without affecting the whole protein structure. For this reason, each of the Trp residues was mutated one by one

and the fluorescence behaviour was observed. This eliminates the need to introduce more than one mutation that may be detrimental to the protein stability. The mutation will replace the chromophore, with the least possible effect on the structural integrity at that mutation site.

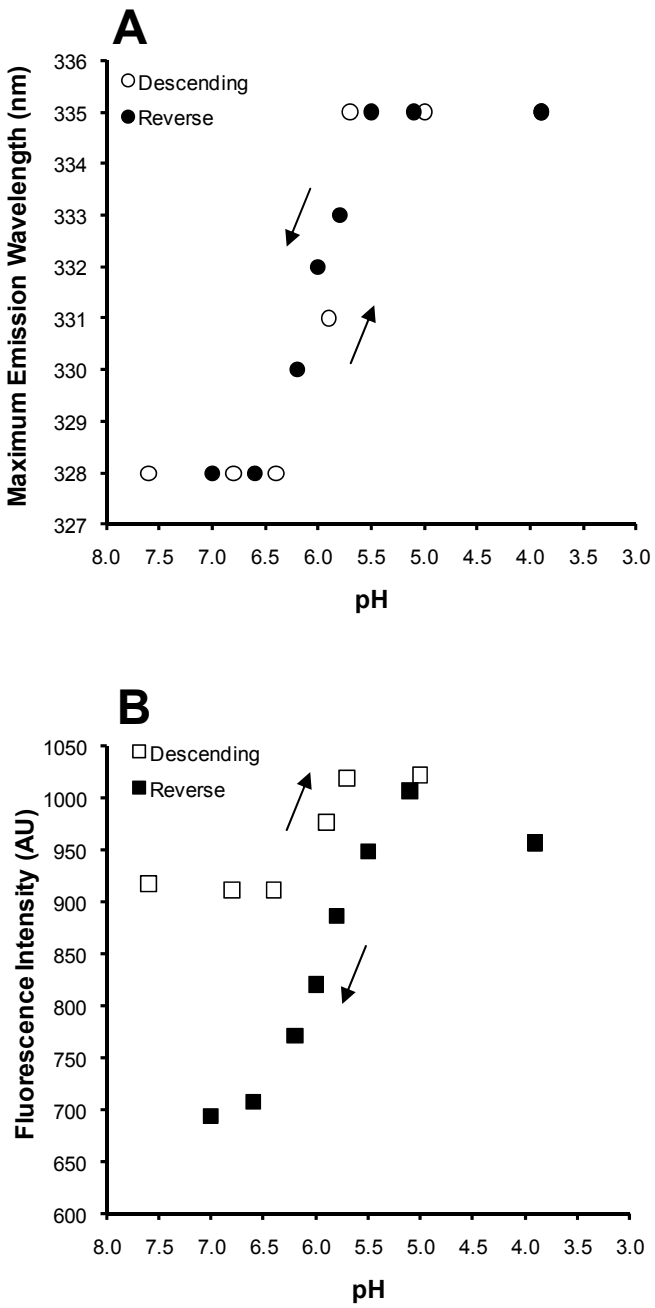


Figure 3.10: Reversibility of the change in (A) maximum emission wavelength (λ_{em}) and (B) fluorescence intensity (I_f) at the initial λ_{em} of HSP47 WT.
Titration was initially done from pH 8.0 to 4.0 (*Descending* plot) before bringing the pH back to 8.0 again (*Reverse* plot) with concentrated NaOH. pH was raised after acid titration with the addition of concentrated NaOH. Values were taken from individual spectra recorded during pH titration. All curves are comparable after several repeat titrations.

3.6 Fluorescence Characterisation of HSP47 WF Mutants

CD spectra of the mutants do not show any significant difference from WT, apart of those attributed to exciton coupling between aromatic side chains, as explained in Section 3.4. This suggests that there is no perturbations in the secondary structure of mutants W110, 192 and 341 (Figure 3.6A). Then by scanning the fluorescence spectra of the mutants at the same excitation wavelength and at the same concentration, variations in the I_f and λ_{em} are apparent (Figure 3.11). Removal of W110 causes a drop in fluorescent intensity by around 10% to that of WT while W192 and W341 probably contribute to similar effect on fluorescence as removal of either chromophore decreases I_f by nearly 20% respectively. The emission spectrum of W110F shows a blue shift which suggests a slight decrease in polarity of the environment of the other four Trp residues or either one of them as a result of W110 mutation. The opposite of this has been observed in the mutagenesis study of *E. coli* DHFR (Ohmae et al., 2001). This also explains the slight perturbation seen in the CD spectrum of W110F (Figure 3.6A).

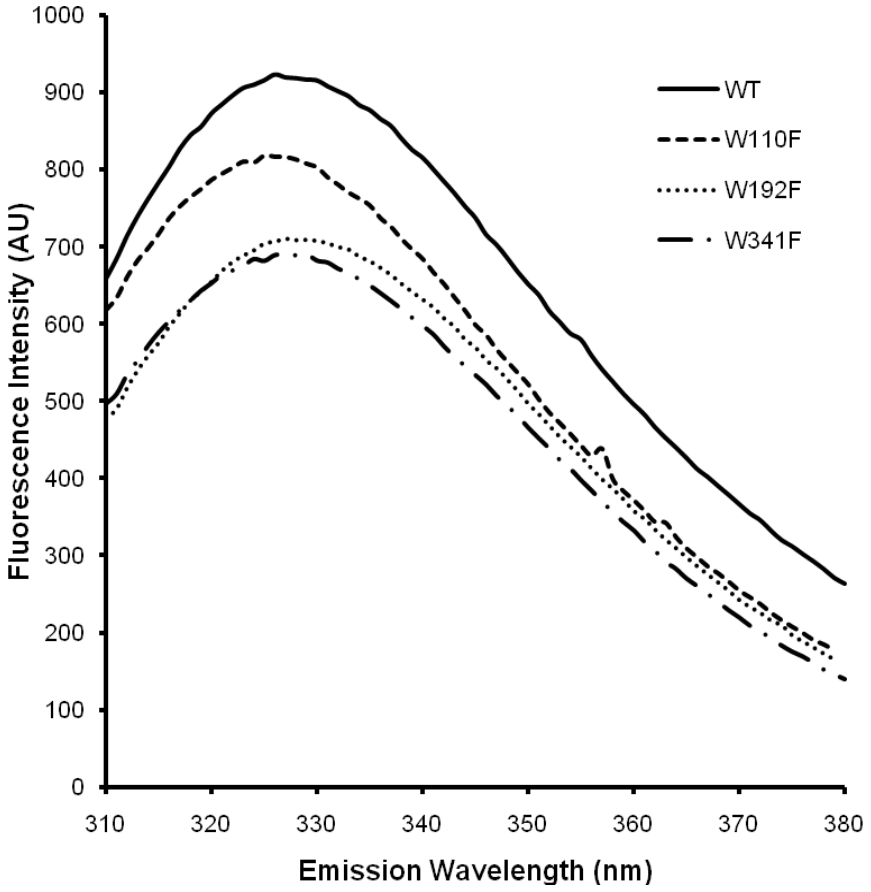


Figure 3.11: Fluorescence emission spectra of HSP47 WT and WF mutants recorded with an excitation wavelength of 295 nm at alkaline pH.
Concentrations were fixed at 1 µM in sodium phosphate buffer pH 8.0 with 100 mM NaCl, at 25°C.

Fluorescence pH-titration of HSP47 WF Mutants

Similar to the experiment carried out on HSP47 WT, the change in maximum emission wavelength (λ_{em}) of the mutants were plotted as a function of pH (Figure 3.12). As demonstrated earlier, this parameter was chosen to probe the conformational rearrangement of HSP47 during pH titration. Monitoring of λ_{em} against pH for WT shows a transition in the collagen-dissociation pH region, and unlike fluorescent intensity, is independent of collisional quenching. The changes in λ_{em} against pH for the mutants were then fitted to single pH ionisation model, with all fits converged well (Figure 3.12). The results of the fitting is summarised in Table 3.1. The pK_a for W110F and W192F agrees well with the pK_a of HSP47 WT, while W341F gives a slight pK_a shift which suggests perturbations in the fluorescence-probed structural transition. The curve for W192F is also slightly deviates from that of WT, though this still could not be fully verified from our current data and tools for analysis.

W341 along with W192 are predicted to be located in the 'breach' cluster from the homology model. It is probably due to their vicinity to the 'breach' His residues that participate actively in triggering pH-controlled release that causes these mutants, especially W341F to experience slight perturbations during pH titrations.

In addition, the fact the λ_{em} data could be fitted to single ionisation model suggest that there is a possibility of single Trp residue mainly responsible for the change. W110, W192 and W341 can be ruled out as mutation of these residues does not abolish this behaviour. W275 is probably the residue involved due to its predicted location at helix H (hH) opposite of the 'breach' cluster. One of the two proposals for HSP47-collagen binding interface also involves the hydrophobic region between β-sheet B, hA and hG/hH (Dafforn et al., 2001; Davids et al., 1995). This means that there is a possibility that W275 experiences the largest change in environment as HSP47 rearranges its structure during pH change. Unfortunately, this currently could not be determined due to the low solubility of this mutant construct.

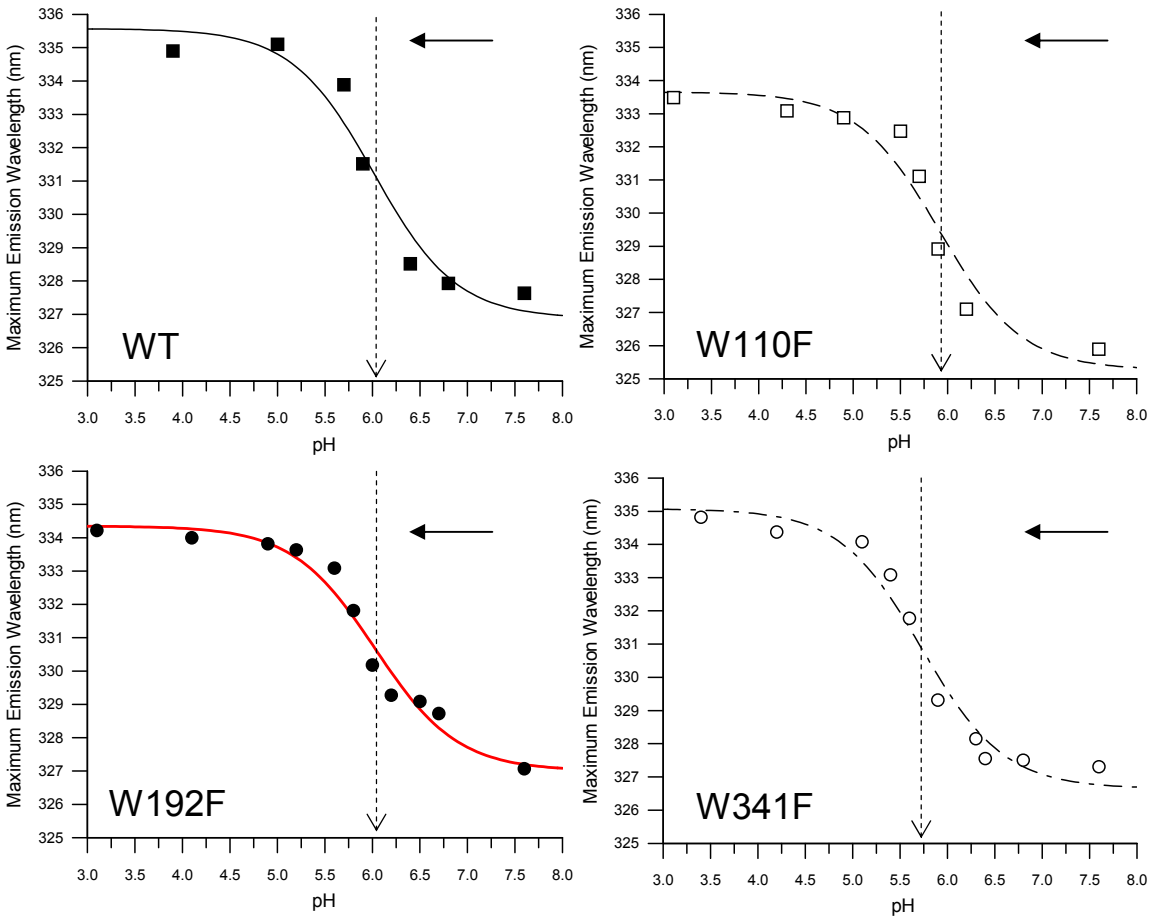


Figure 3.12: The changes in maximum emission wavelength of HSP47 WT and WF mutants during pH titration.
Values were taken from averaged individual spectra recorded during pH titration. Curve fitting was performed with GraFit Pro v7.0.0. All curves are comparable after several repeat titrations. Arrows represents the progress of titration from high to low pH, while broken arrows indicate the inflection points.

Table 3.1: Summary of data for HSP47 WT and WF mutants.

Mutation	Melting temperature ± 0.5°C	Degree of residue solvent exposure[1]	Fitting data
			Inflection point (pH)[2]
WT	52.0	-	6.02±0.18
W110F	53.0	30%	5.91±0.16
W158F	*	40%	*
W192F	52.0	3%	6.03±0.10
W275F	*	90%	*
W341F	54.0	30%	5.73±0.11

*Data for W158F and W275F could not be collected due to low solubility or low purification yield.
[1]Solvent exposure calculated from mouse HSP47 homology model, using ASA-View program (http://gibk21.bse.kyutech.ac.jp/netasa/asaview/). 100% represents a completely exposed residue.
[2]Inflection point was calculated by fitting the change in λ_{em} with pH to a single ionisation model, with GraFit Pro v7.0.0.

It has been suggested in Section 3.5 that the change in λ_{em} upon titration with pH is most probably due to the change in the microenvironment of the Trp residues due to structural rearrangement. Given the complex nature of intrinsic Trp fluorescence emission in proteins, there is also a possibility of interactions between a Trp residue and the neighbouring His residue that are mainly responsible for this instead of the actual structural change. The interaction between the imidazole and indole side chains of His and Trp residues respectively was reported by Shinitzky and Goldman (Shinitzky and Goldman, 1967), where the protonated His was found to quench the fluorescence of Trp moiety. Since then, this Trp-His pair has been demonstrated in various mutagenesis studies. For example in the case of barnase, where mutations of either residue in the neighbouring Trp-His pair abolished the pH dependent change of the intrinsic fluorescence (Loewenthal et al., 1991). With the same protein, Trp was later found to increase the pK_a of a neighbouring His residue (Loewenthal et al., 1992) thus increasing its stability. The distance between the side chains of the two residues are 4.0 Å. The ability of Trp-His interaction to increase protein/peptide stability was then shown in the study using short α-helical peptides where certain positions of Trp and His residues along the peptide are preferable for stabilising effect (Fernández-Recio et al., 1997).

In the case of the hexadecapeptide anantin, protonated imidazole was found to be a better quencher of Trp fluorescence that the non-protonated form (Vos and Engelborghs, 1994) while the study with T4 lysozyme showed the opposite (Van Gilst and Hudson, 1996). This was suggested to be due to the non-ideal stereo arrangement of the neighbouring Trp and His residues during pH change that causes the fluorescence emission to increase instead of decrease.

In order to investigate the possible presence of important Trp-His pair in HSP47, the homology model was used to compute the closest distance between them. Mutation of either residue in the pair is known to completely abolish the pH transition when probed using intrinsic Trp fluorescence (Loewenthal et al., 1991). This means that W110, W192 and W341 probably do not form pairs with neighbouring His residues that are ionised during HSP47 pH transition as mutation of each of them retains this behaviour (Figure 3.12). As for W158, the closest distance to a surface His residue is predicted to be 21.8 Å (H135, side chain-side chain) while the closest His residue is H140 (10.1 Å), while for W275 the closest surface His residue is H244, which is also the closest predicted His residue (20.5 Å) (Figure 3.13). As the effective distances suggested for a His and Trp residues to form pair is less than 8.0 Å (Jas and Kuczera, 2004), these suggest that the pairing between Trp residue and an important His residue might not exist in HSP47. These further suggest that the change in fluorescence properties observed during titration is due to the change in the environment of the Trp residues.

