The HSP47 - Procollagen Interaction: Mechanism of pH-Dependent Client Release and Development of Antifibrotic Inhibitors

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen

Fakültät der Universität zu Köln

Vorgelegt von

Sinan Öcal

aus Istanbul

Köln, September 2018

Gutachter:	Prof. Dr. Ulrich Baumann Institut für Biochemie / Universität zu Köln
	Prof. Dr. Karsten Niefind Institut für Biochemie / Universität zu Köln
Prüfungsvorsitzender:	Prof. Dr. Eric von Elert Institut für Zoologie / Universität zu Köln

Tag der mündlichen Prüfung: 13. 11. 2017

Die Arbeiten und Experimente zur vorliegenden Dissertation wurden zwischen Oktober 2012 und September 2017 unter Betreuung von Prof. Dr. Ulrich Baumann am Institut für Biochemie der Universität zu Köln, Otto-Fischer Str. 12-14, D-50674 Köln, Deutschland, durchgeführt.

Teile dieser Arbeit wurden bereits veröffentlicht:

Oecal S, Socher E, Uthoff M, et al. The pH-dependent Client Release from the Collagen-specific Chaperone HSP47 Is Triggered by a Tandem Histidine Pair. *The Journal of Biological Chemistry*. 2016;291(24):12612-12626.

Table of Contents

Zusammenfassung

Kollagene sind eine äußerst vielseitige Superfamilie von Proteinen, welche primär als strukturgebende Hauptkomponente der Extrazellulären Matrix (EM) von Metazoen bekannt sind. Aufgrund ihrer zentralen Rolle beim Aufbau und der Dynamik der EM sind Kollagene an einer Vielzahl an unterschiedlichen Prozessen wie z.B. der Zelladhäsion, Zellmigration, Angiogenese, Morphogenese, Immunantwort oder Krebsprogression beteiligt. Die Biosynthese dieser großen, stark post-translational modifizierten Proteine wird durch eine fein aufeinander abgestimmte Maschinerie bewerkstelligt. Die Komplexität dieser Maschinerie äußert sich durch ihre Anfälligkeit gegenüber Störungen, welche oft zur Entstehung von schwerwiegenden Defekten und Krankheiten führen. Das molekulare Chaperon von Kollagen, Hitzeschockprotein 47 (HSP47), spielt hierbei eine entscheidende Rolle und gewährleistet sowohl die fehlerfreie Synthese und den Transport von Kollagenen als auch den Schutz des Endoplasmatischen Retikulums (ER) vor durch Aggregation und Akkumulation von Kollagenen bedingtem Stress. Obwohl HSP47 seit mehr als zwei Jahrzehnten Forschungsgegenstand ist, sind viele, teilweise grundlegende Aspekte seiner Funktion immer noch unbekannt; es ist z.B. unbekannt, welchen Vorteil die Serinproteasestruktur für die Proteinfunktion bringt, oder ob alle Kollagentypen zum Substratrepertoire gehören. Die kürzlich gelöste Kristallstruktur von HSP47 im Komplex mit einem Kollagenmodellpeptid hat der Beantwortung der offenen Fragen neuen Anstoß gegeben. Eine betrifft den Mechanismus der Substratfreisetzung: untypisch für Chaperone wird die Funktion von HSP47 nicht durch Austausch oder Hydrolyse von Nukleotiden reguliert, sondern durch die graduelle Abnahme des pH-Wertes im sekretorischen Weg. Die molekularen Details dieser pH-induzierten Substratfreisetzung sind ungeklärt und wurden im ersten Teil der vorliegenden Arbeit untersucht. Ausgehend von der Kristallstruktur und Phylogenetischen Daten wurden gezielt HSP47-Punktmutanten generiert und deren Bindung an Kollagenmodellpeptide mittels Biolayerinterferometrie kinetisch quantifiziert. Besonderes Interesse galt hierbei den 14 Histidinresten von HSP47, welche Aufgrund Ihrer Fähigkeit, bei physiologischen pH-Werten Protonen aufnehmen zu können, als potentielle Auslöser einer Substratfreisetzung in Betracht kamen. Eine systematische Analyse dieser Aminosäuren enthüllte dass His273 und His274, am Rande der Interaktionsfläche gelegen, großen Einfluss auf die pH-Sensitivität der HSP47 - Kollagen Interaktion ausüben. Andere Histidinreste wurden ebenfalls als wichtige Komponenten der Interaktion ausgemacht, wie z.B. His238, welches bei der korrekten Ausrichtung von für die Bindung essentiellen Aminosäureseitenketten eine Rolle spielt. Versuche, HSP47 bei leicht saurem Milieu zu Kristallisieren waren Aufgrund der verminderten Stabilität des Proteins unter solchen Bedingungen nicht erfolgreich; somit steht eine umfassende, strukturelle Erklärung des Substratfreisetzungsmechanismus noch aus.

In den letzten Jahren haben mehrere Studien gezeigt, dass Genablation von HSP47 mittels siRNA zu einer deutlichen Verbesserung bis hin zur Aufhebung von fibrotischen Krankheiten führen kann. HSP47 wurde dadurch als ein vielversprechendes Ziel für die Entwicklung antifibrotischer Medikamente erkannt. Vor diesem Hintergrund wurde in dieser Arbeit ein auf Fluoreszenz basierender Assay entwickelt, welcher mittels eines Hochdurchsatzverfahens zur Identifizierung neuartiger Inhibitoren des HSP47 - Kollagen Komplexes herangezogen werden kann. Der Assay wurde genutzt, um aus einer 40.000 Chemikalien umfassenden Substanzbibliothek heraus 4 potentielle Inhibitorkandidaten zu identifizieren, von welchen mindestens eine Substanz in Validationsexperimenten vielversprechende Ergebnisse lieferte.