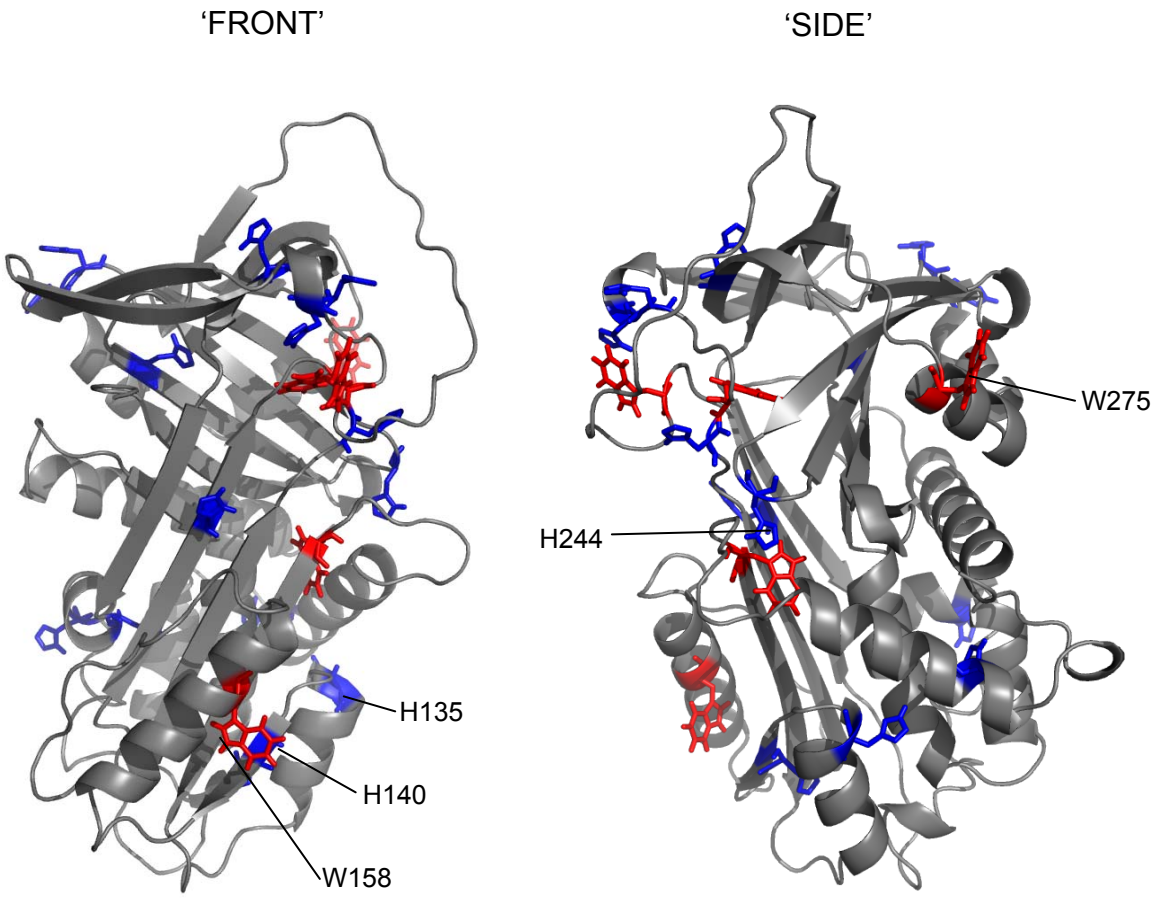


Figure 3.13: HSP47 homology model showing all five Trp residues (red) and all His (blue).
The predicted closest distance between the side chains of W158 and H135 is 21.8 Å while with that of H140 is 10.1 Å. The distance between W275 and H244 is 20.5 Å.
Trp and His residues are shown in 'sticks' representation. Atomic distances computed using Deepview Swiss-pdbViewer v4.0.1.

3.7 Conclusions

In the previous chapter, mutational work on surface His residues revealed that perturbations in biophysical property are mostly observed when residues in the 'breach' cluster are mutated. This suggests the localisation of the residues plays a crucial role in triggering and controlling the transition from the collagen binding state (alkali) to its non binding state (acid). This chapter was aimed at further observing this mechanism by means of intrinsic Trp fluorescence change. Trp residue is widely used in observing structural changes in proteins as it responds to change in its microenvironment, and can be excited independently and separate from that of Tyr or Phe excitation. It turns out that much information could be obtained with this technique, especially for HSP47 which is among the least understood members of serpin superfamily. Firstly, even though Trp-to-Phe mutation is considered to be the most conservative mutation, it has the potential to decrease significantly the stability of the construct when expressed in *E. coli,* as in the cases of W158F and W275F mutations. Mutation of W110 seems to affect the environment of the other Trp residues. Finally, Trp reporter in the 'breach' cluster is possibly located in or close to the dynamic region during HSP47 pH-switch transition. Any behaviour that would go against this proposal was also not observed during the investigation on WF mutants. These further support the findings presented in Chapter 2, which were used to construct a model on HSP47 pH-switch mechanism as explained in Chapter 4. The results presented here could actually be improved should the soluble forms of W158F and W275F are obtained. This could be done by employing one of the techniques available for increasing protein solubility, *e.g.* engineering of surface hydrophobic residues or addition of fusion tags. Double and triple WF mutations have also been attempted to obtain a better identification of Trp residues responding the most to pH-change. This is also restricted by the poor solubility of the mutants constructed; hence it is not reported in the thesis.

Chapter 4

Results and Discussion: A Proposed Model for HSP47 pH-switch Mechanism

CHAPTER 4: A PROPOSED MODEL FOR HSP47 pH-SWITCH MECHANISM

4.1 Introduction

Results from the His-to-Ala (HA) and Trp-to-Phe (WF) mutations are being put together in this chapter to propose a viable model for the pH-triggered release of HSP47 from its only substrate collagen. Full understanding of this behaviour is particularly desirable as it is currently widely-accepted that this is the only mechanism governing HSP47/collagen binding and release. Unlike many serpin superfamily members, HSP47 does not possess any protease inhibitory activity, which suggests a different reactive centre loop (RCL) conformation. But serpins usually share a high similarity in the main structural scaffold, especially the highly conserved β-sheet A, B and C and the α-helices. This similarity is useful in computational prediction of a reliable homology model of HSP47 in which known structures of serpins can act as templates. This provides critical information for proposing a binding and release mechanism of HSP47 while the three dimensional structures in its bound and unbound states have not been worked out yet due to difficulties in obtaining the crystal structure. In this chapter, previous reports on serpin structures and functions are reviewed and compared with our data for HSP47 presented Chapters 2 and 3 to provide a working model on HSP47 pH-switch mechanism. This chapter also includes discussion on other proteins carrying similar functions as HSP47, which could enrich our understanding on the behaviour of this molecular chaperone.

4.2 Role of the 'Breach' Region in Serpin Family Members and HSP47

In Chapter 2, results obtained from our work suggest that both single and double mutations of His residues in the 'gate' and 'shutter' clusters do not significantly affect HSP47 pH-triggerable release and thus unlikely to participate actively in the process. In addition, our study based on intrinsic Trp fluorescence titration in Chapter 3 finds these two clusters less dynamic compared to the 'breach' cluster. The mutation introduced at the 'breach' cluster perturbs normal HSP47/collagen interaction significantly. H191, which is predicted to be located at the s3A-s4C loop close to the A-sheet, suggests a possibly critical role in HSP47

pH-shift. Trp residues located in vicinity to the cluster are predicted to experience the largest change in their environment during secondary structural rearrangement, implying that the 'breach' area is very dynamic, compared to the other two regions of HSP47.

It is interesting to note that the 'breach' area is also considered to be important for various serpin functions, for example α_1-AT. Mutation of glutamic acid (Glu, E) residue at position 342 to lysine (Lys, K) results in the protein being accumulated in the ER of hepatocytes, impeding its transport to Golgi for further processing (Bathurst et al., 1984). E342 is located at the *N*-terminal end of strand 5 of β-sheet A, known as the proximal 'hinge' region of serpin (Kaiserman et al., 2006). This region is the first RCL point-of-entry during insertion into β-sheet A for the conversion from native inhibitory to latent form. E341K mutation increases the propensity of α_1-AT to polymerise probably due to the loss of side chain interactions, *via* insertion of RCL of one molecule into β-sheet A of another molecule (so called 'head' to 'tail' polymerisation) (Lomas et al., 1992).

Another highly conserved 'breach' residue is tryptophan at position 175 of plasminogen activator inhibitor-1 (PAI-1) (Blouse et al., 2003). The residue, which is found in virtually all known serpins, is located at the *N*-terminal end of strand 3 of β-sheet A (s3A). W175 of PAI-1 is homologous to W192 of mature mouse HSP47 in our predicted model. As observed for HSP47, W175F mutation does not cause any significant perturbations in the folding of PAI-1 upon expression. PAI-1 is known to exhibit a spontaneous transition from 'stressed' to 'relax' conformation (*S*-to-*R*) under certain physiological conditions (Carrell et al., 1991; Mottonen et al., 1992). The mutant was found to have a stabilised 'stressed' conformation, even though the association of target proteinase is unaffected. When reacted with several substrate proteinases, it was found that W175F/PAI-1 exhibits a significant drop in the limiting rate of one proteinase inhibition in particular, namely the tissue-type plasminogen activator (tPA). It was postulated that this is due to the disruption of amino acid interactions in the breach region of PAI-1, that are responsible for tight exosite interactions with the basic residues on tPA loop and as a result, controlling the reversible deacylation of P1′ amino acid of RCL and preventing its release from the substrate tPA. The disruptions restrict the forward reaction rates and decrease the limiting rate of tPA inhibition by mutant PAI-1 (Blouse et al., 2003). Even though this behaviour is not studied in HSP47, it demonstrates the importance of the 'breach' cluster to inhibitory serpins, and very likely to any serpin family members in general.

In another study on PAI-1, various mutations introduced in the 'breach' cluster and its adjacent region cause a change in activity of the protein (Sui and Wiman, 2000). Other than the 'breach' cluster, the region targeted is the β-sheet B, particularly strands 2 (s2B) and 3 (s3B). Among the notable effects of 'breach' residue modifications is the increased stability of the resulting mutant, such as Y228S and T232F which are located on the s2B-s3B loop. This is probably due to restriction on RCL insertion caused by the change in the environment around those residues. A more significant finding is the involvement of H229 in the pH stability effect of PAI-1. PAI-1 was reported to have increased stability at slightly acidic pH, than at neutral or alkaline pH. Like HSP47, histidine imidazole group is suggested to be responsible for this (Kvassman et al., 1995). Mutation of H229 decreases the stability of PAI-1 at acidic to about two-fold to that of at alkaline pH, instead of the normal ten-fold increase in stability (Sui and Wiman, 2000). H229 is located in the s2B-s3B loop, very close to the predicted location of H191 in HSP47 at the *C*-terminal end of s3A in our homology model. Although no pH-switch function has been observed with PAI-1 so far, the fact that the 'breach' residue ionised to help stabilise PAI-1 at acidic pH is a notable finding. This could be related to our finding that the His residues in the 'breach' cluster in HSP47 play an important role in the mechanistic behaviour. The pH regulation of PAI-1 is said to be important in stabilising PAI-1 in the cells (pH approximately 6.8) in the event of blood flow obstructions to active tissues that lead to acidification *via* CO_2 accumulation and metabolic shift towards fermentation (Kvassman et al., 1995).

As also discussed in Chapter 3, in antichymotrypsin (ACT), the 'breach' residue W194 is shown to be the main residue that controls the rate of ACT polymerisation (Pearce et al., 2007). W194 is located at the *N*-terminus of RCL, which is also the *C*-terminal end of s3A. Again, this is the exact region where important residues for serpin function should be found. Like the residue E342 in α_1-AT (Lomas et al., 1992), W194 of ACT is probably responsible for preventing serpin polymerisation by increasing the kinetic barrier or misfolding.

It is clear from the works described above that amino acids in the 'breach' cluster are indispensable for normal serpin functions. These include preventing polymerisation in the case of α_1-AT and ACT, controlling the formation of latent conformation and proteinase binding in PAI-1 and being responsible for low pH-induced stability in PAI-1. In many other inhibitory serpins, this region is considered to be the first point-of-contact for RCL

translocation before insertion into β-sheet A despite the absence of a more detailed mutational work. At the same time, the 'breach' residues are also responsible for other secondary functions, as proved by the work explained above. For non-inhibitory serpins, there are limited mutational works performed to confirm a similar role of residues in the 'breach' cluster as with their inhibitory counterparts, though a few computational studies reported tend to support this.

An example is in the study of pigment epithelium derived factor (PEDF), its structure has been discovered relatively recently compared to other serpins (Simonovic et al., 2001). It is found as a component of the retinal inter-photoreceptor matrix as well as in the vitreous and aqueous liquid in the adult eye (Simonovic et al., 2001). It is also reported to have neutrophic and antiangiogenic functions. Computational work done on several sequences of this serpin from different species revealed conserved residues in the area predicted to be located in the 'breach' region (Tombran-Tink et al., 2005). For example Glu at position 342, which is homologous to E342 of α_1-AT, is highly conserved in PEDF from all species except one. It is also found to be conserved in many sequences of non-inhibitory serpins too (Tombran-Tink et al., 2005). Given that non-inhibitory serpins probably evolved from their inhibitory relatives, it is likely that some structural-related functions are preserved.

4.3 A Proposed Mechanistic Hypothesis of HSP47 Release from Collagen

The previous chapters have established the significant involvement of His residues in the 'breach cluster' in HSP47 pH-switch behaviour. It is very likely that side chain protonation triggers re-arrangement of secondary structures, making this cluster more dynamic than others. This then triggers the shift in conformation of the region near collagen binding interface causing HSP47 release.

Homology model predicted that three solvent-accessible His residues are located in the 'breach cluster, namely H191, H197 and H198. A fourth residue, H244 may also be located in this cluster and participate in the active mechanism. Analyses on the H244A mutant (as presented in Chapter 2) though suggest otherwise. No significant physical perturbation can be seen in the structural stability or the pH-induced structural transition. For

this reason, all or one of the His residues mentioned above is very likely to be the active residue(s) involved.

H191 is suggested to influence the release of HSP47 from collagen, as its absence triggers dissociation at higher pH, based on the gelatin column and CD experiments as reported in Chapter 2. This does not abolish dissociation altogether, implying that there are some other residues responsible for triggering the release. In the process of identifying these 'auxiliary' residues, mutations were performed on other surface-predicted His residues covering various locations of HSP47. In the 'gate' cluster, mutations of H255 and H256 do not significantly alter the properties of the mutant. This includes its secondary structural content and gelatin/collagen binding and release property. This cluster is located in vicinity to the flexible loop connecting s3C and s4C critical in the conversion of PAI-1 from native to latent structure (Stein and Carrell, 1995). Its opposite location to the 'breach' cluster suggests that it may play an important part, but experimental result reveals otherwise. Finally the 'shutter' cluster, which is found near the centre of sheet A centred at the start of helix B (Stein and Carrell, 1995) accommodating H302 and H135 was found to have little influence on collagen binding/release behaviour of HSP47.

These suggest that only His residues in the 'breach' cluster are responsible for the pH-triggered conformational rearrangement of HSP47 that leads to dissociation from collagen. Other than H191, it is likely that H197 and H198 also participate actively in this process on two bases: first, the vicinity of these two residues to H191 in HSP47 homology model; and second, the effect of mutation of these residues on HSP47 collagen-specific recognition to withstand thorough washing in column experiment. However, the direct involvement of these residues in the release mechanism could not currently be confirmed as the mutants are not available for further analysis.

The proposed mechanism for HSP47 pH-switch action that triggers dissociation from collagen is depicted in Figure 4.1. The simplified diagram only shows the residues pertinent for the proposed mechanism, namely H191, in its non-ionised and ionised states. It is also based on the prediction of surface His residues according to the constructed homology model described earlier in the thesis. At alkaline pH, HSP47 is in its binding state and binds collagen at the binding groove/interface. The interface is probably at a different site than the 'breach' cluster, based on the results from the collagen/gelatin binding assay of the mutants

(Chapter 2). As proton concentration increases, certain His residues begin to be ionised at pH 5.9-6.0. Ionisation triggers electrostatic repulsion of their neighbouring positively-charged residue(s) and exposure of the side chains to more hydrophobic environment. This triggers the formation of a partially non-binding state but is insufficient to proceed to a dissociation of collagen molecule. As pH decreases further to 5.7, H191 is in turn ionised and can now trigger a sufficient conformational rearrangement causing its release from collagen. This is supported by the analyses made with H191A mutant, where it is found to release collagen binding at a slightly higher pH value. This conformational rearrangement is complete as pH reaches 5.3 where full acid conformation is adopted, based on the minimal secondary structural changes at pH lower than 5.3 from CD experiment.