Abstract

Collagens are a multifaceted superfamily of proteins which constitute the principal structural component of the extracellular matrix (ECM) of metazoan organisms. Intimately tied to ECM architecture and dynamics, collagens are involved in a multitude of processes such as cell adhesion and migration, angiogenesis, morphogenesis, the immune response and cancer progression. The biosynthesis of these large and complex molecules is reliant on a finely tuned machinery, disruptions of which is often causative to severe disease. Heat-shock protein 47 (HSP47), the molecular chaperone of collagen, plays a central role in ensuring proper processing and trafficking of collagens as well as the protection of the endoplasmatic reticulum (ER) from stress induced by aggregation and accumulation of its client. Although HSP47 has been first described more than 20 years ago, critical aspects of its function are still shrouded in mystery, ranging from why it is a member of the serpin superfamily of proteins to whether its client repertoire includes all types of collagen. The rather recently solved crystal structure of HSP47 in complex with a collagen model peptide has provided a new impetus for answering the many open questions. One of these is how client-release is achieved: untypical for chaperones, HSP47 function is not coupled to nucleotide hydrolysis or exchange, but governed by the gradual decrease in pH along the secretory pathway. The mechanism by which the pH-shift induces client release is unclear. In this work, it was investigated whether this process is based on conformational re-arrangements, more subtle distortions of the binding site or electrostatic repulsion. Of particular interest were the 14 histidine residues in HSP47, which have long been considered as potential trigger residues, since their imidazole side-chains can serve as a proton acceptor at physiological pH. Systematic analysis of these histidines in context of this thesis has revealed that His273 and His274, located at the fringe of the binding interface, exert considerable influence on the pH-sensitivity of the HSP47 - collagen complex. Other histidines have also been found to be important for the interaction; most notably, His238 was shown to be an essential actor in the pre-arrangement of key residues in a client-binding competent conformation.

Over the recent years, studies using gene ablation via siRNA have shown that interfering with the HSP47 - collagen complex can resolve a variety of fibrotic diseases. HSP47 has thus enjoyed increasing attention as a potential target for anti-fibrotic drugs. In light of this, one part of this thesis has focused on developing a fluorescence based, high-throughput screening compatible assay to be utilized for the identification of novel inhibitors of the HSP47 - collagen interaction. Interrogation of a compound library using the assay has yielded 4 potential inhibitor candidates, at least one of which having shown promising results in initial validation studies.

1. Introduction

1.1 Collagen

1.1.1 The Collagen Superfamily: An Overview

Collagens are a remarkably diverse superfamily of proteins which are best known for their prominent role in the architecture of the extracellular matrix (ECM) of animals. Collagen is the most abundant protein in vertebrates, constituting almost a third of total protein mass - a consequence of the ubiquity of the extracellular matrix and the markedly low turnover of many collagens, some of which being practically permanent, as for instance collagens in adult cartilaginous tissue[1]. By and large the principal function of collagens is to provide structural integrity and scaffolding to the extracellular space. Collagens fulfil this role via assembling into various superstructures with extraordinary mechanical properties: the elastic modulus of a single collagen I fibril from rat tail, for instance, has been determined to be in the range of 1-10 GPa, coming close to that of typical metals and showcasing formidable tensile strength at a much lower density[2]. Besides these well-characterized structural roles, collagens are also involved in a multitude of dynamic phenomena such as cellular adhesion, migration or chemotaxis. Products of collagen processing are often important biochemical effectors: endostatin, a C-terminal domain of multiplexin collagens, has been shown to inhibit angiogenesis upon proteolytic release and is currently in clinical trials as an anti-tumour agent[3,4], while the C-propeptide of collagen type I is believed to play a key regulatory role both in fibrillogenesis and collagen biosynthesis[5]. Similarly, membrane collagens are known not only to function as cell surface receptors but also serve as a reservoir for effectors in the form of shed ectodomains. Another example for the multifaceted nature of collagens is collagen type VI, which besides forming beaded filaments, an important structural element of the ECM, is also capable of disrupting bacterial membranes and as such involved in innate host defense of the ECM[6].

Despite its omnipresence and biochemical significance, the definition of what is a collagen and what is not is still blurry at best. Three features are widely considered to be hallmarks of collagens[7]:

(i) The presence of at least one so-called collagenous domain, characterized by a triple-helical structure comprised of three separate polypeptide chains with repeating triplets of Gly-Xaa-Yaa (Xaa, Yaa = any amino acid),
(ii) Localization in the extracellular space, and
(iii) Participation in the assembly of superstructures.

These hallmarks are not always sufficient to fully characterize collagens, though, since there are exceptions aplenty: membrane collagens, for example, do not assemble into superstructures and are

strictly seen components of the cell membrane. The triple-helical structural motif, on the other hand, is also encountered in other proteins such as adiponectin, the asymetric form of acetylcholinesterase, C1q, macrophage scavenger receptors (MARCO), ficolins, collectins and many others[7]. Emilins and emu proteins fulfil all three criteria of being collagens, but are not (yet) classified as such[8].

Collagens and related molecules are not exclusive to vertebrates and have been discovered in many different multicellular organisms, including primordial forms of life such as Sponges[9] and Hydra[10], or rather extraordinary ones, such as the annelid *Alvinella pompeiana,* one of the most heat-tolerant metazoan known to date[11]. A single, 54 bp long sequence is believed to be ancestral to fibrillar collagens, which subsequently arose via multiple duplications of this basic genetic unit[10,12]. Such collagen-related structural motifs (CSM), although rare, are also found in unicellular eukaryotes and bacteria[13]; the streptococcal cell surface proteins Scl1 and Scl2, for instance, contain extended Gly-Xaa-Yaa repeats in their amino acid sequences[10], as does the exosporium filament BclA of *Bacillus anthracis*[14,15]. An intact collagen gene has even been discovered in the viral shrimp pathogen White Spot Bacilliform Virus[16]. It is quite possible that prokaryotic and viral collagen genes are the result of horizontal gene transfer and that the appearance of collagens is tied to the emergence of a hypothetical ancestor of metazoans[9]; this would imply that the acquisition and retention of such genes by unicellular organisms provides advantages in the interaction with multicellular hosts.

1.1.2 Composition and Classification of Collagens

Collagens are multidomain proteins comprised of three individual polypeptide chains, called collagen α-chains. The primary structure of α-chains characteristically includes extended regions consisting of repeating Gly-Xaa-Yaa triplets, where Xaa is frequently proline (28%) and Yaa (2S,4R)-4-hydroxyproline (38%)[17]. As will be discussed in more detail below, these proline-rich regions are responsible for the formation of the distinctive collagen triple-helix. In collagens, extensive triple-helical stretches, called collagenous domains, are typically interspersed with short non-collagenous (i.e. non triple-helical) domains, referred to as interruptions, and flanked by globular N- and C-terminal domains. A testament to the complexity of collagens, the identity and processing of these N- and C-terminal domains can vary greatly between different types of collagens.

As of today, 28 types of collagens (designated with Roman numerals I-XXVIII) with a total of 46 unique* α-chains have been characterized in humans (The "novel" collagen α-chain XXIX α1 turned out to be identical to collagen VI α5). For most collagen types, only one α-chain is known; others have multiple unique α-chains (up to 6 for collagen IV, and collagen VI in some mammals) which are distinguished via Arabic numerals. Since some collagens have several isoforms with differing chain compositions and others even form hybrid structures with α-chains of different collagens, the exact definition of a collagen molecule requires the specification of all three incorporated α-chains (e.g.

*α1(II) and α3(XI) share the same sequence but diverge in posttranslational processing and cross-linking[18].

$[\alpha1(I)]_3$ for homotrimeric collagen type I and $[\alpha1(I)]_2$, $\alpha2(I)]$ for heterotrimeric collagen type I).

Collagens are commonly classified according to their domain organization and/or the superstructures they form. The subfamilies include fibril forming collagens (I, II, III, V, XI, XXIV and XXVII), fibril associated collagens with interrupted triple-helices (FACIT) (IX, XII, XIV, XVI, XIX, XX, XXI, XXII), membrane associated collagens with interrupted triple-helices (MACIT) (XIII, XVII, XXIII, XXV), collagens with multiple triple-helical domains with interruptions (Multiplexins) (XV, XVIII), and network forming collagens (IV, VIII, X). Furthermore, there are collagens forming beaded filaments (collagen VI) and anchoring fibrils (collagen VII) as well as some which do not quite fit into any category (XXVI, XXVIII). Since a detailed discussion of the differences in structure and biosynthesis between these classes of collagens would be beyond the scope of this work, the following sections will focus on fibril forming collagens, which are the best characterized and have been subject to study in this work.

Fibril Forming Collagens

Fibrillar collagens are the most abundant in terms of total protein mass: collagen type I alone comprises more than 90% of all collagen in the body. The subfamily encompasses collagens of type I, II and III (type A clade, also called major fibrillar collagens in regard to their quantity), collagens type V and XI (type B clade, minor fibrillar collagens) and finally the novel collagens type XXIV and XXVII (type C clade)[19]. The latter are rather similar to collagens found in invertebrates, being shorter and containing imperfections/interruptions in their triple-helical domain[20].

Figure 1.1 shows the domain organization of a prototypical fibrillar collagen. Collagenous domains (COL) and non-collagenous domains (NC) are numbered proceeding from C- to N-terminus (this is consistent for all fibrillar collagens, but there are cases in which the numbering is reversed, such as MACIT collagen type XIII). Fibrillar collagens of the same clade typically share N-terminal NC domains: these are homologous to von Willebrand factor type C (vWFC, also called chordin-like cysteine rich repeats) in members of clade A and to thrombospondin-1 N-terminus-like domain (TSPN)

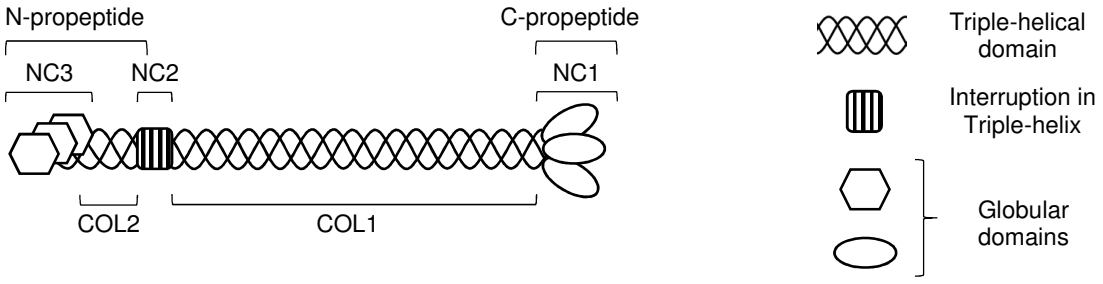


Figure 1.1 | Domain organization of a prototypical fibrillar collagen. An extended and in most cases uninterrupted collagenous domain (COL1) is flanked N- and C-terminally by non-collagenous domains (NC2 and NC1, respectively). Proteolytic cleavage within the NC2 region results in release of the N-propeptide, consisting of a short collagenous domain (COL2) and a globular domain (NC3), usually TSPN or vWFC. The C-propeptide is synonymous with the NC1 domain.

in members of clades B and C[19]. Unsurprisingly for collagens, there are exceptions: the N-terminal NC domain seems to have been deleted in case of α2(I), and α2(V) harbors a cysteine-rich repeat even though it belongs to clade B. The C-terminal NC1 domain, however, is highly conserved among fibrillar collagens and is important for trimerization and correct registration during collagen maturation.

1.1.3 Structure and Stability of the Collagenous Domain

The Polyproline Helix Type II

The PPII helix is a much neglected secondary structure element that is not only encountered in fibrillar proteins such as collagen, but is also widely dispersed in natively folded globular proteins, most often over short stretches of 4 or 5 amino acids and frequently preceding/transitioning into α-helices, β-sheets, 3_{10}-helices or reverse turns[21]. Analysis of protein structures deposited to the Protein Data Bank archive (PDB) of the Research Collaboratory for Structural Bioinformatics (RCSB) shows that around 2% of amino acids are incorporated into PPII helices[22]. The PPII helix is particularly important as the structural motif which is recognized by the abundant proline recognition domains such as SH3, WW or EVH1[23,24]. The PPII helical conformation is furthermore believed to be dominant in unfolded proteins[25], short polypeptides and regions of proteins previously assumed unstructured[26]. Raman spectroscopic studies also suggest that the PPII helix is an important transitory element in protein folding and denaturation, having been observed during the melting of α-helices in peptides[27] as well as the transformation of an α-helix in native human lysozyme into a β-sheet strand involved in amyloidogenesis[28].

The PPII helix is an elongated, left handed helix encompassing 3 residues per turn with a helical pitch of 3.1 Å per residue. In its ideal form, it is defined by the backbone dihedral angles $\varphi = -75°$, $\psi = 145°$ and $\omega = 180°$ (all-trans conformation). As the name suggests, proline residues have a high propensity to form PPII helices. This is a consequence of the pyrrolidine ring restricting φ to a region suitable for PPII helix formation ($-75 \pm 15°$), thus leading to a decreased loss of conformational entropy upon assuming the ordered structure. It should be noted though that the above mentioned angles correspond to Ramachandran regions which are populated by all amino acids, meaning that proline residues are not an absolute requirement for PPII helix formation.