The mechanism is suggested from the observations made from our HSP47 HA mutation work, in particular of H191A. There is also a possibility that full acid state is not adopted in the *cis*-Golgi before HSP47 is recycled back to the ER *in vivo*. This is because the cycle demands rapid structural rearrangement back to alkali, collagen-binding state for HSP47 to continue its chaperone function in the ER. The ER-Golgi transport cycle is illustrated in Figure 4.2.

WT

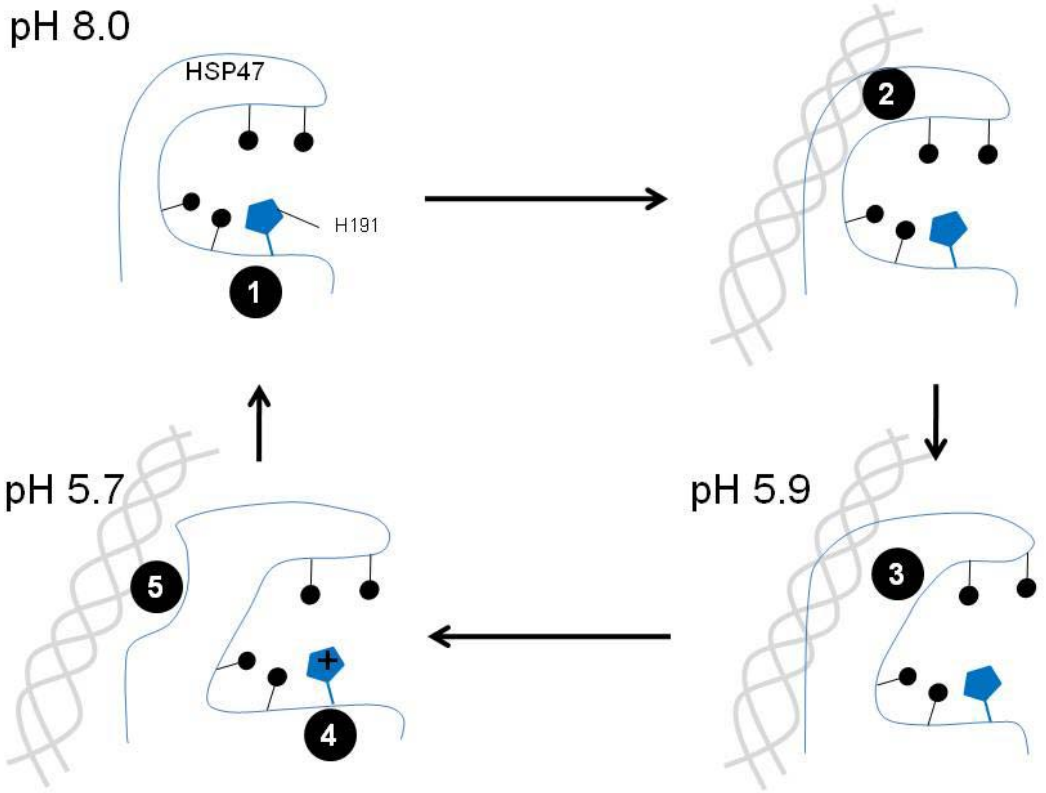


Figure 4.1: Schematic representation for the proposed mechanism of the concerted action of His residues in the breach cluster in controlling HSP47 pH-switch mechanism. The mechanism is based on the mutanegesis work performed on all surface His residues predicted from homology model. pH values might differ in different experimental conditions though the magnitude is unlikely to. **1**. At alkaline pH, all the 'breach' residues are non-ionised and allow the cluster to stay intact. Hydrophobic interactions with neighbouring side chains are also possible (black). **2**. Binding to collagen triple helix at alkaline condition. **3**. Upon decrease in pH to 5.9, the side chains of other residues start to ionise and perturb the region around the cluster slightly but insufficient to allow dissociation. **4**. As pH reaches 5.7, H191 subsequently ionises and allows full rearrangement to non-binding state. **5**. HSP47 then dissociates from collagen, before adopting binding state again at suitable pH condition.

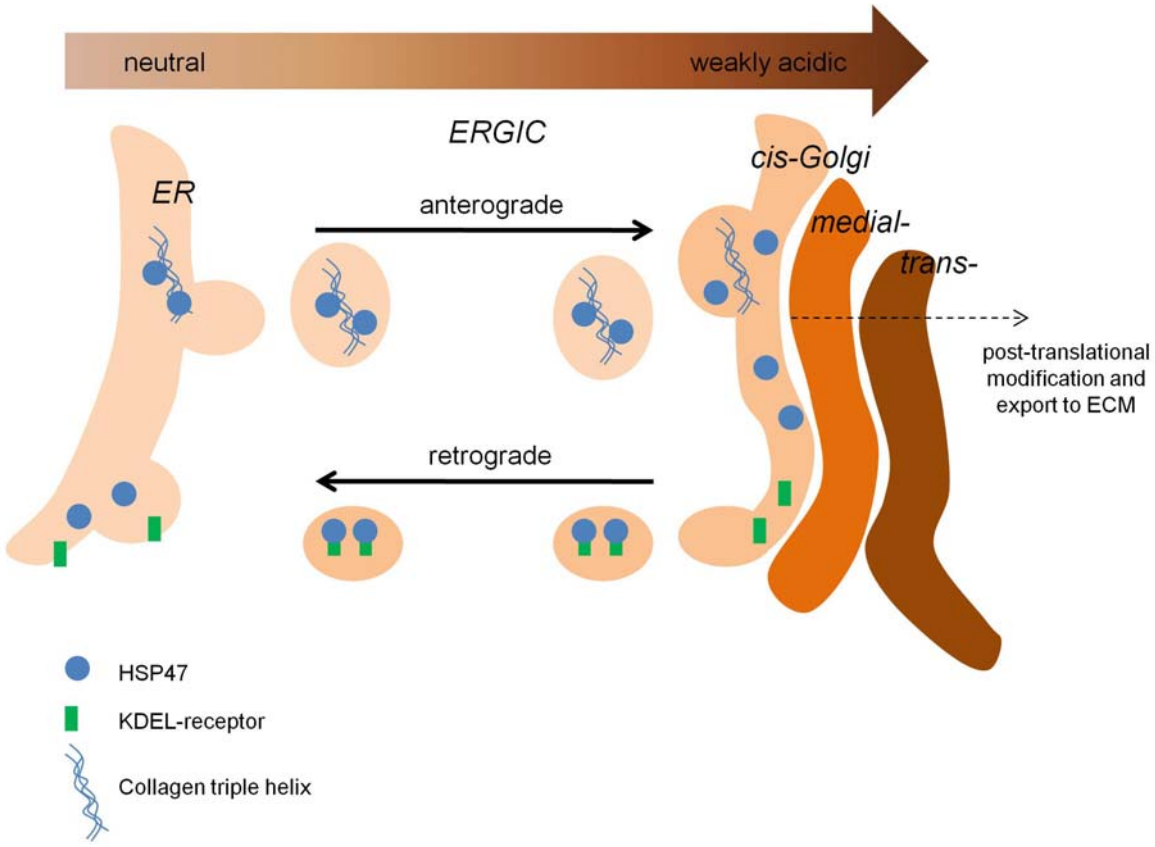


Figure 4.2: Anterograde transport of collagen from ER to Golgi apparatus and retrograde transport mechanism to recycle collagen-bound HSP47 back to ER. Adapted from (Paroutis et al., 2004; Vitale and Denecke, 1999).

4.4 Possible Effects of 'Breach' His Residue Mutations to Cellular Processes: Case Study on Other Proteins

Mutations of the His triad in the 'breach' cluster were found to perturb HSP47 mechanism significantly. H191 is one of the important residues crucial for pH release as its absence increases HSP47 sensitivity to pH and triggers release from collagen at a higher pH values. As accurate and concerted release of HSP47 from collagen triple helices is crucial during biosynthesis, by extrapolation, mutation of H191 can be detrimental to the process. Although cellular work is out of the scope of this thesis, it would be very comprehensible to summarise the findings up-to-date by discussing some similar studies conducted elsewhere to suggest effects of mutations on HSP47-expressing cells.

It is known that HSP47 binds collagen chain in the ER and assists in the stabilisation of the nascent triple helix. Then the HSP47-collagen complex is transported to the Golgi apparatus, before HSP47 dissociates from the properly folded collagen. Although the exact dissociation location is not yet determined, it has been speculated that it occurs either in the cisternal space of *cis*-Golgi or in the ER-Golgi intermediate compartment (ERGIC) (Nagata, 2003). HSP47 is then recycled back to the ER *via* retrograde transport pathway by receptors that recognise the *C*-terminal RDEL ER-retention signal sequence (Figure 4.3). One of the receptors that have been identified is erd2, a 26 kDa Golgi membrane protein that was initially discovered as the *ERD2* gene product in yeast *Saccharomyces cerevisiae.* The name is an acronym of ER retention-defective complementary group 2 (Lewis et al., 1990). erd2 has been demonstrated to be localised mainly in the *cis*-Golgi compartment (Elmendorf and Haldar, 1993).

The pH of the Golgi apparatus is variably reported to be in the slightly acidic to near neutral region (pH 5.7 to 6.7) (Kim et al., 1996; Paroutis et al., 2004; Wu et al., 2001). The accurate method for the measurement of the actual pH value in the *cis*-Golgi, and the secretory pathway generally, is still being actively investigated. One report suggests that pH of the ER is around 7.1 demonstrated in resting HeLa cells while that in the Golgi complex is reported to be between pH 6.3-6.5 (Kim et al., 1996; Kim et al., 1998). The affinity of erd2, as an important component of the secretory pathway, is known to be also affected by pH. Its binding is most optimum at pH between 5.0 and 5.5 (Wilson et al., 1993). This explains its

function in retrieving ER resident proteins in Golgi. Upon retrograde transport back to ER, the proteins are liberated from the receptor due to the neutral pH condition of ER compartment (Figure 4.3). This drop in pH along the secretory pathway is said to be due to the increase in active proton pump density (Wu et al., 2001). Also as the mechanism of binding-and-release action of erd2 to its partners is mediated by pH difference, it is interesting to note that this behaviour is completely opposite to that of HSP47 in terms of pH condition (Wilson et al., 1993). In addition, this complex is influenced by other factors such as aggregation state of the receptor in the ER and Golgi, or the regulation by ionic constituents in Golgi. The latter is taken into consideration as Asp residue in the hydrophilic KDEL-binding pocket of erd receptor has been shown to form direct pairing with positively-charged residue at position -4 from the *C*-terminus of the ligand (position **K** in **K**DEL sequence). This position is interchangeably occupied by Arg, Lys or His in different ligands (Scheel and Pelham, 1998).

The multiple function of His residue in pH-control, substrate binding and others has been observed before, even in non-ER/Golgi related proteins. For example in the human $P2X_4$ purinoreceptor, where mutation of one His to Ala not only abolishes the pH sensitivity of the receptor, but also has some effect on its ligand binding. P2X channels mediate fast excitatory actions of ATP when it is co-released with other transmitters in the central and peripheral nervous systems. There are many different isoforms of the P2X receptor that have been found, which involve in the modulation of ATP-gated channels and therefore regulate synaptic transmission. Mutant $P2X_4$ carrying H286A mutation, located on the extracellular side of the channel, abolishes its sensitivity of the ATP-activated current upon decrease in extracellular pH (Clarke et al., 2000).

A more significant case on ER resident proteins has been reported concerning the receptor associated protein (RAP), which is a chaperone of the low density lipoprotein receptor (LDLR) family. It is a three-domain, 39-kDa chaperone and each domain carries a different function. Generally, it inhibits undesirable premature interaction of the LDL receptor family and their ligands in the ER and also facilitates the delivery of newly-synthesised LDLRs to the cell surface (Willnow et al., 1995; Willnow et al., 1996). The latter function is reported to involve a single domain of RAP, namely domain 3 (D3). Very much like HSP47, RAP dissociates from LDLR in the Golgi apparatus prompting the suggestion of His-controlled mechanism as its possible mechanism of release. The structure of D3 has been

solved using NMR, and it revealed the highly-helical content of the chaperone. Its surface charge distribution is found to increase significantly at pH lower than 6.5 further supporting the His switch hypothesis (Lee et al., 2006). In addition, mutation of His residues in D3 impairs the dissociation from its client at low pH. This then affects the secretion of LDLR outside of the cell. The reason of this probably involves cellular degradation process of the RAP-LDLR complex (Estrada et al., 2008). It was also found that in cells transfected with RAP carrying D3 mutant His, RAP and LDLR co-localise in the ER. There is a suggestion that mutant RAP-LDLR complex is recycled back to the ER *via* anterograde transport action of KDEL-receptors, as its export to the Golgi is shown to be unaffected (Lee et al., 2006). It is still not known whether the degradation of the aberrant complex involves the ER-associated degradation (ERAD) pathway or another degradation mechanism in the cytoplasm.

From these studies, there could be an adverse effect on collagen biosynthetic pathway of HSP47 upon mutating its critical His residues. Dissociation of HSP47 from correctly folded collagen triple helix usually occurs in the Golgi complex or in the ER-Golgi intermediate compartment. These compartments definitely have specific pH values, just like many other cellular compartments, in order to regulate cellular functions. For that reason, premature release of mutant HSP47 from collagen will probably expose the folded collagen molecules to degradation due to lack of chaperone to help in stabilisation. This could in turn affect the transport of collagen *in vivo* leading to disrupted formation of extracellular matrices, which is potentially lethal.

4.5 Conclusion

This chapter compiled the data obtained from the experimental work presented in the two previous chapters. After examining the significance of the 'breach' region to serpin structure and function, a working model on the mechanistic behaviour of HSP47 during pH-release from collagen was proposed. The model is built on the structural rearrangement experienced by the region around His residues located in the 'breach' cluster, based on findings described from Chapters 2 and 3. This chapter also presents the proposed effect of the rearrangement, induced by pH change, to HSP47 binding and release behaviour. This model is still at its infancy and is constructed based on the knowledge gained from experiments using the techniques available in our laboratory. Although far from being a complete model, this is the first to date explaining HSP47 pH-switch mechanism, which is the only mechanism known to govern HSP47 release from collagen. Finally, a thorough literature review was conducted to outline the possible effects of His mutations on cellular behaviour in order to help guiding future experiments and investigations on cellular mechanism of HSP47 during collagen biosynthesis, until a new study comes to light.

Chapter 5

Future Directions

CHAPTER 5: FUTURE DIRECTIONS

5.1 Construction of More His Mutants to Validate pH-switch Model

The model presented in Chapter 4 is proposed based on the observations made on eight His-to-Ala mutants targeting eight residues (six single and two double mutants). These are the residues predicted to reside at solvent accessible locations of HSP47 molecule. In order to validate and confirm the model, a further six mutants could be constructed targeting the other residues (H90, 140, 220, 297, 335 and 368). Analyses on the mutants will provide more data for considerations towards the pH-switch model, and confirm their predicted locations at the non-solvent accessible locations of HSP47 homology model.

5.2 Determination of HSP47-collagen Binding Interface

Like many non-inhibitory serpins, the binding interface of HSP47 to its partner is still unknown. Earlier on, Chapter 2 presented two proposals made for HSP47/collagen binding site and interface (Dafforn et al., 2001; Koide et al., 2006a). These two models still have plenty of room for improvements. Hence it is of interest to investigate this further, using either mutagenesis technique; or co-crystallisation of HSP47 with short collagen mimetic peptides with desirable binding motif. In the first approach, all hydrophobic residues predicted to be located near the 'breach' region and the putative collagen binding groove of sB-hA-hG-hH could be mutated. This will probably involve 10-15 mutants.

In the second approach, the binding of HSP47 to CMPs is known to be quite weak (K_d in the μM to nM range). So in order to withstand crystallisation, a stronger HSP47 binding motif can be engineered into the CMPs. One residue, namely Arg, is now known to be the preferred residue for HSP47 recognition (Koide et al., 2002). The knowledge of collagen binding interface will be very useful to guide research on the development of small molecules to inhibit collagen expression by blocking the chaperoning action of HSP47. This is actually one of the strategies actively investigated in the treatment of fibrotic diseases, alongside HSP47 siRNA and oligonucleotides.