The PPII helix lacks the backbone hydrogen bonds characteristic of the other secondary structure elements, and the factors governing its stability have long been disputed. One important aspect are steric interactions and restrictions: the region of φ and ψ populated by amino acids in a PPII helix characteristically lacks unfavourable steric interactions[29], as would be intuitive considering the elongated, well-spaced structure of the helix with all side chains pointing away from the helical axis. Furthermore, steric effects alone have been sufficient to drive PPII helix formation in Monte Carlo simulations, with the important interactions having been pinpointed to occur between the pyrrolidine moiety of proline residues and the backbone of the preceding amino acid[30]. Another factor contributing to the stability is the high degree of solvation, especially of the backbone carbonyl and amide groups,

which are aligned perpendicular to the helical axis and highly exposed. This is underlined, inter alia, by the correlation observed between solvent-accessible surface area and PPII formation propensity[31]. The side chains of the amino acids are also observed to influence the stability of the helix: the strong tendency of Gln residues to form PPII helices, for instance, has been explained with intermolecular hydrogen bonds formed between the amide hydrogen of the side-chain and the backbone carbonyl of the C-terminally neighboring amino acid.[32] Another stabilizing effect yet comes in the form of a $n \rightarrow \pi^*$ interaction, in which overlap between the non-bonding orbital of a carbonyl-oxygen with the anti-bonding π^* orbital of the carbonyl moiety of a C-terminally neighbouring residue leads to a stabilization of an estimated 0.7 kcal/mol via electron delocalization[33,34].

The Collagen Triple-helix

The molecular structure of this defining element of collagens has been extensively characterized in the past - the classical collagen model peptide $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$ has even made it aboard the space shuttle Discovery, where crystals were grown under micro-gravity[35], which back on earth resulted in structure determination with a respectable 1.3Å resolution and an R-factor of 0.18[36]. The collagen triple-helix is comprised of three parallel PPII helices, supercoiled along a common axis to form a right handed triple helix with a helical pitch ranging from 7/2 for proline-rich regions to 10/3 for proline-poor regions[17]. The α-chains are axially staggered by one amino acid and thus can be distinguished as leading (+0), middle (+1) and trailing (+2) strands in order of decreasing C-terminal overhang, as first defined by Emsley *et al*[37]. The triple-helix is stabilized by periodic hydrogen bonds between the amide-nitrogen of glycine residues and the amide-oxygen of residues in position Xaa on the strand with -1 stagger. The estimated contribution of each hydrogen bond to the Gibbs free energy of trimerization is between -1.4 to -1.8 kcal[17].

The observed tight packing of the triple-helix explains the necessity for the staggered assembly of α-chains as well as the strict requirement for glycine residues in every third position: the side-chains of residues in these positions are oriented such that any amino acid other than glycine would experience steric clashes with the neighbouring strand with +1 stagger.

Structural Aspects of Triple-helix Stability

The collagen triple helix is not stable at body-temperature[38], which is rather surprising considering its significance as a structural element in the extracellular matrix. This meta-stability underlines that collagens, although often being regarded merely as rigid scaffolding proteins, are remarkably dynamic and able to utilize their structural "imperfections" for biological activity. Transient structural perturbations in the collagen triple-helix are indeed essential for many processes such as recognition of collagens by matrix metalloproteases (MMPs) or the binding of heparin to the collagenous domain of acetylcholinesterase, determining its anchoring location[39].

Although the interaction interface of triple-helix forming PPII helices is dominated by the polypeptide backbone, triple-helix stability is greatly influenced by α-chain sequence. As is often the case with polymeric molecules, the loss of conformational entropy upon adopting an ordered structure is significant for collagen α-chains. This underlies the fact that the stability is intrinsically governed by the propensity of α-chains to pre-organize and adopt a PPII helical conformation, and further of the PPII helices to intertwine to form a triple-helix.

The strict requirement for glycine residues at every third position, as explained above based on the crystal structure, is also showcased by the fact that many diseases, such osteogenesis imperfecta (OI) or epidermolysis bullosa, are linked to substitutions of these key residues in collagen α-chains[40,41]. While the disruptive impact of such mutations on triple-helix stability is also influenced by the identity of adjacent amino acids and is for instance attenuated in proline-rich regions[42], they can also have more indirect effects: during collagen biosynthesis, re-nucleation of the triple helix beyond interruptions induced by glycine substitutions occurs after a certain delay, during which the collagen α-chains are overmodified. Such overmodifications alter triple-helix as well as fibril stability and can be detected in patients suffering from OI induced by such glycine substitutions[43]. This is further substantiated by the observation that glycine mutations closer to the C-terminus are causative to more severe forms of OI: triple-helix formation proceeds from the C- towards the N-terminus (see 1.1.4), and thus a larger proportion of the molecule is exposed to overmodification during the presumably identical delay[44].

The influence of amino acids in positions Xaa and Yaa on triple-helix stability has been thoroughly characterized in the past. Fortunately, natural collagens only contain relatively few of the theoretically possible different triplets (more than 400), which somewhat limited the effort to map the triple-helix formation propensities of amino acids. The effect of amino acid substitutions on thermal stability is usually assessed using host-guest peptides. These consist of a variable region (guest), which is flanked N- and C- terminally by trimeric regions of $(GPP)_n$ or $(GPO)_n$ triplets (host). The T_m of peptides is commonly determined using circular dichroism spectropolarimetry, a valuable tool for distinguishing between and quantifying monomeric and triple-helical content. Persikov *et al.*[45] have studied substitutions at either the Xaa or the Yaa site and made the following observations:

(i) Charged residues generally show the least destabilizing effect at either position. This might indicate the presence of polar interactions of the sidechains with neighbouring α-chains. Arginine, for instance, is capable of forming a hydrogen bond with the backbone-carbonyl of an adjacent chain, stabilizing the triple helix[46].

(ii) Glycine and hydrophobic residues show the largest destabilizing effect at either position. This is most probably due to the conformational restrictions imposed by the PPII helix geometry in case of the former, which enjoys access to a large area in φ, ψ-

space; the hydrophobic residues, on the other hand, are possibly too bulky and block solvent access to the backbone of neighbouring strands.

(iii) The enthalpic contributions (ΔH°) of Pro (in position Xaa) as well as Hyp (in position Yaa) to triple-helix formation were noted to be among the lowest of all amino acids, supporting the idea that entropic aspects and preorganization are critical to triple-helix stability. This was further corroborated by the medium correlation observed between the propensities of PPII helix and triple-helix formation.

(iv) Disregarding GPP and GPO, the most stable triplets were GEO and GPR for substitutions at position Xaa and Yaa, respectively. Complementary studies where both positions were substituted simultaneously have identified GER as the most stable triplet lacking Pro and Hyp, which concurs with the above findings. It was furthermore observed that triplet stability correlates positively with its occurrence in natural collagen.[47]

Proline hydroxylation is known to exert profound influence on the stability of collagens, and the incidence of proline hydroxylation in different organisms clearly correlates with the average environmental temperature these experience[48]. The influence of hydroxylation on thermal stability is dependent on location and stereochemistry: the vast majority of hydroxyprolines are typically found in the Yaa position and with (2S,4R) configuration, with exceptions being very rare (collagen type IV incorporates (2S,3R)-hydroxyproline in both positions)[49,50]. Due to the supercoiling of the PPII helices, residues in positions Xaa and Yaa become distinguishable in terms of solvent exposed area as well as preferred main-chain dihedral angles. Experimentally determined[36] φ, ψ and ω values for proline residues in Xaa or Yaa position of a triple helical α-chain are contrasted with those of an ideal PPII helix in table 1.1. These differences, especially in φ, lead to a position dependent discrimination against different pyrrazolidine pucker conformations of proline and derivative residues. The dihedral angles of the Xaa position favour the Cγ-*endo* (also called "down") pucker, which is slightly preferred in prolines, while in the Yaa position the Cγ-*exo* ("up") pucker is preferred, which is the predominant form in hydroxyprolines due to the *gauche* effect of the electron withdrawing hydroxyl group (Fig 1.2)[51,52]. The *gauche* effect describes the tendency of molecules to adopt a conformation with a dihedral angle of ± 60° between vicinal polar bonds, which in case of hydroxyproline allows for a larger overlap between

Table 1.1 | Comparison of backbone dihedral angles between the PPII helix and the collagen triple-helix.

	φ / °	ψ / °	ω / °
Xaa (Triple-helix)	-74.5	164.3	176.0
Yaa (Triple-helix)	-60.1	152.4	175.4
Xaa (PPII-helix)	-75	145	180
Yaa (PPII-helix)	-75	145	180

the Cδ-H σ and Cγ-O σ* orbitals and thus leads to increased stabilization via hyperconjugation. Studies with a range of proline derivatives have supported the view that hydroxylation influences stability primarily via this stereoelectronic effect, and that the participation of the hydroxyl-group in H_2O networks, long believed to be the major contributor to stability, is secondary to this[53,54,55].

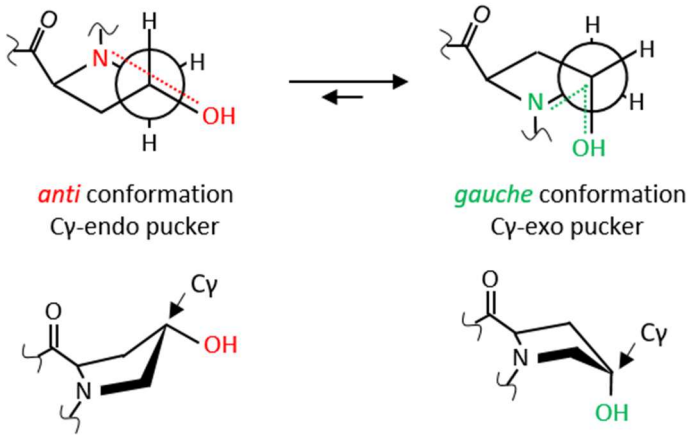


Figure 1.2 | Gauche effect and preferred pucker conformation in 4S-hydroxyproline. The Newman projection (proximal atom: Cγ, distal atom: Cδ) of hydroxyproline conformers is shown in the top. The *gauche* effect, in this case attributable to hyperconjugation between the Cδ-H σ and Cγ-O σ* orbitals, leads to a preference of the Cγ-*exo* pucker over the Cγ-*endo* pucker, despite increased steric clash of the 4-hydroxyl group with the vicinal amide group. The Cγ-exo pucker corresponding to this conformation is thus favored in 4S-hydroxyproline.

One requirement for the formation of the collagen triple-helix is an all-*trans* configuration of α-chain peptide bonds. Proline residues, though, induce a high population of *cis*-configuration in peptide bonds with N-terminally preceding amino acids. This is primarily due to the steric clash between Cα of the N-terminally adjacent amino acid and Cδ of proline in *trans*-configuration, increasing its energy state to become similar to that of the *cis*-configuration. The incorporation of hydroxyproline, however, affects both the thermodynamics and kinetics of the *cis-trans* isomerisation: it shifts the ratio towards the *trans*-configuration due to the large $n \rightarrow \pi^*$ overlap seen in the Cγ-*exo* pucker[56] and increases the isomerization rate via weakening of the neighboring amide resonance[57].

1.1.4 Collagen Biosynthesis

Like all classical secretory proteins, procollagen α-chains are co-translationally translocated into the lumen of the rough ER. The nascent polypeptide chains are immediately subjected to a range of post-translational modifications and interactions with molecular chaperones, which orchestrate the proper processing, folding and trafficking of the maturing collagens. As is evident from the deleterious impact of α-chain overmodification on the stability and downstream processing of collagens, the regulation of the trimerization of unfolded chains is crucial, since the tightly packed triple-helix renders the α-chains inert towards further enzymatic modification. One consistent observation with collagen modifying enzymes is that they often have a secondary chaperone function. This is a sensible solution considering the sheer abundancy and size of the intrinsically meta-stable and aggregation-prone collagens. Many collagen modifying enzymes are not exclusive to collagens; furthermore, they often act as part of heterocomplexes with more than one function. While the complex P3H1/CRTAP/CypB, for instance, is responsible for (3S) hydroxylation of prolines (*vide infra*), defects in any member of the complex not only abolish hydroxylation but also lead to a general overmodification of collagens, suggesting that it fulfils a range of chaperoning roles that are not yet fully understood[58]. One proposed function of this complex is to transiently stabilize junctions between triple-helical and unfolded regions until other chaperones such as HSP47 can take over[59]. Lysine hydroxylases similarly act as member of a multifunctional complex together with FK506-binding protein 65 (FKBP65), 78kDa Glucose-regulated protein (GRP78, or BiP) and HSP47 (*vide infra*)[60], suggesting a complex interplay between the proteins.

Hydroxylation

The three isoforms of proline-4-hydroxylase (P4H) catalyse the oxidation of proline to (2S,4R)-hydroxyproline. The enzyme acts as part of a heterotetramer, in which two molecules P4H associate with two molecules of protein disulphide isomerase (PDI). The latter is not involved in the hydroxylation, but is required for the solubility and ER retention of P4H[61]. The catalysed reaction is stereospecific and occurs at the Yaa position of Gly-Xaa-Yaa triplets; it requires Fe^{2+} as cofactor and O_2 as well as α-ketoglutarate as cosubstrates. The latter is oxidatively decarboxylated to yield succinate and CO_2. Ascorbate, while not involved in proline hydroxylation per se, is required as a reducing agent for the regeneration of Fe^{2+}. P4H also seems to act as a chaperone, since it can still associate with hydroxylated α-chains, albeit at a lower affinity than with its natural substrate[62].