5.3 Crystallisation and Three-dimensional Structure Determination

Before co-crystallisation of HSP47 with collagen peptide can be done, the crystallisation of HSP47 alone should be attempted. HSP47 is known to be highly resistant to crystallisation, mainly due to aggregation problem at high concentration, from our experience. So in order to overcome this problem, several strategies can be attempted. Firstly, engineering of HSP47 to allow expression in high concentration but retains solubility. Several surface residues can be targeted, without affecting the important regions such as the 'breach' region known to be important for its function. Surface hydrophobic residues can be engineered to hydrophilic residues, for example. Secondly, by optimising HSP47 solubility using various protein tags commercially available, such as the NusA tag (Novagen). Various additives, such as amino acids, can also be attempted during cell lysis. Once high solubility is ensured, crystallisation can be attempted. NMR spectroscopy can also be tested, especially on the quality of signal when HSP47 sample is run. With the advent of state-of-the-art NMR techniques, more proteins can be investigated in this way.

5.3 Investigation on the Effect of Specific His-to-Ala Mutation of HSP47 Behaviour in the Cell and Animal Model

As discussed at the end of Chapter 4, it is interesting to know the effect of His mutation of HSP47 on normal collagen synthetic pathway. This is achievable using collagen expressing cells, for example the mouse embryonic fibroblast (MEF) is among the most widely-used cells for HSP47-related cell work. Then mutant plasmids can be engineered before transfection into the cells. The localisation of collagen and HSP47 molecules can be detected using cell stains labelled with collagen/HSP47 specific antibodies. Once the behaviour in the cell is established, the work can be further extended to include animal model. A good starting point for such work is the study reported by Nagai and co-workers (Nagai et al., 2000), which demonstrates the requirement of HSP47 for the normal development of mice, using gene targeting vector delivered into mouse strain C57BL/6J. Although the study uses gene *knock-out* technique targeting *hsp47* gene, site-specific mutations on mouse model has been documented on other diseases (Bedell et al., 1997; Simmons, 2008). If successful, the finding

will corroborate results obtained from *in vitro* experiments detailed in this thesis, with various potential contributions to the study of collagen-related diseases and abnormalities.

Chapter 6

Materials and Methods

CHAPTER 6: MATERIALS AND METHODS

6.1 Materials

Unless otherwise indicated, all commonly used chemicals and reagents were of analytical grade and obtained either from Sigma-Aldrich, VWR, Fluka or Merck.

6.2 Bacterial Strains

Three strains of *E. coli* were used in this study. Their genotypes are listed below, obtained from EcoliWiki database (http://ecoliwiki.net/colipedia/index.php/Welcome_to_EcoliWiki) and also from OpenWetWare (http://openwetware.org/wiki/E._coli_genotypes#BL21.28DE3.29):

DH5α: *huA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*. Used mainly in mutagenesis for plasmid propagation and mutant screening works, and also for long-term plasmid maintenance.

BL21(DE3)pLysS: *F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS(cm^R).* A high stringency expression host cell, contains the pLysS plasmid that encodes T7 phage lysozyme, which is an inhibitor for T7 polymerase which reduces and almost eliminates basal expression from transformed T7 promoter containing plasmids when not induced with IPTG.

BL21(DE3): *F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*. Non pLysS-containing host. Less stringent than BL21(DE3)pLysS, for higher expression level.

6.3 Site Directed Mutagenesis of Mouse HSP47 Gene

This study used the mouse HSP47 gene, the sequence of which has been deposited to GeneBank with an accession number X60676.1 (Hosokawa et al., 1993; Takechi et al., 1992). The mouse HSP47 cDNA was amplified by PCR using 5′ and 3′ primers with *Hin*dIII and *Nde*I restriction sites, respectively. Subsequently, the gene was cloned into the pET-24b(+) vector (Novagen) for HA mutants or pET-21a(+) (Novagen) for WF mutants, which contained the equivalent restriction sites. pET-21a vector was engineered to allow a co-expression of a six-histidine tag (His-tag) to the *N*-terminal of the recombinant protein while pET-24b was engineered for expression of a non His-tagged protein. The former carries ampicillin-resistance (amp^R) gene while the latter is kanamycin resistant (kan^R). Both are low copy number vectors to avoid too high expression which can result in HSP47 aggregation. The presence of T7*lac* promoter system allows the use of IPTG to induce protein expression. All the cloning works have been done previously and the plasmids were kindly supplied by Dr Takayuki Homma (pET-21a) and Dr Dee Olerenshaw (for pET-24b) (Imperial College Genetic Therapies Centre, UK). In pET-24b(+), the protein is expressed with an initial vector-derived Met before the start of the first amino acid of the mature HSP47 sequence Ala, with 400 amino acids altogether from the first Ala to the final Leu. This is similar in pET-21a(+) vector, with an additional six His tag at the *N*-terminus, and a linker region between the tag and the Met residue.

Mutations were introduced using the method developed by Sawano and Miyawaki (Sawano and Miyawaki, 2000) with DH5α as the *E. coli* host. Table 6.1 shows the primer sequences to construct the mutants. Primers were initially phosphorylated at their 5′ end using T4 polynucleotide kinase (New England Biolabs). PCR reactions were carried out in a 50 µL total reaction volume using 1 µL of 40-50 ng/µL template plasmid, 7 µL of 2 pmol/ µL of primer, 1 µL of 2.5 mM of dNTPs, 1 µL of 2.5 U/µL of *Pfu* DNA polymerase (Stratagene) in 0.5× *Pfu* polymerase reaction buffer, and 0.5 µL of 20 U/µL of *Taq* DNA ligase (New England Biolabs, MA, USA) in 0.5× *Taq* DNA ligase buffer. The thermal cycler (Mastercycler Gradient, Eppendorf) was programmed as follows: pre-incubation at 65°C for 5 min allowing the ligase to repair any nicks in the template; initial denaturation at 95°C for 2 min; then 23 amplification cycles (denaturation at 95°C for 10 sec, annealing at 55°C for 30 sec and elongation at 65°C for 13.5 min); and finally post-incubation at 75°C for 7 min. Then

1 μL (10 U/μL) of *Dpn*I (Stratagene) was added to the 50 μL reaction mixture, and incubated at 37°C for 1 h. For non-mutant strand synthesis, the reaction mixture was then subjected to denaturation at 95°C for 30 sec, followed by 2 cycles of (95°C for 30 sec, 55°C for 1 min and 70°C for 13.5 min), then finally 75°C for 7 min for final extension (Sawano and Miyawaki, 2000). 10 μL of the final sample was used for transformation into chemically-competent *E.coli* DH5α cells.

Single *E. coli* colonies were picked from LB agar plate with appropriate antibiotics and 0.4%(w/v) of glucose, and inoculated into sterile 100 mL LB broth containing the same antibiotics. Plasmid isolation and purification were carried out using the EZNA Plasmid Isolation Kit Midi (Omega Biotek) according to manufacturer's instruction. DNA concentration was measured using the NanoDrop ND-1000 (NanoDrop). Typical DNA yield is usually around 100-200 ng/μL with good purity ($A_{260}/A_{280} > 1.8$). The plasmid samples were sequenced to screen for positive mutants. Each positive mutant construct was then sequenced again to confirm the presence of a unique mutation and no other mutations between the promoter and terminator regions. Sequencing service was provided by Beckman Coulter Genomics (UK). Mutagenesis efficiency varies, but normally between 10-50%.

HA mutants were expressed in *E. coli* BL21(DE3)pLysS (Promega) and WF mutants were expressed in BL21(DE3) (Invitrogen) hosts .

Table 6.1: Primers used for mutagenesis reactions.
Target amino acids mutated are underlined. Primers were custom synthesised by MWG Biotech (Germany) or Invitrogen (UK).

Target mutation	Primer sequence
H135A	5´- CAGCAGCAAGCAA**GCA**TACAACTGCGAAC
H191A	5´- ATGTTCTTTAAGCCA**GCA**TGGGATGAGAAGTTT
H197A	5´- CACACTGGGATGAGAAGTTTCAC**GCA**AAGATGGTGG
H198A	5´- CACACTGGGATGAGAAGTTT**GCA**CACAAGATGGTGG
H197A:H198A	5´- CACACTGGGATGAGAAGTTT**GCAGCA**AAGATGGTGG
H244A	5´- GATGCCCCTGGCT**GCA**AAGCTCTCCAGC
H255A:H256A	5´- GCCTCATCATCCTCATGCCC**GCAGCA**GTGGAGCCGC
H302A	5´- CATGACCTGCAGAAA**GCC**CTGGCAGGA
W110F	5´- ACTGCGCGCAACGTGACC**TTT**AAACTGGGCAGCCGC
W158F	5´- TCCATCAACGAG**TTT**GCCTCGCAGACCACG
W192F	5´- TTCTTTAAGCCACAC**TTT**GATGAGAAGTTTCAG
W275F	5´- GAGCAGCTGAAGGCC**TTT**ATGGGAAAGATGCAGAAG
W341F	5´- ACTGCCTTCGAG**TTC**GACACCGAGGGCAAC

6.4 Preparation of Chemically-competent *E. coli* Cells and Transformation

E. coli cell stock was first cultured in 2 mL of 2×YT medium (tryptone 16 g, yeast extract 10 g, NaCl 5 g, deionised water 1 L) or LB medium (tryptone 10 g, yeast extract 5 g, NaCl 10 g, deionised water 1 L) for DH5α, while supplemented with 0.2%(w/v) glucose (for BL21(DE3)) and 20 μg/mL chloramphenicol (for BL21(DE3)pLysS). The culture was grown at 37°C with shaking overnight before 1 mL of the culture being transferred into 100 mL of fresh medium with glucose and/or antibiotics. Cultivation was continued at 37°C until optical density at 600 nm (OD_{600}) reached 0.3. The culture was then aliquoted out into 50 mL Falcon tubes and placed on ice for 5 min. *E. coli* cells were collected by centrifugation at 4000 rpm with ALC T534 rotor (ALC) for 10 min. The resulting pellet was resuspended in 20 mL of sterile TfbI solution (30 mM potassium acetate, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15%(v/v) glycerol, pH 5.8) by vortexing. The suspended cell solution was placed on ice for 5 min, before being centrifuged as above. The pellet was again resuspended in 2 mL of sterile TfbII solution (10 mM PIPES, 75 mM $CaCl_2$, 10 mM RbCl, 15%(v/v) glycerol, pH 6.5) before being placed on ice for 15 min. The cell suspension was finally dispensed into sterile, ice-cold 1.5 mL Eppendorf tubes in 100 μL aliquots and quick-frozen in liquid nitrogen before storage at -80°C.

Plasmid was transformed into the competent *E. coli* by adding 0.5-1 μL (or 10 μL from mutagenesis reaction) into the frozen cells and placed on ice for 30 min. The tube was heat shocked at 42°C for 1 min and returned to ice for another 2 min. Then 1 mL of LB broth was added, supplemented with 0.2%(w/v) glucose and 20 mM $MgSO_4$ and grown at 37°C for 1 h with shaking. The cells were finally plated out on LB agar with glucose and appropriate antibiotics and incubated at 37°C overnight.

6.5 Bacterial Cultivation and Preparation of HSP47 Crude Sample

Recombinant *E. coli* containing WT and mutagenic HSP47 gene was grown on media with 25 μg/mL kanamycin for BL21(DE3) and both kanamycin and 20 μg/mL chloramphenicol for BL21(DE3)pLysS. A loopful of glycerol stock, originally stored at -80°C was scraped off and inoculated into a pre-culture of 2×YT medium supplied with appropriate antibiotics and 0.2%(w/v) of glucose. The culture was incubated at 37°C with shaking at 180 rpm overnight. The pre-culture was then inoculated into a larger culture in a volume ratio of 1:20 of a 1 L of the same medium with the same antibiotics but without glucose. The main culture was incubated at 20°C, 180 rpm until the OD_{600} reached 0.3 before IPTG was added (0.1 mM of final concentration) and continued for further 12-18 h. The cells were then harvested by centrifugation at 4000 rpm with GS-3 rotor (Sorvall) for 10 min. When necessary, the pellet was resuspended in saline solution (8.4 g/L) and centrifuged at 4,000 rpm with ALC T534 rotor (ALC) for 15 min before storing at -80°C. For lysis, cell pellet was resuspended in lysis buffer (50 mM sodium phosphate pH 8.0, 500 mM NaCl, 0.02%(v/v) Triton X-100, 3 mM β-mercaptoethanol), supplemented with Complete Mini EDTA-free protease inhibitor tablet (Roche, Germany). The inhibitor cocktail contains broad spectrum inhibitors of serine and cysteine proteases. Bacterial lysis was carried out using sonication on ice (total sonication time of 1.5 min per cycle and the cycle repeated twice, with 5 min of resting period between cycles, at 20% amplitude) with a VCX400 sonicator (Sonics and Materials, Inc). When needed, DNaseI (Roche) was added to 0.1 mg/mL final concentration with 2 mM $MgCl_2$ followed by 15 min incubation at on ice. The resulting lysate was then centrifuged at 16,000 rpm with SS-34 rotor attached to RC-5B Plus centrifuge (both are from Sorvall) for 40 min at 4°C. The supernatant was filtered through a 0.2 μm syringe filter and purified using the gelatin-agarose column, or the HisTrap HP immobilised nickel affinity column (GE Healthcare).

6.6 Gelatin-agarose Purification of HSP47

HA mutants were purified using the gelatin-agarose affinity chromatography. Gelatin-agarose (Sigma-Aldrich) was packed into a 5 mL HR16/5 column (GE Healthcare) and packed columns were stored at 4°C until further use. Before purification, the columns were equilibrated with cold buffer A (50 mM sodium phosphate, 100 mM NaCl, pH 8.0) for about 5 column volumes (CV). All buffers were degassed. The flow rate was fixed at 1 mL/min for this step and all the subsequent steps, using the Fast Protein Liquid Chromatography (FPLC) system (Pharmacia Biotech). Filtered supernatant was loaded into the column (kept on ice) and washed using buffer A before being eluted with a 14 CV gradient of buffer B (50 mM sodium phosphate, 100 mM NaCl, pH 4.1). Fractions were collected at 2 mL per fraction and the high pH was restored with 200 μL of 1 M sodium phosphate pH 8 before being stored at 4°C. The columns were then washed and re-equilibrated with buffer A for about 5 CV before being stored at 4°C in 20%(v/v) ethanol.

6.7 Immobilised Nickel Affinity Chromatography Purification of HSP47

WF mutants were purified using the HisTrap HP nickel sepharose column (1 mL) (GE Healthcare) according to the manufacturer's protocol. Before use, the column was first washed with 3-5 column volumes (CV) of distilled water. Then it was equilibrated with at least 5 CV of binding buffer 50 mM sodium phosphate, 150 mM NaCl, 20 mM imidazole, pH 8.0) at a flow rate of 1.5 mL/min. Then the filtered supernatant was applied before being washed with binding buffer for at least 10-15 CV. Finally, the bound protein was eluted with elution buffer (50 mM sodium phosphate, 150 mM NaCl, 250 mM imidazole, pH 8.0) using a linear gradient of 30 CV. After 3-4 purifications, the column was stripped with stripping buffer (20 mM sodium phosphate, 500 mM NaCl, 50 mM EDTA, pH 7.4) and recharged with 100 mM $NiSO_4$.

Imidazole was removed using HiPrep 26/10 desalting column (GE Healthcare) equilibrated and eluted with same buffer as the gelatin-agarose column wash buffer.

6.8 Determination of Protein Concentration

Protein concentration was determined using either Bichinchoninic Acid (BCA) Assay Kit (Sigma-Aldrich) or calculated from the absorbance at 280 nm using the Beer-Lambert law with the molar extinction coefficient of 42,400 $M^{-1}cm^{-1}$ calculated using ProtParam (Gasteiger et al., 2005).

6.9 Protein Visualisation using Denaturing SDS-PAGE

Samples for SDS-PAGE analysis were initially mixed with the 2× Novex Tris-glycine SDS-PAGE sample suffer (Invitrogen) and 5%(v/v) of β-mercaptoethanol (β-ME) followed by heating at 95°C for 5 min before loading into the wells of a pre-cast 12% SDS-PAGE gel and electrophoresed at 200 V for 70 min in 1× Tris-glycine SDS-PAGE tank buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) (Fermentas). The gel was then stained with Coomassie Brilliant Blue stain for 1 h and de-stained with MilliQ water overnight.