Proline-3-hydroxylases catalyse the oxidation of proline residues in Gly-Pro-Hyp sequences to (2S,3S)-hydroxyproline. P3H forms a heterotrimer with cartilage-associated protein (CRTAP), which shares homology with P3H but lacks the monooxygenase domain, and the peptidyl-prolyl cis-trans isomerase cyclophilin B (CypB)[59].

Lysine hydroxylases are luminally oriented, peripheral membrane proteins which catalyse the hydroxylation of lysine residues in Gly-Xaa-Lys sequences. 5-hydroxylysine is a pre-requisite for O-

glycosylation and fiber-crosslinking during later stages of collagen maturation[63]. The active enzyme, which requires the same cofactor and cosubstrates as P4H and P3H, is a homodimer which forms with the aid of the peptidyl-prolyl *cis-trans* isomerase FKBP65[64]. Latest research has suggested that GRP78 and HSP47 are also involved in the chaperone complex, with the former, an ER homologue of HSP70, acting as a scaffold for the complex and the latter acting as a negative regulator for lysine hydroxylation[60,65]. Lysine hydroxylation occurs both throughout the collagenous domain (catalysed by LH1) as well as telopeptide domains (LH2), and its extent varies strongly between different collagen types and tissues[63].

Glycosylation

A number of hydroxylysine residues are further modified via O-glycosylation. This involves the glycosyltransferases GLT25D1 and GLT25D2, which catalyse the attachment of β-galactose, and LH3, which (primarily) catalyses the attachment of α-glucose[58]. The biological significance of these O-linked sugars has not been fully elucidated yet; one study indicates that they are, among other roles, involved in recognition by the endocytic collagen receptor uPARAP/Endo180[66].

Collagens are also subjected to N-glycosylation with mannose-rich oligosachharides in their telopeptide domains[67]. The consensus sequence for recognition by ER-resident glycosyltransferases in collagens, Asn-Ile-Thr, is highly conserved within fibrillar collagens; collagens lacking N-glycosylation, however, surprisingly do not show any abnormalities in assembly, secretion or deposition[68]. The discovery that endocytosis of cleaved collagen type I C-propeptides occurs via the mannose-receptor[69] renders it likely that N-glycolysation is utilized in recognition and clearance of cleaved proopeptides.

Cis-trans isomerization

The isomerization of *cis* peptide bonds to *trans*, which is the rate limiting step in triple-helix formation[70], is catalysed by the peptidyl-prolyl *cis-trans* isomerases (PPI) CypB, FKBP22 and FKBP65[58]. CypB is believed to be the major catalyst in triple-helix formation; in addition to the above mentioned complex with P3H and CRTAP it also interacts with other collagen modifying enzymes and chaperones such as PDI, Calnexin/Calreticulin, LH1 and HSP47[58]. Data on the role of FKBP22 is scarce, but it has been reported to be involved in the processing of collagens type III, VI and X[71]. In addition to its *cis-trans* isomerase activity, FKBP65 also displays properties of a chaperone, since it can bind unfolded as well as triple-helical collagen[72]; furthermore, it is a positive regulator for lysyl hydroxylation[73,74].

Folding of collagens

Chain selection, trimerization and proper registration of fibrillar collagens is steered by their C-terminal propeptides. The folding of these propeptides is aided by rER-residing general chaperones such as GRP78, PDI, Calnexin/Calreticulin and CypB[58]; furthermore, these domains undergo enzymatic modifications such as LH2 catalysed lysine hydroxylation and PDI catalysed formation of intra- and

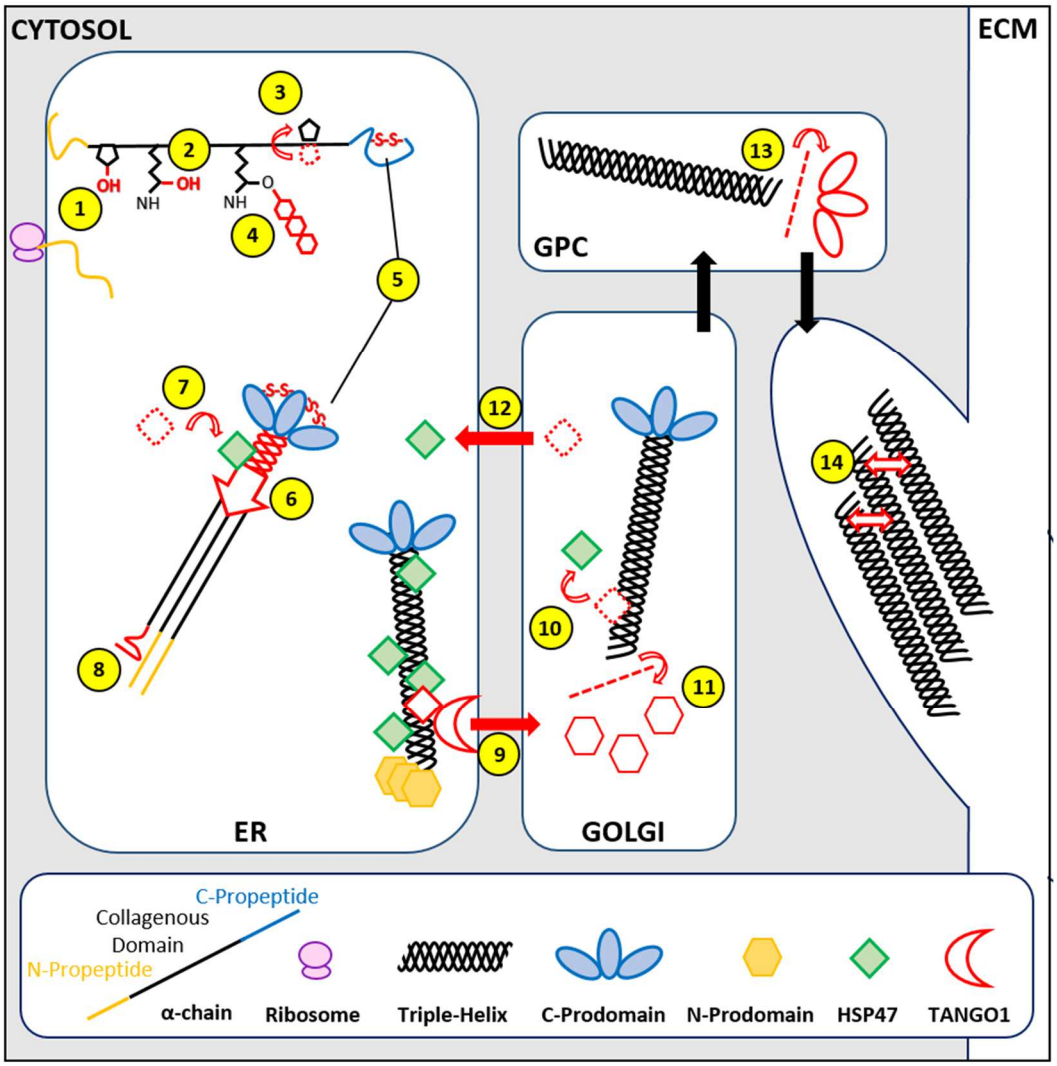
interchain disulphide bonds. Although this covalent cross-linking stabilizes tertiary structure of the subunits and prevents dissociation of the trimer, it has been shown that collagen type IV mutants lacking critical cysteine residues are still capable of producing mature collagen[75]. It thus seems plausible that the disulphide bridges rather serve to impose structure than stabilize it. Besides its thiol-shuffling function, PDI also serves as a chaperone and associates to nascent collagen α-chains, preventing aggregation and improper trimerization[62]. The C-propeptides are believed to be closely associated to the rER membrane, facilitating trimerization due to the higher probability of a trimolecular binding event in two-dimensional space[76]. Upon assembly of three C-terminal propeptides, triple-helix formation is initiated by prolyl-4-hydroxylation in at least two Gly-Xaa-Pro triplet repeats localized at the C-terminal end of the collagenous domain[77], and proceeds towards the N-terminus in what is often called a zipper-like fashion[78].