6.10 Immunoblotting Analysis

Protein samples electrophoresed on 12% SDS-PAGE gel was first transferred to ECL Hybond nitrocellulose membrane (GE Healthcare). A wet transfer system was used, at a constant voltage of 100 V for 1 h in 5 mM sodium tetraborate buffer kept on ice. The membrane was blocked with 5% skimmed milk in PBS(-) buffer (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) at 4°C overnight or at room temperature for 1 h. Then, after discarding the blocking solution, anti-HSP47 primary antibody (Stressgen) in skimmed milk (1:1000 dilution) was added and incubated on a rocking platform at room temperature for another 1 h. After washing twice with 0.1% Tween-20 in PBS(-) for 10 min, the membrane was probed with goat anti-mouse secondary antibody, alkaline phosphatase (AP) conjugate (Biosource) (1:2000 dilution) for an additional 1 h before a 5-min washing with Tween-20 in PBS(-) four times and once in MilliQ water for 3 min for the final wash. Finally, detection was performed using the 5-bromo-4-chloro-3-indolylphosphate/nitro blue

tetrazolium (BCIP/NBT) tablet (Sigma-Aldrich). The substrate solution when dissolved in 10 mL MilliQ water contains BCIP (0.15 mg/mL), NBT (0.30 mg/mL), Tris buffer (100 mM), and $MgCl_2$ (5 mM), pH 9.25-9.75. BCIP is hydrolysed by AP forming an intermediate that undergoes dimerisation to produce an indigo dye. Then NBT is reduced to the NBT-formazan by two reducing equivalents generated by the dimerisation. Together, two compounds generated an intense black-purple precipitate. Where necessary, band quantification was performed using NIH ImageJ software (Rasband, 1997-2009) with a built-in gel analysis program.

6.11 Gelatin-agarose Pull Down Assay

This assay was used to determine the binding ability of HSP47 to gelatin. 5.4 mL lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 15% glycerol, 0.2% Nonidet P-40, protease inhibitor tablet, pH 8.0) was added to a pelleted cell culture (from about 12 mL of culture) and re-suspended. The cell suspension was then sonicated on ice and 1500 µL of lysate was aliquoted out and centrifuged at 13,000 rpm with Biofuge Pico (Sorvall), for 20 min at 4°C. 400 µL of the supernatant was added to 20 µL (bed volume) of gelatin-agarose beads (Sigma-Aldrich). The mixture was then rotated for 1 h at 4°C before being centrifuged at 6,000 rpm (Biofuge Pico, Sorvall) for 3 min at 4°C. The pellet was washed with 200 µL of wash buffer, and centrifuged again. This was repeated twice before 100 µL of MilliQ water was added, with 100 µL of 2× sample buffer, 10 µL of β-ME, and heated at 95°C for 5 min. The sample was centrifuged before loading to SDS-PAGE, followed by immunoblotting analysis for visualisation and quantification.

6.12 Collagen Mimetic Peptide (CMP) Synthesis

The peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ [$(POG)_{10}$] and $(Pro\text{-}Hyp\text{-}Gly)_4$-Pro-Arg-Gly-$(Pro\text{-}Hyp\text{-}Gly)_5$ [$(POG)_4$-PRG-$(POG)_5$] were synthesized using a Symphony Quartet peptide synthesiser (Protein Technologies Inc.) using *N*-(9-fluorenyl)methoxycarbonyl (Fmoc)-protected $_L$-amino acids on Gly-wang resin support. The peptides were purified on a reverse-phase high pressure

liquid chromatography (HPLC) system with a C_{18} column. Purity of the peptides was determined by mass spectrometry using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) and electrospray ionisation (ESI). The service was provided by the Department of Chemistry Mass Spectrometry Service Unit, Imperial College London.

6.13 Coupling of CMP to Cyanogen Bromide (CNBr)-activated Sepharose Beads

The coupling of CMP to CNBr-activated sepharose was done according to manufacturer's instruction. First, the required amount of freeze dried CNBr-sepharose powder (GE Healthcare) was weighed and resuspended in 1 mM HCl. The medium swelled immediately and was washed for 15 min with 1 mM HCl on a sintered glass filter, added in several aliquots. Then the peptide to be coupled was dissolved in coupling buffer (0.1 M $NaHCO_3$ pH 8.3, 0.5 M NaCl). 5 mL of coupling solution was used per gram of freeze dried powder. The coupling solution containing the peptide was mixed with the medium in a stoppered vessel, rotated end-over-end for 1 h at room temperature or overnight at 4 °C. Excess ligand was washed away with at least 5 medium (gel) volumes of coupling buffer. Finally any remaining active groups were blocked with 1 M ethanolamine, pH 8.0. Coupling efficiency was monitored with HPLC.

Analytical HPLC was performed on a Hitachi-LaChrom L-7150 equipped with a Polymer Laboratories PL-ELS 1000 evaporative light-scattering detector. A Vydac C_{18} column was used (5 µM, 300 Å, 4.6 × 250 mm) (Grace) with a retention time of 60 min. Acetonitrile and water were used as solvents.

6.14 HSP47-Collagen Anti-fibrillisation Assay

The protocol used for collagen fibril formation was similar to that described previously (Thomson and Ananthanarayanan, 2000; Williams et al., 1978). 2 µM (final concentration) of pepsin-solubilised type I collagen from porcine skin (Nitta Gelatin) was brought to pH 7.3-7.4 by the addition of phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 8.0).

The formation of collagen fibrils was then monitored by measuring the turbidity of the sample at A_{313} using a Thermo Varioskan microplate reader (Thermo Fisher Scientific) in the presence of different amounts of HSP47 WT and mutants. The absorbance was background-corrected and buffer-subtracted. Each experiment was monitored over a period of 120 min at a constant temperature of 34°C. Measurements were made in triplicates.

6.15 Circular Dichroism Spectroscopy

Spectrum and Temperature Titration

Circular dichroism (CD) spectra were recorded on a JASCO J715 spectropolarimeter using a thermostatically controlled 10 mm quartz cuvette. HSP47 was studied in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 8.0). Far UV CD spectra were recorded from 300-190 nm at 25°C. Each scan was set at 100 nm/min with a continuous scan mode and a standard sensitivity (100 mdeg). The response was set at 4 sec with 2 nm of band width. Each spectrum was an accumulation of 16 scans. For the temperature titration experiments, $\Delta\Delta A_{222}$ was monitored while the temperature was increased at a constant rate of 1°C per minutes from 25 to 90°C. Melting temperatures were estimated from the temperature titration profile with HSP47 samples at concentrations ranging from 2.8-3.0 μM. To calculate T_m, the fraction unfolded was obtained using the following equation, modified from (Persikov et al., 2004):

$$F(T) = \frac{\theta(T) - \theta_F(T)}{\theta_D(T) - \theta_F(T)} \quad \text{(Eq. 6.1)}$$

where θ is the observed ellipticity and θ_D and θ_F are the ellipticities for the denatured and folded. T_m was obtained at the midpoint of transition, where $F(T_m) = ½$.

pH Titration

CD pH titrations were carried out using an autotitrator connected to the spectropolarimeter. Sample was added to a stirred, thermostatically controlled 1 cm quartz cuvette. HSP47 WT and mutants were studied at a final protein concentration of 3 μM in phosphate buffer (50 mM sodium phosphate buffer, 100 mM NaCl, pH 8.0) at 25^{o}C. The pH was then progressively decreased by the addition of 2.5 μL aliquots of phosphoric acid using the autotitrator. The change in ΔA_{222} was then measured. The actual pH in the cuvette was determined using an ultraslim pH electrode (Hamilton). Equation 6.1 was used to estimate the inflection point of transition by replacing T, θ_D and θ_F with pH, θ_{Ac} and θ_{Al} which are the ellipticities at the pHs of full alkaline and acid states. CD spectra for each variant were also measured before and after titration to identify any structural variation following the change in pH.

HSP47-Peptide Binding Assay

The binding of HSP47 to PRG peptide was observed by monitoring the changes in CD signal as a function of peptide concentration. CD spectra were recorded in the presence of different peptide concentration at different pH values and then subtraction was performed with concentration matched-peptide spectra from blank titration. Ellipticity at 216 nm was taken instead of at 225 nm as the change in CD signal at this wavelength is significantly larger and better titration data can be obtained. The ellipticity values were then plotted as a function of peptide concentration and the binding isotherms were fitted to a single site binding model using OriginPro v8.5 (OriginLab).

Unit Conversion

Conversion between CD units was performed based on the guidance notes of the DichroWeb spectra deconvolution server (http://dichroweb.cryst.bbk.ac.uk). Throughout this thesis, two CD units are used, namely ΔA and mean residue ellipticity (MRE). The relationships between ΔA, MRE and the machine unit, theta (θ) is explained below.

ΔA is directly related to θ and converted using the equation, with θ in milidegrees:

$$\Delta A = \frac{\theta}{32.98} \quad \text{(Eq. 6.2)}$$

while $\Delta\varepsilon$, also known as molar circular dichroism (mdeg M^{-1} cm^{-1}) is expressed by this equation:

$$\Delta\varepsilon = \frac{[\theta]}{3298} \quad \text{(Eq. 6.3)}$$

where $[\theta]$ is the MRE (deg cm^2 $dmol^{-1}$ $residue^{-1}$). In most cases the unit is only displayed as (deg cm^2 $dmol^{-1}$). $\Delta\varepsilon$ also related to θ based on the following equation:

$$\Delta\varepsilon = \theta \times \frac{(0.1 * MRW)}{(P * CONC) * 3298} \quad \text{(Eq. 6.4)}$$

where *MRW* is the mean residue weight (protein mean weight [Dalton]/number of residues), *P* is the pathlength (cm) and *CONC* is protein concentration (mg/mL).

6.16 Fluorescence Anisotropy

Conjugation of BODIPY to PRG Peptide

Similar to the binding experiment with CD, PRG peptide was also used for this experiment. Peptide was synthesised using the standard Fmoc method on automated peptide synthesiser and analysed using MALDI-MS and reversed-phase HPLC. A modified, two-steps EDC/NHS coupling was used to avoid head-to-tail coupling of peptides, adapted from (Hermanson, 1996). 0.5 mg of BODIPY dye (Invitrogen) was initially dissolved in 0.05 M 2-(*N*-morpholino)ethane sulfonic acid (MES) buffer pH 6. Then EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide HCl) (Sigma-Aldrich, MO, USA) and sulfo-NHS (*N*-hydroxysulfosuccinimide) (Sigma-Aldrich) were added to the peptide solution to obtain a concentration of 2 mM for EDC and 5 mM for sulfo-NHS. The reaction was mixed for 15 min at room temperature to activate the dye molecules. 2-mercaptoethanol was then added to the reaction solution at a final concentration of 20 mM, before being mixed for a further 10 min at room temperature. This quenches the unreacted EDC and sulfo-NHS in the dye solution. Then, 0.46 mg of peptide (10-fold less molar amount than BODIPY) was added to the solution, and reaction was allowed to proceed to at least 2 hr at room temperature with continuous mixing. Conjugate was purified using size-exclusion chromatography with PD-10 column (GE Healthcare). PRG-BODIPY conjugate was analysed with reversed-phase HPLC system (Gilson) equipped with a UV-vis detector, using a Chromolith RP-18 C_{18} column (2 μM, 130 Å, 4.6 × 100 mm) (Merck). Acetonitrile and water were used as solvents. (Beauchamp et al., 1998)

Anisotropy Measurement

Anisotropy experiment was performed on a Perkin-Elmer LS-50B fluorimeter equipped with a thermostat. Temperature was maintained at 25°C throughout the measurements. Samples were excited at 430 nm and monitored at 510 nm. The excitation slit width was set to 2.5 nm and the emission slit was 10 nm, with 2 s integration time. Measurements were performed in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 8.0). HSP47 concentration was determined by the absorbance at 280 nm with an extinction coefficient of 42,400 $M^{-1}cm^{-}$

[1], and conjugate concentration was estimated from initial coupling reaction, taking into account all dilution steps assuming more than 95% conjugation efficiency. This is in good agreement with the value determined using the extinction coefficient of BODIPY, which is 68,000 $M^{-1}cm^{-1}$ at 504 nm (Invitrogen). For anisotropy measurement of conjugate in the presence of different HSP47 concentrations, 0.5 μM was incubated with HSP47 in same phosphate buffer for up to 5 mins at 25°C. Measurement was performed according to the instrumental parameters stated above.

Anisotropy measurement in the presence of different concentrations of glycerol were performed by addition of 90% (v/v) glycerol to 0.5 μM solution of conjugate to obtain final glycerol concentrations ranging from 0-36% (v/v), in 50 mM sodium phosphate buffer pH 8.0. Solution viscosities were calculated from the glycerol concentration, as the concentration of conjugate and buffer are very low, using the data on glycerol-water viscosity obtained from the Handbook of Chemistry and Physics (www.hbcpnetbase.com). This database supplies absolute viscosity values (η) in mPa s which is equivalent to centipoise (cP) against mass %. The viscosity of pure water at 20°C is given as 1.002 mPa s. Samples were equilibrated at 20°C for 20 min before measurement. Curve fitting was then performed using OriginPro v8.5 (OriginLab).

6.17 Steady-state Intrinsic Tryptophan Fluorescence Spectroscopy

Intrinsic tryptophan fluorescence emission spectra were recorded on a Shimadzu RF-5301PC spectrofluorophotometer (Shimadzu) using a thermostatically controlled 10 mm quartz cuvette. HSP47 was studied in phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 8.0) at 25^{o}C. Samples were excited at 295 nm and emission detected from 310-410 nm. The excitation slit width was set at 3 nm, and the emission slit widths were varied between 10-15 nm. The pH-titration experiments were performed by the addition of small aliquots of phosphoric acid (1-2 μL each) with constant stirring. Protein concentration was fixed to 1 μM for all analysis. Each spectrum was the average of three repetitions. All experiments were background corrected and performed within 12 hr after purification.

Averaged individual spectra recorded at various pH values were then fitted to polynomial model second order for the purpose of data extraction using OriginPro v8.5 software (OriginLab). Fitting ensures more accurate and smoother data for maximum fluorescence intensity (I_f) and wavelength at maximum intensity (λ_{em}). λ_{em} values against pH generated from the extraction were then fitted to single pH ionisation model using GraFit Pro v7.0.0 (Erithacus Software Ltd.) based on Equation 6.5. The y values vary from one value at low pH to another at high pH depending upon the ionization of a group with a certain pK_a.

$$y = L_1 + \frac{(L_2 - L_1).10^{(pH-pK_a)}}{10^{(pH-pK_a)} + 1} \quad \text{(Eq. 6.5)}$$

L_1 represents Limit 1 or the upper limit of the y values while L_2 is the lower limit of y values. In our case the λ_{em} values were taken as the limit values (Leatherbarrow, 2009).

References

Ahmad, S., M. Gromiha, H. Fawareh, and A. Sarai. 2004. ASAView: Database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics*. 5:51.

Anderson, K.S., S. Sibani, G. Wallstrom, J. Qiu, E.A. Mendoza, J. Raphael, E. Hainsworth, W.R. Montor, J. Wong, J.G. Park, N. Lokko, T. Logvinenko, N. Ramachandran, A.K. Godwin, J. Marks, P. Engstrom, and J. LaBaer. 2011. Protein Microarray Signature of Autoantibody Biomarkers for the Early Detection of Breast Cancer. *J. Proteome Res.* 10:85-96.

Anfinsen, C.B. 1973. Principles that Govern the Folding of Protein Chains. *Science*. 181:223.

Anfinsen, C.B., E. Haber, M. Sela, and J. White, F. H. 1961. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A*. 47:1309-1314.

Araki, K., T. Mikami, T. Yoshida, M. Kikuchi, Y. Sato, M. Oh-ishi, Y. Kodera, T. Maeda, and I. Okayasu. 2009. High expression of HSP47 in ulcerative colitis-associated carcinomas: proteomic approach. *Br J Cancer*. 101:492-497.

Asada, S., T. Koide, H. Yasui, and K. Nagata. 1999. Effect of HSP47 on prolyl 4-hydroxylation of collagen model peptides. *Cell Struct. Funct.* 24:187-96.

Bathurst, I.C., J. Travis, P.M. George, and R.W. Carrell. 1984. Structural and functional characterization of the abnormal Z α_1-antitrypsin isolated from human liver. *FEBS Lett.* 177:179-183.

Beauchamp, N.J., R.N. Pike, M. Daly, L. Butler, M. Makris, T.R. Dafforn, A. Zhou, H.L. Fitton, F.E. Preston, I.R. Peake, and R.W. Carrell. 1998. Antithrombins Wibble and Wobble (T85M/K): archetypal conformational diseases with in vivo latent-transition, thrombosis, and heparin activation. *Blood*. 92:2696-706.

Beck, K., and B. Brodsky. 1998. Supercoiled Protein Motifs: The Collagen Triple-Helix and the α-Helical Coiled Coil. *J. Struct. Biol.* 122:17.

Bedell, M.A., N.A. Jenkins, and N.G. Copeland. 1997. Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. *Genes Dev.* 11:1-10.

Bella, J., B. Brodsky, and H.M. Berman. 1995. Hydration structure of a collagen peptide. *Structure*. 3:893-906.

Bella, J., M. Eaton, B. Brodsky, and H.M. Berman. 1994. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science*. 266:75-81.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235-242.

Blond-Elguindi, S., S.E. Cwirla, W.J. Dower, R.J. Lipshutz, S.R. Sprang, J.F. Sambrook, and M.-J.H. Gething. 1993. Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell*. 75:717-728.

Blouse, G.E., M.J. Perron, J.-O. Kvassman, S. Yunus, J.H. Thompson, R.L. Betts, L.C. Lutter, and J.D. Shore. 2003. Mutation of the Highly Conserved Tryptophan in the Serpin Breach Region Alters the Inhibitory Mechanism of Plasminogen Activator Inhibitor-1 *Biochemistry*. 42:12260-12272.

Bordoli, L., F. Kiefer, K. Arnold, P. Benkert, J. Battey, and T. Schwede. 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4:1-13.

Bornstein, P. 1974. The Biosynthesis of Collagen. *Annu. Rev. Biochem.* 43:567-603.

Brodsky, B., and A.V. Persikov. 2005. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* 70:301-39.
Brown, K.E., K.A. Broadhurst, M.M. Mathahs, E.M. Brunt, and W.N. Schmidt. 2005. Expression of HSP47, a collagen-specific chaperone, in normal and diseased human liver. *Lab. Invest.* 85:789-97.
Canty, E.G., and K.E. Kadler. 2005. Procollagen trafficking, processing and fibrillogenesis. *J. Cell Sci.* 118:1341.
Carrell, R.W., D.L. Evans, and P.E. Stein. 1991. Mobile reactive centre of serpins and the control of thrombosis. *Nature*. 353:576-578.
Cejas, M.A., W.A. Kinney, C. Chen, G.C. Leo, B.A. Tounge, J.G. Vinter, P.P. Joshi, and B.E. Maryanoff. 2007. Collagen-Related Peptides: Self-Assembly of Short, Single Strands into a Functional Biomaterial of Micrometer Scale. *J. Am. Chem. Soc.* 129:2202-2203.
Chan, P., and J. Warwicker. 2009. Evidence for the adaptation of protein pH-dependence to subcellular pH. *BMC Biology*. 7:69.
Chen, X., D.P. Knight, and F. Vollrath. 2002. Rheological characterization of Nephila spidroin solution. *Biomacromolecules*. 3:644.
Christiansen, H.E., U. Schwarze, S.M. Pyott, A. AlSwaid, M. Al Balwi, S. Alrasheed, M.G. Pepin, M.A. Weis, D.R. Eyre, and P.H. Byers. 2010. Homozygosity for a Missense Mutation in SERPINH1, which Encodes the Collagen Chaperone Protein HSP47, Results in Severe Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86:389-398.
Clarke, C.E., C.D. Benham, A. Bridges, A.R. George, and H.J. Meadows. 2000. Mutation of histidine 286 of the human P2X4 purinoceptor removes extracellular pH sensitivity. *J. Physiol.* 523:697-703.
Clarke, E.P., G.A. Cates, E.H. Ball, and B.D. Sanwal. 1991. A collagen-binding protein in the endoplasmic reticulum of myoblasts exhibits relationship with serine protease inhibitors. *J. Biol. Chem.* 266:17230-5.
Creighton, T.E., D.A. Hillson, and R.B. Freedman. 1980. Catalysis by protein-disulphide isomerase of the unfolding and refolding of proteins with disulphide bonds. *J. Mol. Biol.* 142:43-62.
Cubellis, M.V., F. Caillez, T.L. Blundell, and S.C. Lovell. 2005. Properties of polyproline II, a secondary structure element implicated in protein–protein interactions. *Proteins*. 58:880-892.
Czabotar, P.E., S.R. Martin, and A.J. Hay. 2004. Studies of structural changes in the M2 proton channel of influenza A virus by tryptophan fluorescence. *Virus Res.* 99:57-61.
Dafforn, T.R., M. Della, and A.D. Miller. 2001. The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J. Biol. Chem.* 276:49310-9.
Dafforn, T.R., R.N. Pike, and S.P. Bottomley. 2004. Physical characterization of serpin conformations. *Methods*. 32:150.
Darby, N., and T.E. Creighton. 1995. Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry*. 34:16770-16780.
Davids, J.W., S.H. El-Thaher, A. Nakai, K. Nagata, and A.D. Miller. 1995. Modeling the Three-Dimensional Structure of Serpin/Molecular Chaperone HSP47. *Bioorg. Chem.* 23:427.
De Lauder, W.B., and P. Wahl. 1970. pH dependence of the fluorescence decay of tryptophan. *Biochemistry*. 9:2750-2754.

Degen, E., and D.B. Williams. 1991. Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility molecules. *J. Cell Biol.* 112:1099.
DeLano, W.L. 2004. PyMol User's Guide, San Carlos, California.
Dias, N., and C.A. Stein. 2002. Antisense Oligonucleotides: Basic Concepts and Mechanisms. *Mol. Cancer Ther.* 1:347-355.
Drögemüller, C., D. Becker, A. Brunner, B. Haase, P. Kircher, F. Seeliger, M. Fehr, U. Baumann, K. Lindblad-Toh, and T. Leeb. 2009. A Missense Mutation in the *SERPINH1* Gene in Dachshunds with Osteogenesis Imperfecta. *PLoS Genet.* 5:e1000579.
Eberhardt, E.S., N. Panasik, and R.T. Raines. 1996. Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *J. Am. Chem. Soc.* 118:12261-12266.
Edward, J.T. 1970. Molecular volumes and the Stokes-Einstein equation. *J. Chem. Educ.* 47:261-270.
El-Thaher, S.H., A.F. Drake, S. Yokota, A. Nakai, K. Nagata, and A.D. Miller. 1996. The pH-dependent, ATP-independent interaction of collagen specific serpin/stress protein HSP47. *Protein Peptide Lett.* 3:1-8.
Elliott, P.R., J.-P. Abrahams, and D.A. Lomas. 1998. Wild-type α_1-antitrypsin is in the canonical inhibitory conformation. *J. Mol. Biol.* 275:419-425.
Elliott, P.R., D.A. Lomas, R.W. Carrell, and J.-P. Abrahams. 1996. Inhibitory conformation of the reactive loop of α_1-antitrypsin. *Nat. Struct. Biol.* 3:676 - 681.
Elliott, P.R., X.Y. Pei, T.R. Dafforn, and D.A. Lomas. 2000. Topography of a 2.0 Å structure of α1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci.* 9:1274-1281.
Ellis, J. 1987. Proteins as Molecular Chaperones. *Nature*. 328:378-379.
Ellis, R.J., and A.P. Minton. 2006. Protein aggregation in crowded environments. *Biol. Chem.* 387:485-497.
Ellis, R.J., and S.M. van der Vies. 1991. Molecular Chaperones. *Annu. Rev. Biochem.* 60:321-347.
Elmendorf, H.G., and K. Haldar. 1993. Identification and localization of ERD2 in the malaria parasite Plasmodium falciparum: separation from sites of sphingomyelin synthesis and implications for organization of the Golgi. *EMBO J.* 12:4763.
Engh, R., H. Löbermann, M. Schneider, G. Wiegand, R. Huber, and C.-B. Laurell. 1989. The S variant of human α_1-antitrypsin, structure and implications for function and metabolism. *Protein Eng.* 2:407-415.
Estrada, K., C. Fisher, and S.C. Blacklow. 2008. Unfolding of the RAP-D3 Helical Bundle Facilitates Dissociation of RAP-Receptor Complexes *Biochemistry*. 47:1532-1539.
Farnum, M., and C. Zukoski. 1999. Effect of Glycerol on the Interactions and Solubility of Bovine Pancreatic Trypsin Inhibitor. *Biophys. J.* 76:2716-2726.
Feng, Y., G. Melacini, J.P. Taulane, and M. Goodman. 1996. Acetyl-Terminated and Template-Assembled Collagen-Based Polypeptides Composed of Gly-Pro-Hyp Sequences. 2. Synthesis and Conformational Analysis by Circular Dichroism, Ultraviolet Absorbance, and Optical Rotation. *J. Am. Chem. Soc.* 118:10351-10358.
Fernández-Recio, J., A. Vázquez, C. Civera, P. Sevilla, and J. Sancho. 1997. The Tryptophan/Histidine interaction in [alpha]-helices. *J. Mol. Biol.* 267:184-197.
Ferreira, L.R., K. Norris, T. Smith, C. Hebert, and J.J. Sauk. 1994. Association of Hsp47, Grp78, and Grp94 with procollagen supports the successive or coupled action of molecular chaperones. *J. Cell. Biochem.* 56:518-26.

Ferreira, L.R., K. Norris, T. Smith, C. Hebert, and J.J. Sauk. 1996. Hsp47 and Other ER-Resident Molecular Chaperones Form Heterocomplexes with Each Other and with Collagen Type IV Chains. *Connect. Tissue Res.* 33:265-273.

Fields, C.G., B. Grab, J.L. Lauer, A.J. Miles, Y.-C. Yu, and G.B. Fields. 1996. Solid-phase synthesis of triple-helical collagen-model peptides. *Lett. Pept. Sci.* 3:3-16.

Fields, G.B. 2010. Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* 8:1237-1258.

Fields, G.B., and D.J. Prockop. 1996. Perspectives on the synthesis and application of triple-helical, collagen-model peptides. *Pept. Sci.* 40:345-357.

Fraser, R.D.B., T.P. MacRae, and E. Suzuki. 1979. Chain conformation in the collagen molecule. *J. Mol. Biol.* 129:463-481.

Garrett, D.S., Y.-J. Seok, A. Peterkofsky, G.M. Clore, and A.M. Gronenborn. 1998. Tautomeric state and pK_a of the phosphorylated active site histidine in the N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate: sugar phosphotransferase system. *Protein Sci.* 7:789.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In* The Proteomics Protocols Handbook. J.M. Walker, editor. Humana Press, New Jersey. 571-607

Gelman, R.A., D.C. Poppke, and K.A. Piez. 1979. Collagen fibril formation in vitro. The role of the nonhelical terminal regions. *J. Biol. Chem.* 254:11741-11745.

Gelse, K., E. Pöschl, and T. Aigner. 2003. Collagens—structure, function, and biosynthesis. *Adv. Drug Deliver. Rev.* 55:1531.

Georgopoulos, C., and W.J. Welch. 1993. Role of the Major Heat Shock Proteins as Molecular Chaperones. *Annu. Rev. Cell Biol.* 9:601-634.

Gerlsma, S.Y. 1968. Reversible Denaturation of Ribonuclease in Aqueous Solutions As Influenced by Polyhydric Alcohols and Some Other Additives. *J. Biol. Chem.* 243:957-961.

Gettins, P.G.P.G.W. 2002. Serpin structure, mechanism, and function. *Chem. Rev.* 102:4751-804.

González Flecha, F.L., and V. Levi. 2003. Determination of the molecular size of BSA by fluorescence anisotropy. *Biochem. Mol. Biol. Edu.* 31:319-322.

Goodman, M., and J. Kwak. 1999. Design, synthesis and conformations of novel triple helical collagen mimetic structures. *J. Chem. Sci.* 111:35-49.

Greenfield, N.J. 2007. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protocols*. 1:2876-2890.

Guex, N., and M.C. Peitsch. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis*. 18:2714-2723.

Guo, J.S., C.H. Cho, J.Y. Wang, and M.W.L. Koo. 2002. Expression and Immunolocalization of Heat Shock Proteins in the Healing of Gastric Ulcers in Rats. *Scand. J. Gastroenterol.* 37:17-22.

Guzman, N.A., P.N. Graves, and D.J. Prockop. 1978. Addition of mannose to both the amino- and carboxy-terminal propeptides of type II procollagen occurs without formation of a triple helix. *Biochem. Biophys. Res. Commun.* 84:691-698.

Hagiwara, S., H. Iwasaka, S. Matsumoto, and T. Noguchi. 2007a. Introduction of antisense oligonucleotides to heat shock protein 47 prevents pulmonary fibrosis in lipopolysaccharide-induced pneumopathy of the rat. *Eur. J. Pharmacol.* 564:174-180.

Hagiwara, S.S., H.H. Iwasaka, S.S. Matsumoto, and T.T. Noguchi. 2007b. Antisense oligonucleotide inhibition of heat shock protein (HSP) 47 improves bleomycin-induced pulmonary fibrosis in rats. *Respir. Res.* 8:37.

Hagiwara, S.S., H.H. Iwasaka, S.S. Matsumoto, and T.T. Noguchi. 2007c. An antisense oligonucleotide to HSP47 inhibits paraquat-induced pulmonary fibrosis in rats. *Toxicology*. 236:199-207.

Hartl, F.U., and M. Hayer-Hartl. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16:574-581.

Harwood, R., M.E. Grant, and D.S. Jackson. 1975. Studies on the glycosylation of hydroxylysine residues during collagen biosynthesis and the subcellular localization of collagen galactosyltransferase and collagen glucosyltransferase in tendon and cartilage cells. *Biochem. J.* 152:291.

Hasselbacher, C.A., E. Rusinova, E. Waxman, R. Rusinova, R.A. Kohanski, W. Lam, A. Guha, J. Du, T.C. Lin, and I. Polikarpov. 1995. Environments of the four tryptophans in the extracellular domain of human tissue factor: comparison of results from absorption and fluorescence difference spectra of tryptophan replacement mutants with the crystal structure of the wild-type protein. *Biophys. J.* 69:20-29.

Helenius, A., and M. Aebi. 2001. Intracellular Functions of N-Linked Glycans. *Science*. 291:2364-2369.

Henderson, N.C., and J.P. Iredale. 2007. Liver fibrosis: cellular mechanisms of progression and resolution. *Clin. Sci.* 112:265.

Hermanson, G.T. 1996. Bioconjugate Techniques. Academic Press, Inc, San Diego, California, USA.

Hirayoshi, K., H. Kudo, H. Takechi, A. Nakai, A. Iwamatsu, K.M. Yamada, and K. Nagata. 1991. HSP47: a tissue-specific, transformation-sensitive, collagen-binding heat shock protein of chicken embryo fibroblasts. *Mol. Cell. Biol.* 11:4036-44.

Hirst, J.D., and C.L. Brooks. 1994. Helicity, Circular Dichroism and Molecular Dynamics of Proteins. *J. Mol. Biol.* 243:173-178.

Hochstenbach, F., V. David, S. Watkins, and M.B. Brenner. 1992. Endoplasmic reticulum resident protein of 90 kilodaltons associates with the T-and B-cell antigen receptors and major histocompatibility complex antigens during their assembly. *Proc. Natl. Acad. Sci. U. S. A.* 89:4734.

Hopkins, P.C., R.W. Carrell, and S.R. Stone. 1993. Effects of mutations in the hinge region of serpins. *Biochemistry*. 32:7650-7.

Horwich, A.L., E.U. Weber-Ban, and D. Finley. 1999. Chaperone rings in protein folding and degradation. *Proc. Natl. Acad. Sci. U. S. A.* 96:11033-11040.

Hosokawa, N., H. Takechi, S. Yokota, K. Hirayoshi, and K. Nagata. 1993. Structure of the Gene Encoding the Mouse 47-kDa Heat Shock Protein 47 (HSP47). *Gene*. 126:187-193.

Huber, R., and R.W. Carrell. 1989. Implications of the three-dimensional structure of .alpha.1-antitrypsin for structure and function of serpins. *Biochemistry*. 28:8951-8966.

Im, H., and M.-H. Yu. 2000. Role of Lys335 in the Metastability and Function of Inhibitory Serpins. *Protein Sci.* 9:934-941.

Invitrogen. 2006. Fluorescence Polarization: Technical Resource Guide. Invitrogen Corporation, Madison, WI, USA.

Ishida, Y., H. Kubota, A. Yamamoto, A. Kitamura, H.P. Bächinger, and K. Nagata. 2006. Type I collagen in Hsp47-null cells is aggregated in endoplasmic reticulum and deficient in N-propeptide processing and fibrillogenesis. *Mol. Biol. Cell*. 17:2346-55.

Jain, N., A. Brickenden, E.H. Ball, and B.D. Sanwal. 1994a. Inhibition of procollagen I degradation by colligin: a collagen-binding serpin. *Arch. Biochem. Biophys.* 314:23-30.

Jain, N., A. Brickenden, I. Lorimer, E.H. Ball, and B.D. Sanwal. 1994b. Interaction of procollagen I and other collagens with colligin. *Biochem. J.* 304 (Pt 1):61-8.