It should be mentioned that this folding process is not uniform for all collagens: in the ectodomains of membrane collagens type XIII and XVII, for example, trimerization proceeds in the opposite direction[7]; furthermore, in FACIT collagen type IX, the NC2 domain was identified as being responsible for chain selection and trimerization of neighboring collagenous domains[79].

Trafficking, processing and secretion

Collagens are very large proteins with a length often exceeding 300 nm, while vesicles budding from the ER membrane typically have a diameter of 60 - 80 nm. The transport of procollagen from the rER to the Golgi apparatus thus requires special vesicles shaped to accommodate rigid, rod-like cargo of such size. These large vesicles share the coat-protein Complex 2 (COPII) coat associated with ER anterograde transport vesicles, but the principles governing their formation and cargo selection are not fully understood. The ubiquitin ligase CUL3-KLHL12 has been reported promote assembly of large vesicles via monoubiquitinylation of SEC31, a component of the outer COPII coat[80]. Loading of the cargo vesicle with collagen is mediated by transmembrane protein transport and Golgi organization protein 1 (TANGO1), an ER exit-site residing protein operating in complex with cutaneous T-cell lymphoma-associated antigen 5 (cTAGE5). TANGO1 interacts with the COPII inner coat components Sec23A and Sec24C, and furthermore with the guanine nucleotide-exchange factor Sec12, thus possibly modulating the Sar1 GTPase cycle which initiates vesicle formation[81,82]. In this way TANGO1 resembles Sedlin, another protein associated with secretion of large proteins, which also acts via regulation of nucleotide exchange in Sar1. Recent experiments have revealed that the SH3-domain of TANGO1 barely recognizes collagens on its own and that the binding is mediated by HSP47[83]. The importance of TANGO1 for collagen shuttling is corroborated by experiments in which knockout of TANGO1 in chondrocytes, fibroblasts, endothelial and mural cells was observed to hamper secretion of collagen types I, II, III, IV, VII and IX[84].

In the early stages of the secretory pathway (ERGIC and Golgi apparatus), gradually decreasing pH induces full dissociation of HSP47 from procollagen, while members of the "a disintegrin and metalloproteinase with thrombospondin motifs" (ADAMTS) family of proteases cleave off the N-terminal prodomains. The truncated procollagen is then transported to the extracellular space via large Golgi-to-plasma membrane carriers (GPCs) which originate from the *trans*-Golgi[85]. During this transit, the C-terminal prodomains are removed by members of the tolloid family of metalloproteinases such as bone morphogenic protein 1 (BMP1) or tolloid-like 1 (TLL1) to yield tropocollagen consisting only of a collagenous domain[86]. This results in a marked decrease in solubility (up to five orders of magnitude) and promotes lateral aggregation and assembly into fibrils[87]. While it is understood that fibrillogenesis is initiated by the removal of the C-propeptides, there is uncertainty concerning the exact location of it. In the Kadler model of fibrillogenesis, tropocollagen generation and its subsequent lateral association takes place in GPCs, while in the Birk model these steps occur extracellularly in a cavernous invagination of the plasma membrane[87]. Fibrillogenesis is nucleated by collagens V and XI[88], regulated by small leucine-rich proteoglycans (SLRPs) such as decorin, biglycan, lumican or fibromodulin[87] and aided by adaptor proteins such as the TSPN family member cartilage oligomeric matrix protein (COMP). Assembly of tropocollagen typically occurs in a staggered, head-to-tail fashion, the former resulting in the characteristic banded pattern seen in collagen fibrils[89]. The fibrils are stabilized by intra- as well as inter-molecular crosslinking between allysine and/or hydroxyallysine residues, generated by lysyl oxidases (LOXs) via oxidative deamination of specific lysine and hydroxylysine residues, respectively. Lateral and longitudinal growth leads to formation of mature collagen fibers, which in turn also form increasingly complex structures such as parallel bundles or basket waves[90].