James, E.L., and S.P. Bottomley. 1998. The Mechanism of α_1-Antitrypsin Polymerization Probed by Fluorescence Spectroscopy. *Arch. Biochem. Biophys.* 356:296-300.

Jas, G.S., and K. Kuczera. 2004. Equilibrium Structure and Folding of a Helix-Forming Peptide: Circular Dichroism Measurements and Replica-Exchange Molecular Dynamics Simulations. *Biophys. J.* 87:3786-3798.

Kadler, K.E., C. Baldock, J. Bella, and R.P. Boot-Handford. 2007. Collagens at a glance. *J. Cell Sci.* 120:1955-1958.

Kadler, K.E., Y. Hojima, and D.J. Prockop. 1988. Assembly of type I collagen fibrils de novo. Between 37 and 41 degrees C the process is limited by micro-unfolding of monomers. *J. Biol. Chem.* 263:10517-10523.

Kadler, K.E., D.F. Holmes, J.A. Trotter, and J.A. Chapman. 1996. Collagen fibril formation. *Biochem. J.* 316:1-11.

Kaiserman, D., J.C. Whisstock, and P.I. Bird. 2006. Mechanisms of serpin dysfunction in disease. *Expert Rev. Mol. Med.* 8:1.

Kar, K., P. Amin, M.A. Bryan, A.V. Persikov, A. Mohs, Y.-H. Wang, and B. Brodsky. 2006. Self-association of Collagen Triple Helix Peptides into Higher Order Structures. *J. Biol. Chem.* 281:33283-33290.

Karolin, J., L.B.A. Johansson, L. Strandberg, and T. Ny. 1994. Fluorescence and Absorption Spectroscopic Properties of Dipyrromethene boron Difluoride (BODIPY) Derivatives in Liquids, Lipid Membranes, and Proteins. *J. Am. Chem. Soc.* 116:7801-7806.

Kelly, S.M., T.J. Jess, and N.C. Price. 2005. How to study proteins by circular dichroism. *BBA-Proteins Proteom.* 1751:119-139.

Kim, J.H., A.L. Clifford, D.B. Williams, W. Furuya, M.F. Manolson, and S. Grinstein. 1996. Dynamic measurement of the pH of the Golgi complex in living cells using retrograde transport of the verotoxin receptor. *J. Cell Biol.* 134:1387.

Kim, J.H., L. Johannes, B. Goud, C. Antony, C.A. Lingwood, R. Daneman, and S. Grinstein. 1998. Noninvasive measurement of the pH of the endoplasmic reticulum at rest and during calcium release. *Proc. Natl. Acad. Sci. U. S. A.* 95:2997-3002.

King, S.M. 2000. The dynein microtubule motor. *BBA-Mol. Cel. Res.* 1496:60-75.

Kobayashi, T., and M. Uchiyama. 2010. Effect of HSP47 expression levels on heterotrimer formation among type IV collagen a3, a4 and a5 chains. *Biomed. Res.* 31:371-377.

Kohn, J., and M. Wilchek. 1982. Mechanism of activation of Sepharose and Sephadex by cyanogen bromide. *Enzyme Microb. Technol.* 4:161-163.

Koide, T., S. Asada, and K. Nagata. 1999. Substrate recognition of collagen-specific molecular chaperone HSP47. Structural requirements and binding regulation. *J. Biol. Chem.* 274:34523-6.

Koide, T., S. Asada, Y. Takahara, Y. Nishikawa, K. Nagata, and K. Kitagawa. 2006a. Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation. *J. Biol. Chem.* 281:3432-8.

Koide, T., A. Aso, T. Yorihuzi, and K. Nagata. 2000. Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47. *J. Biol. Chem.* 275:27957-63.

Koide, T., and K. Nagata. 2005. Collagen biosynthesis. *Top. Curr. Chem.* 247:85.

Koide, T., Y. Nishikawa, S. Asada, C.M. Yamazaki, Y. Takahara, D.L. Homma, A. Otaka, K. Ohtani, N. Wakamiya, K. Nagata, and K. Kitagawa. 2006b. Specific recognition of the collagen triple helix by chaperone HSP47. II. The HSP47-binding structural motif in collagens and related proteins. *J. Biol. Chem.* 281:11177-85.

Koide, T., Y. Takahara, S. Asada, and K. Nagata. 2002. Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J. Biol. Chem.* 277:6178-82.

Koivu, J., R. Myllylä, T. Helaakoski, T. Pihlajaniemi, K. Tasanen, and K.I. Kivirikko. 1987. A single polypeptide acts both as the beta subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase. *J. Biol. Chem.* 262:6447-6449.

Koloczek, H., A. Banbula, J. Potempa, and G.S. Salvesen. 1996. Serpin α_1-proteinase inhibitor probed by intrinsic tryptophan fluorescence spectroscopy. *Protein Sci.* 5:2226-2235.

Kotch, F.W., and R.T. Raines. 2006. Self-assembly of synthetic collagen triple helices. *Proc. Natl. Acad. Sci. U. S. A.* 103:3028-3033.

Kramer, R.Z., J. Bella, B. Brodsky, and H.M. Berman. 2001. The crystal and molecular structure of a collagen-like peptide with A biologically relevant sequence. *J. Mol. Biol.* 311:131-147.

Kurkinen, M., A. Taylor, J.I. Garrels, and B.L. Hogan. 1984. Cell surface-associated proteins which bind native type IV collagen or gelatin. *J. Biol. Chem.* 259:5915-22.

Kvassman, J.-O., D.A. Lawrence, and J.D. Shore. 1995. The Acid Stabilization of Plasminogen Activator Inhibitor-1 Depends on Protonation of a Single Group That Affects Loop Insertion into β-Sheet A. *J. Biol. Chem.* 270:27942-27947.

Lakowicz, J.R. 2006. Principles of Fluorescence Spectroscopy. Springer, Baltimore, MD, USA.

Lamandé, S.R., S.D. Chessler, S.B. Golub, P.H. Byers, C. Chan, W.G. Cole, D.O. Sillence, and J.F. Bateman. 1995. Endoplasmic Reticulum-mediated Quality Control of Type I Collagen Production by Cells from Osteogenesis Imperfecta Patients with Mutations in the pro1(I) Chain Carboxyl-terminal Propeptide which Impair Subunit Assembly. *J. Biol. Chem.* 270:8642-8649.

Lang, K., F.X. Schmid, and G. Fischer. 1987. Catalysis of protein folding by prolyl isomerase. *Nature*. 329:268-270.

Leatherbarrow, R.J. 2009. GraFit Version 7. Erithacus Software Ltd., Horley, U.K.

Lee, D., J.D. Walsh, I. Mikhailenko, P. Yu, M. Migliorini, Y. Wu, S. Krueger, J.E. Curtis, B. Harris, S. Lockett, S.C. Blacklow, D.K. Strickland, and Y.-X. Wang. 2006. RAP Uses a Histidine Switch to Regulate Its Interaction with LRP in the ER and Golgi. *Mol. Cell*. 22:423-430.

Lee, S.-S., L.-H. Tseng, Y.-C. Li, C.-H. Tsai, and Y.-C. Chang. 2011. Heat shock protein 47 expression in oral squamous cell carcinomas and upregulated by arecoline in human oral epithelial cells. *J. Oral Pathol. Med.* 40:390-396.

Leikina, E., M.V. Mertts, N. Kuznetsova, and S. Leikin. 2002. Type I collagen is thermally unstable at body temperature. *Proc. Natl. Acad. Sci. U. S. A.* 99:1314-1318.

Lewis, M.J., D.J. Sweet, and H.R.B. Pelham. 1990. The ERD2 gene determines the specificity of the luminal ER protein retention system. *Cell*. 61:1359-1363.

Li, D., W. Guang, W.M. Abuzeid, S. Roy, G.-P. Gao, J.J. Sauk, and B.W. O'Malley. 2008. Novel adenoviral gene delivery system targeted against head and neck cancer. *Laryngoscope*. 118:650-8.

Li, H., A.D. Robertson, and J.H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. 61:704-721.

Liberek, K., J. Marszalek, D. Ang, C. Georgopoulos, and M. Zylicz. 1991. Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc. Natl. Acad. Sci. U. S. A.* 88:2874-2878.

Lodish, H., D. Baltimore, A. Berk, S.L. Zipursky, P. Matsudaira, and J. Darnell. 1995. Molecular Cell Biology. Scientific American Books, New York.

Loebermann, H., R. Tokuoka, J. Deisenhofer, and R. Huber. 1984. Human α_1-proteinase inhibitor : Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J. Mol. Biol.* 177:531-557.

Loewenthal, R., J. Sancho, and A.R. Fersht. 1991. Fluorescence spectrum of barnase: contributions of three tryptophan residues and a histidine-related pH dependence. *Biochemistry*. 30:6775-6779.

Loewenthal, R., J. Sancho, and A.R. Fersht. 1992. Histidine-aromatic interactions in barnase : Elevation of histidine pKa and contribution to protein stability. *J. Mol. Biol.* 224:759-770.

Lomas, D.A., D. Li-Evans, J.T. Finch, and R.W. Carrell. 1992. The mechanism of Z α_1-antitrypsin accumulation in the liver. *Nature*. 357:605-607.

Macdonald, J.R., and H.P. Bächinger. 2001. HSP47 binds cooperatively to triple helical type I collagen but has little effect on the thermal stability or rate of refolding. *J. Biol. Chem.* 276:25399-403.

Maitra, A., C. Iacobuzio-Donahue, A. Rahman, T.A. Sohn, P. Argani, R. Meyer, C.J. Yeo, J.L. Cameron, M. Goggins, S.E. Kern, R. Ashfaq, R.H. Hruban, and R.E. Wilentz. 2002. Immunohistochemical Validation of a Novel Epithelial and a Novel Stromal Marker of Pancreatic Ductal Adenocarcinoma Identified by Global Expression Microarrays. *Am. J. Clin. Pathol.* 118:52-59.

Makareeva, E., and S. Leikin. 2007. Procollagen Triple Helix Assembly: An Unconventional Chaperone-Assisted Folding Paradigm. *PLoS ONE.* 2:e1029.

Marassi, F.M., and S.J. Opella. 1998. NMR structural studies of membrane proteins. *Curr. Opin. Struct. Biol.* 8:640-648.

Masuda, H.H., M.M. Fukumoto, K.K. Hirayoshi, and K.K. Nagata. 1994. 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.* 94:2481-8.

Mathews, C.K., K.E. van Holde, and K.G. Ahern. 1999. Biochemistry. Benjamin/Cummings, San Francisco, USA.

Meagher, J.L., J.M. Beechem, S.T. Olson, and P.G. Gettins. 1998. Deconvolution of the fluorescence emission spectrum of human antithrombin and identification of the tryptophan residues that are responsive to heparin binding. *J. Biol. Chem.* 273:23283-9.

Mikhalyov, I., N. Gretskaya, F. Bergstrom, and L.B.A. Johansson. 2002. Electronic ground and excited state properties of dipyrrometheneboron difluoride (BODIPY): Dimers with application to biosciences. *Phys. Chem. Chem. Phys.* 4:5663-5670.

Moens, P.D.J., M.K. Helms, and D.M. Jameson. 2004. Detection of Tryptophan to Tryptophan Energy Transfer in Proteins. *Protein J.* 23:79-83.

Mohan, P.M.K., M. Barve, A. Chatterjee, and R.V. Hosur. 2006. pH driven conformational dynamics and dimer-to-monomer transition in DLC8. *Protein Sci.* 15:335-342.

Monera, O.D., T.J. Sereda, N.E. Zhou, C.M. Kay, and R.S. Hodges. 1995. Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix. *J. Pept. Sci.* 1:319-329.

Mottonen, J., A. Strand, J. Symersky, R.M. Sweet, D.E. Danley, K.F. Geoghegan, R.D. Gerard, and E.J. Goldsmith. 1992. Structural basis of latency in plasminogen activator inhibitor-1. *Nature*. 355:270.

Munch, M., H. Christian, P.H. Jensen, and P.A. Andreasen. 1991. Type-1 inhibitor of plasminogen activators : Distinction between latent, activated and reactive centre-cleaved forms with thermal stability and monoclonal antibodies. *FEBS Lett.* 295:102-106.

Myllylä, R., L. Tuderman, and K.I. Kivirikko. 1977. Mechanism of the Prolyl Hydroxylase Reaction. *Eur. J. Biochem.* 80:349-357.

Nagai, N., M. Hosokawa, S. Itohara, E. Adachi, T. Matsushita, N. Hosokawa, and K. Nagata. 2000. Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J. Cell Biol.* 150:1499-506.

Nagata, K. 2003. HSP47 as a collagen-specific molecular chaperone: function and expression in normal mouse development. *Semin. Cell Dev. Biol.* 14:275-82.

Nagata, K., S. Saga, and K.M. Yamada. 1986. A major collagen-binding protein of chick embryo fibroblasts is a novel heat shock protein. *J. Cell Biol.* 103:223.

Nakai, A., M. Satoh, K. Hirayoshi, and K. Nagata. 1992. Involvement of the stress protein HSP47 in procollagen processing in the endoplasmic reticulum. *J. Cell Biol.* 117:903-14.

Natsume, T., T. Koide, S. Yokota, K. Hirayoshi, and K. Nagata. 1994. Interactions between collagen-binding stress protein HSP47 and collagen. Analysis of kinetic parameters by surface plasmon resonance biosensor. *J. Biol. Chem.* 269:31224-8.

Nishikawa, Y., Y. Takahara, S. Asada, A. Shigenaga, A. Otaka, K. Kitagawa, and T. Koide. 2010. A structure-activity relationship study elucidating the mechanism of sequence-specific collagen recognition by the chaperone HSP47. *Bioorg. Med. Chem.* 18:3767-3775.

Nishino, T., M. Miyazaki, K. Abe, A. Furusu, Y. Mishima, T. Harada, Y. Ozono, T. Koji, and S. Kohno. 2003. Antisense oligonucleotides against collagen-binding stress protein HSP47 suppress peritoneal fibrosis in rats. *Kidney Int*. 64:887-896.

O'Farrell, P.A., and L. Joshua-tor. 2007. Mutagenesis and crystallographic studies of the catalytic residues of the papain family protease bleomycin hydrolase: new insights into active-site structure. *Biochem J*. 401:421-428.

Ogawa, Y., M.S. Razzaque, K. Kameyama, G. Hasegawa, S. Shimmura, M. Kawai, S. Okamoto, Y. Ikeda, K. Tsubota, Y. Kawakami, and M. Kuwana. 2007. Role of Heat Shock Protein 47, a Collagen-Binding Chaperone, in Lacrimal Gland Pathology in Patients with cGVHD. *Invest. Ophthalmol. Vis. Sci.* 48:1079.

Ohmae, E., Y. Sasaki, and K. Gekko. 2001. Effects of Five-Tryptophan Mutations on Structure, Stability and Function of Escherichia coli Dihydrofolate Reductase. *J. Biochem.* 130:439-447.

Okano-Kosugi, H., O. Matsushita, S. Asada, A.B. Herr, K. Kitagawa, and T. Koide. 2009. Development of a high-throughput screening system for the compounds that inhibit collagen-protein interactions. *Anal. Biochem.* 394:125-131.

Okuyama, K., C. Hongo, R. Fukushima, G. Wu, H. Narita, K. Noguchi, Y. Tanaka, and N. Nishino. 2004. Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Pept. Sci.* 76:367-377.

Okuyama, K., K. Okuyama, S. Arnott, M. Takayanagi, and M. Kakudo. 1981. Crystal and molecular structure of a collagen-like polypeptide (Pro-Pro-Gly)10. *J. Mol. Biol.* 152:427-443.

Ou, W.-J., P.H. Cameron, D.Y. Thomas, and J.J.M. Bergeron. 1993. Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature*. 364:771-776.
Pain, R.H. 2004. Determining the fluorescence spectrum of a protein. *Current Protocols in Protein Science*:7.7.1 - 7.7.20.
Paroutis, P., N. Touret, and S. Grinstein. 2004. The pH of the Secretory Pathway: Measurement, Determinants, and Regulation. *Physiology*. 19:207-215.
Pearce, M.C., L.D. Cabrita, A.M. Ellisdon, and S.P. Bottomley. 2007. The loss of tryptophan 194 in antichymotrypsin lowers the kinetic barrier to misfolding. *FEBS J.* 274:3622-3632.
Persikov, A.V., Y. Xu, and B. Brodsky. 2004. Equilibrium thermal transitions of collagen model peptides. *Protein Sci.* 13:893-902.
Pires, M.M., and J. Chmielewski. 2009. Self-assembly of Collagen Peptides into Microflorettes via Metal Coordination. *J. Am. Chem. Soc.* 131:2706-2712.
Potempa, J., E. Korzus, and J. Travis. 1994. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* 269:15957-60.
Prockop, D.J., K.I. Kivirikko, L. Tuderman, and N.A. Guzman. 1979. The biosynthesis of collagen and its disorders (first of two parts). *N. Engl. J. Med.* 301:13-23.
Przybyla, D.E., and J. Chmielewski. 2010. Higher-Order Assembly of Collagen Peptides into Nano- and Microscale Materials. *Biochemistry*. 49:4411-4419.
Ramachandran, G.N., and R. Chandrasekharan. 1968. Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers*. 6:1649-1658.
Ramachandran, G.N., and G. Kartha. 1955. Structure of Collagen. *Nature*. 176:593-595.
Rasband, W.S. 1997-2009. ImageJ. *In* http://rsb.info.nih.gov/ij/. U. S. National Institutes of Health, Bethesda, Maryland, USA.
Razzaque, M.S., M.A. Hossain, and T. Taguchi. 1998. Bleomycin-induced pulmonary fibrosis in rat is associated with increased expression of collagen-binding heat shock protein (HSP) 47. *Virchows Arch.* 432:455-460.
Rich, A., and F.H.C. Crick. 1955. The Structure of Collagen. *Nature*. 176:915-916.
Rodriguez, R., G. Chinea, N. Lopez, T. Pons, and G. Vriend. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics*. 14:523-528.
Rosen, C.G., and G. Weber. 1969. Dimer formation from 1-anilino-8-naphthalenesulfonate catalyzed by bovine serum albumin. Fluorescent molecule with exceptional binding properties. *Biochemistry*. 8:3915-3920.
Saga, S., K. Nagata, W.T. Chen, and K.M. Yamada. 1987. pH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein. *J. Cell Biol.* 105:517-27.
Sakakibara, S., K. Inouye, K. Shudo, Y. Kishida, Y. Kobayashi, and D.J. Prockop. 1973. Synthesis of (Pro-Hyp-Gly)n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. *BBA-Protein Struct.* 303:198-202.
Sakakibara, S., Y. Kishida, K. Okuyama, N. Tanaka, T. Ashida, and M. Kakudo. 1972. Single crystals of (Pro-Pro-Gly)10, a synthetic polypeptide model of collagen. *J. Mol. Biol.* 65:371-372.
Sarkar, S.K., P.E. Young, C.E. Sullivan, and D.A. Torchia. 1984. Detection of cis and trans X-Pro peptide bonds in proteins by 13C NMR: application to collagen. *Proc. Natl. Acad. Sci. U. S. A.* 81:4800-4803.
Sato, Y., K. Murase, J. Kato, M. Kobune, T. Sato, Y. Kawano, R. Takimoto, K. Takada, K. Miyanishi, T. Matsunaga, T. Takayama, and Y. Niitsu. 2008. Resolution of liver

cirrhosis using vitamin A-coupled liposomes to deliver siRNA against a collagen-specific chaperone. *Nat Biotech*. 26:431-442.
Satoh, M., K. Hirayoshi, S. Yokota, N. Hosokawa, and K. Nagata. 1996. Intracellular interaction of collagen-specific stress protein HSP47 with newly synthesized procollagen. *J. Cell Biol.* 133:469-83.
Sawano, A., and A. Miyawaki. 2000. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* 28:E78.
Scheel, A.A., and H.R.B. Pelham. 1998. Identification of Amino Acids in the Binding Pocket of the Human KDEL Receptor. *J. Biol. Chem.* 273:2467-2472.
Schmid, F.X. 1997. Optical spectroscopy to characterize protein conformation and conformational changes. *In* Protein structure: a practical approach. T.E. Creighton, editor. Oxford University Press, New York.
Schreuder, H.A., B. de Boer, R. Dijkema, J. Mulders, H.J.M. Theunissen, P.D.J. Grootenhuis, and W.G.J. Hol. 1994. The intact and cleaved human antithrombin III complex as a model for serpin–proteinase interactions. *Nat. Struct. Mol. Biol.* 1:48.
Schwede, T., A. Sali, N. Eswar, and M.C. Peitsch. 2008. Protein Structure Modeling. *In* Computational Structural Biology: Methods and Applications. T. Schwede and M.C. Peitsch, editors. World Scientific Publishing Co.
Sedgwick, H., J.E. Cameron, W.C.K. Poon, and S.U. Egelhaaf. 2007. Protein phase behavior and crystallization: Effect of glycerol. *J. Chem. Phys.* 127:125102.
Shinitzky, M., and R. Goldman. 1967. Fluorometric Detection of Histidine-Tryptophan Complexes in Peptides and Proteins. *Eur. J. Biochem.* 3:139-144.
Silverman, G.A., P.I. Bird, R.W. Carrell, F.C. Church, P.B. Coughlin, P.G.W. Gettins, J.A. Irving, D.A. Lomas, C.J. Luke, R.W. Moyer, P.A. Pemberton, E. Remold-O'Donnell, G.S. Salvesen, J. Travis, and J.C. Whisstock. 2001. 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.* 276:33293-33296.
Simmons, D. 2008. The use of animal models in studying genetic disease: transgenesis and induced mutation. *Nature Education*. 1.
Simonovic, M., P.G.W. Gettins, and K. Volz. 2001. Crystal structure of human PEDF, a potent anti-angiogenic and neurite growth-promoting factor. *Proc. Natl. Acad. Sci. U. S. A.* 98:11131-11135.
Simpson, R.B., and W. Kauzmann. 1953. The Kinetics of Protein Denaturation. I. The Behavior of the Optical Rotation of Ovalbumin in Urea Solutions. *J. Am. Chem. Soc.* 75:5139-5152.
Stampfer, S.D., H. Lou, G.H. Cohen, R.J. Eisenberg, and E.E. Heldwein. 2010. Structural Basis of Local, pH-Dependent Conformational Changes in Glycoprotein B from Herpes Simplex Virus Type 1. *J. Virol.* 84:12924-12933.
Stein, P.E., and R.W. Carrell. 1995. What do dysfunctional serpins tell us about molecular mobility and disease? *Nat. Struct. Biol.* 2:96–113.
Stein, P.E., A.G.W. Leslie, J.T. Finch, W.G. Turnell, P.J. McLaughlin, and R.W. Carrell. 1990. Crystal structure of ovalbumin as a model for the reactive centre of serpins. *Nature*. 347:99-102.
Steinberg, I.Z. 1971. Long-Range Nonradiative Transfer of Electronic Excitation Energy in Proteins and Polypeptides. *Annu. Rev. Biochem.* 40:83-114.

Steinmann, B., P. Bruckner, and A. Superti-Furga. 1991. Cyclosporin A slows collagen triple-helix formation in vivo: indirect evidence for a physiologic role of peptidyl-prolyl cis-trans-isomerase. *J. Biol. Chem.* 266:1299-1303.

Stratikos, E., and P.G.W. Gettins. 1999. 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 β-sheet A. *Proc. Natl. Acad. Sci. U. S. A.* 96:4808-4813.

Strickland, E.H., and S. Beychok. 1974. Aromatic Contributions To Circular Dichroism Spectra Of Protein. *Crit. Rev. Biochem. Mol. Biol.* 2:113-175.

Stryer, L. 1995. Biochemistry. W. H. Freeman and Company, New York.

Sui, G.-C., and B. Wiman. 2000. The B β-sheet in the PAI-1 Molecule Plays an Important Role for its Stability. *Thromb. Haemost.* 83:896-901.

Takashi, R., Y. Tonomura, and M.F. Morales. 1977. 4,4'-Bis (1-anilinonaphthalene 8-sulfonate) (bis-ANS): a new probe of the active site of myosin. *Proc. Natl. Acad. Sci. U. S. A.* 74:2334-2338.

Takechi, H., K. Hirayoshi, A. Nakai, H. Kudo, S. Saga, and K. Nagata. 1992. Molecular cloning of a mouse 47-kDa heat-shock protein (HSP47), a collagen-binding stress protein, and its expression during the differentiation of F9 teratocarcinoma cells. *FEBS J.* 206:323.

Tanford, C., C.E. Buckley, P.K. De, and E.P. Lively. 1962. Effect of Ethylene Glycol on the Conformation of γ-Globulin and β-Lactoglobulin. *J. Biol. Chem.* 237:1168-1171.

Tettamanzi, M.C., C. Keeler, S. Meshack, and M.E. Hodsdon. 2008. Analysis of Site-Specific Histidine Protonation in Human Prolactin. *Biochemistry*. 47:8638-8647.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.

Thomson, C.A., and V.S. Ananthanarayanan. 2000. Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro. *Biochem. J.* 349 Pt 3:877-83.

Thomson, C.A., H.M. Atkinson, and V.S. Ananthanarayanan. 2005. Identification of small molecule chemical inhibitors of the collagen-specific chaperone Hsp47. *J. Med. Chem.* 48:1680-4.

Thomson, C.A., R. Tenni, and V.S. Ananthanarayanan. 2003. Mapping Hsp47 binding site(s) using CNBr peptides derived from type I and type II collagen. *Protein Sci.* 12:1792-800.

Tifany, M.L., and S. Krimm. 1972. Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers*. 11:2309-2316.

Tombran-Tink, J., S. Aparicio, X. Xu, A.R. Tink, N. Lara, S. Sawant, C.J. Barnstable, and S.S.-M. Zhang. 2005. PEDF and the serpins: Phylogeny, sequence conservation, and functional domains. *J. Struct. Biol.* 151:130-150.

Tucker, H.M., J. Mottonen, E.J. Goldsmith, and R.D. Gerard. 1995. Engineering of plasminogen activator inhibitor-1 to reduce the rate of latency transition. *Nat. Struct. Biol.* 2:442 - 445.

Tuderman, L., R. Myllylä, and K.I. Kivirikko. 1977. Mechanism of the Prolyl Hydroxylase Reaction. 1. Role of Co-substrates. *Eur. J. Biochem.* 80:341-348.

Van Gilst, M., and B.S. Hudson. 1996. Histidine-tryptophan interactions in T4 lysozyme: 'Anomalous' pH dependence of fluorescence. *Biophys. Chem.* 63:17-25.

Vasques, M.T., M.A. Alves, V., J. Cerqueira Luz, G. de, and L. Correa. 2010. Immunolocalization of heat shock proteins 27 and 47 during repair of induced oral ulcers. *J. Oral Sci.* 52:623-631.
Vitale, A., and J. Denecke. 1999. The Endoplasmic Reticulum—Gateway of the Secretory Pathway. *Plant Cell*. 11:615-628.
Voet, D., and J.D. Voet. 1995. Biochemistry. John Wiley & Sons, New York.
Voet, D., and J.D. Voet. 2005. Biochemistry. John Wiley & Sons, New York.
Vos, R., and Y. Engelborghs. 1994. A Fluorescence Study of Tryptophan-Histidine Interactions in the Peptide Anantin and in Solution. *Photochem. Photobiol.* 60:24-32.
Vuilleumier, S., J. Sancho, R. Loewenthal, and A.R. Fersht. 1993. Circular dichroism studies of barnase and its mutants: Characterization of the contribution of aromatic side chains. *Biochemistry*. 32:10303-10313.
Wada, I., D. Rindress, P.H. Cameron, W.J. Ou, J.J. Doherty, D. Louvard, A.W. Bell, D. Dignard, D.Y. Thomas, and J.J. Bergeron. 1991. SSR alpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. *J. Biol. Chem.* 266:19599-19610.
Walmsley, A.R., M.R. Batten, U. Lad, and N.J. Bulleid. 1999. Intracellular Retention of Procollagen within the Endoplasmic Reticulum Is Mediated by Prolyl 4-Hydroxylase. *J. Biol. Chem.* 274:14884-14892.
Wang, S.Y., and L.J. Gudas. 1990. A retinoic acid-inducible mRNA from F9 teratocarcinoma cells encodes a novel protease inhibitor homologue. *J. Biol. Chem.* 265:15818-22.
Wang, Z.L., T. Inokuchi, H. Ikeda, T.T. Baba, M. Uehara, N. Kamasaki, K. Sano, T.K. Nemoto, and T. Taguchi. 2002. Collagen-binding heat shock protein HSP47 expression during healing of fetal skin wounds. *Int. J. Oral Maxillofac. Surg.* 31:179-184.
Wei, A., H. Rubin, B.S. Cooperman, and D.W. Christianson. 1994. Crystal structure of an uncleaved serpin reveals the conformation of an inhibitory reactive loop. *Nat. Struct. Mol. Biol.* 1:251.
Whisstock, J.C., R. Skinner, R.W. Carrell, and A.M. Lesk. 2000. Conformational changes in serpins: I. the native and cleaved conformations of α_1-antitrypsin. *J. Mol. Biol.* 295:651-665.
Whitmore, L., and B.A. Wallace. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32:W668-W673.
Whitmore, L., and B.A. Wallace. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*. 89:392-400.
Willaert, K., and Y. Engelborghs. 1991. The quenching of tryptophan fluorescence by protonated and unprotonated imidazole. *Eur. Biophys. J.* 20:177-182.
Williams, B.R., R.A. Gelman, D.C. Poppke, and K.A. Piez. 1978. Collagen fibril formation. Optimal in vitro conditions and preliminary kinetic results. *J. Biol. Chem.* 253:6578-85.
Willnow, T.E., S.A. Armstrong, R.E. Hammer, and J. Herz. 1995. Functional expression of low density lipoprotein receptor-related protein is controlled by receptor-associated protein in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 92:4537-4541.
Willnow, T.E., A. Rohlmann, J. Horton, H. Otani, J.R. Braun, R.E. Hammer, and J. Herz. 1996. RAP, a specialized chaperone, prevents ligand-induced ER retention and degradation of LDL receptor-related endocytic receptors. *EMBO J.* 15:2632.

Wilson, D.W., M.J. Lewis, and H.R. Pelham. 1993. pH-dependent binding of KDEL to its receptor in vitro. *J. Biol. Chem.* 268:7465-7468.
Wilson, R., J.F. Lees, and N.J. Bulleid. 1998. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *J. Biol. Chem.* 273:9637.
Wood, G.C. 1960. The formation of fibrils from collagen solutions. 2. A mechanism of collagen-fibril formation. *Biochem. J.* 75:598-0.
Wu, M.M., M. Grabe, S. Adams, R.Y. Tsien, H.-P.H. Moore, and T.E. Machen. 2001. Mechanisms of pH Regulation in the Regulated Secretory Pathway. *J. Biol. Chem.* 276:33027-33035.
Xu, K., J. Cui, V. Olman, Q. Yang, D. Puett, and Y. Xu. 2010. A Comparative Analysis of Gene-Expression Data of Multiple Cancer Types. *PLoS ONE.* 5:e13696.
Yamazaki, C.M., S. Asada, K. Kitagawa, and T. Koide. 2008. Artificial collagen gels via self-assembly of de novo designed peptides. *Pept. Sci.* 90:816-823.
Yang, Q., S. Liu, Y. Tian, C. Hasan, D. Kersey, H.R. Salwen, A. Chlenski, E.J. Perlman, and S.L. Cohn. 2004. Methylation-Associated Silencing of the Heat Shock Protein 47 Gene in Human Neuroblastoma. *Cancer Res.* 64:4531-4538.
Yang, S.F., C.-H. Tsai, and Y.-C. Chang. 2008. The upregulation of heat shock protein 47 expression in human buccal fibroblasts stimulated with arecoline. *J. Oral Pathol. Med.* 37:206.
Yannariello-Brown, J., and J.A. Madri. 1990. A 48 kDa collagen-binding phosphoprotein isolated from bovine aortic endothelial cells interacts with the collagenous domain, but not the globular domain, of collagen type IV. *Biochem. J.* 265:383.
Yonath, A., and W. Traub. 1969. Polymers of tripeptides as collagen models : IV. Structure analysis of poly(-prolyl-glycyl--proline). *J. Mol. Biol.* 43:461-462.
Zhong, D., S.K. Pal, D. Zhang, S.I. Chan, and A.H. Zewail. 2002. Femtosecond dynamics of rubredoxin: Tryptophan solvation and resonance energy transfer in the protein. *Proc. Natl. Acad. Sci. U. S. A.* 99:13-18.
Zhou, A., Z. Wei, R.J. Read, and R.W. Carrell. 2006. Structural mechanism for the carriage and release of thyroxine in the blood. *Proc. Natl. Acad. Sci. U. S. A.* 103:13321-13326.