Step	Description	Associated Proteins / Complexes
1	Proline hydroxylation	[P3H1/CRTAP/CypB], [P4H/PDI]
2	Lysine hydroxylation	[LH2/FKBP65/GRP74/HSP47]
3	Peptidyl-prolyl *cis-trans* isomerisation	CypB, FKBP22, FKBP65
4	Glycosylation	GLT25D1, GLT25D2, LH3
5	Folding of C-Propeptide and disulphide bond formation	PDI, GRP78, Calnexin, Calreticulin, CypB
6	Triple-helix formation and progression	-
7	HSP47 binding to triple-helix	-
8	Folding of N-Propeptide	PDI, GRP78, Calnexin, Calreticulin, CypB
9	Packaging and anterograde transport to Golgi body	[TANGO1/cTAGE/Sec12/Sec23A/Sec24C], HSP47
10	HSP47 dissociation	-
11	N-Propeptide cleavage	ADAMTS
12	Retrograde transport of HSP47	KDEL-receptors
13	C-Propeptide cleavage	BMP1, TLL1
14	Fibril formation and crosslinking	SLRPs, Comp, LOXs

Figure 1.3 | Schematic overview of collagen biosynthesis. The assembly of a prototypical fibrillar collagen molecule is depicted up to fibril formation. Post-translational processing is grouped into 14 chronologically arranged steps which are indicated by yellow circles. An overview of these steps and the involved enzymes and chaperones is provided in the legend.

1.2 Heat Shock Protein 47

1.2.1 An Introduction to HSP47

Heat-shock protein 47 (HSP47, also known as Colligin or CBP1) is an ER-residing molecular chaperone exclusive to procollagen*. The glycoprotein is a member of the Serine-Protease Inhibitor (Serpin) family (Clade H, member 1), albeit a rather superficial one, since unlike the majority of its paralogues it is not secreted, lacks inhibitory activity against proteases and does not undergo Serpin-typical conformational transitions. Untypical for chaperones, HSP47 is substrate-specific and recognizes the already folded conformation of its substrate, i.e. triple-helical collagen; client binding and release is furthermore not coupled to nucleotide exchange or hydrolysis. HSP47 is also unique in that it is the only heat-shock inducible protein in the ER, all other chaperones operating in context of other forms of ER-stress stimuli.

HSP47 is an essential component of the finely tuned collagen biosynthesis machinery. Knockdown of HSP47 is lethal: HSP47$^{-/-}$ mice display grievous phenotypes such as ruptured blood vessels, defective basement membranes and aberrancies in the orientation of epithelial tissues. These deficiencies result in severe growth retardation and are lethal 11.5 days post-coitus[91]. The loss of the chaperone precipitates as various abnormalities in collagen biosynthesis, ranging from the accumulation and aggregation of overmodified and improperly processed collagens (mainly of type I) in the ER to the complete absence of some collagens (type IV), while the integrity of collagen type III was intriguingly found to be intact although it is a known substrate. The importance of HSP47 is also manifested by the implication of the protein in various collagen-related diseases: destabilizing missense mutations (Leu78Pro in humans, Leu326Pro in dachshund), for instance, result in severe forms of osteogenesis imperfecta[92,93]. Upregulation of HSP47, on the other hand, is commonly linked with diseases characterized by increased collagen deposition, such as arteriosclerosis, myocardial infarction or fibrosis[94,95].

HSP47 can be found in ancient vertebrate classes such as Sarcopterygii and Actinopterygii, and given its central role in collagen biosynthesis, it is presumably common to all vertebrates. Constitutive expression of HSP47 is only observed in collagen producing cells such as fibroblasts, myoblasts or adipocytes. Expression levels are tightly correlated to that of collagens, with up- or down-regulation of the latter being mirrored by HSP47[96], and curiously respond to changes in gravity: while the expression is stimulated by hypergravity, it is attenuated under microgravity conditions[97].

*The client-repertoire of HSP47 is not fully explored; currently, direct evidence for HSP47 binding is available in form of SPR experiments for collagens I through V[98]. Unpublished data (PhD Thesis Anna Köhler, research group Prof..Zaucke) is available which indicates HSP47 also recognizes other non-fibrillar collagens besides collagen IV.

1.2.2 Serpins and Structure of HSP47

The superfamily of Serpins is perhaps better defined by structure than function, since contrary to what their name suggests not all Serpins are inhibitors of serine proteases. Some Serpins are able to inactivate Cys-proteases (SerpinB9 is an inhibitor of Caspase-1[99]) or show cross-class inhibition (Endopin 2A[100]), and quite a number are not inhibitory at all, as is the case with the collagen chaperone HSP47 or the storage protein Ovalbumin. Serpins have been found in all branches of life, partaking in diverse functions ranging from host defence in insects (Serpin27A)[101] to blood coagulation[102]. Although the size of serpins varies considerably (from 292 amino acids in case of the viral Serpin Cytokine Response Modifier 1 (CRM1) to 478 amino acids in case of C1-Inhibitor), and the pairwise identity of their amino acid sequence can be as low as 25%[103], they all share a structurally conserved core domain consisting of seven to nine α-helices, three β-sheets and a flexible, solvent exposed loop called the reactive center loop (RCL). These typical structure formed by these elements constitutes the so-called Serpin-fold.

The Serpin-typical structural elements are highlighted in figure 1.4 using the structure of HSP47. The largest of the β-sheets (βA) consists of five strands aligned with the long axis of the protein. All structural elements lie to one side of the plane formed by βA ("inner side"), the sole exception being hF, which in HSP47 is tethered to the "outer side" of βA via a salt-bridge between Asp182, located at its C-terminal end, and Lys129, located at the N-terminal end of s2A. In inhibitory Serpins, this salt bridge is absent since the inhibitory mechanism requires transient dislocation of hF to make room for the baited protease (*vide infra*)[104]. The longest two strands of β-sheet A, s3A and s5A, are noteworthy in that they are aligned in parallel. This creates a natural weak point in the β-sheet which can be mitigated via insertion of a new, anti-parallel strand. This feature lies at the heart of Serpin-typical conformational transitions involving the insertion of the RCL (or a part of it) as s4A.

β-sheets B and C form the bulk of the C-terminal part of Serpins. The latter comprises the greatest part of the collagen-binding interface of HSP47 (see 1.2.4) and is also utilized by other Serpins for interaction with effector molecules; one example would be the binding of Protein Z by Protein Z-dependent Inhibitor, a Serpin involved in the coagulation cascade[105]. In the majority of ligand-dependent Serpins, binding of glycosaminoglycan cofactors occurs near or at α-helix D, inducing large conformational re-arrangements relayed by the cluster of α-helices surrounding hD. Protein C inhibitor (PCI), as one of the exceptions, binds heparin and heparan sulfate using α-helix H[106].

The RCL is a flexible loop protruding from the protein core in the native Serpin fold. It is tethered to s5A via its N-terminus and to s1C via its C-terminus. The key feature of the RCL (at least in an overwhelming majority of Serpins) is its capacity to insert itself in between the parallel β-strands s3A and s5A as an additional, anti-parallel strand s4A. There are many variants of RCL insertion: partial or full, intramolecular or intermolecular, spontaneous insertion of the intact loop or insertion triggered via

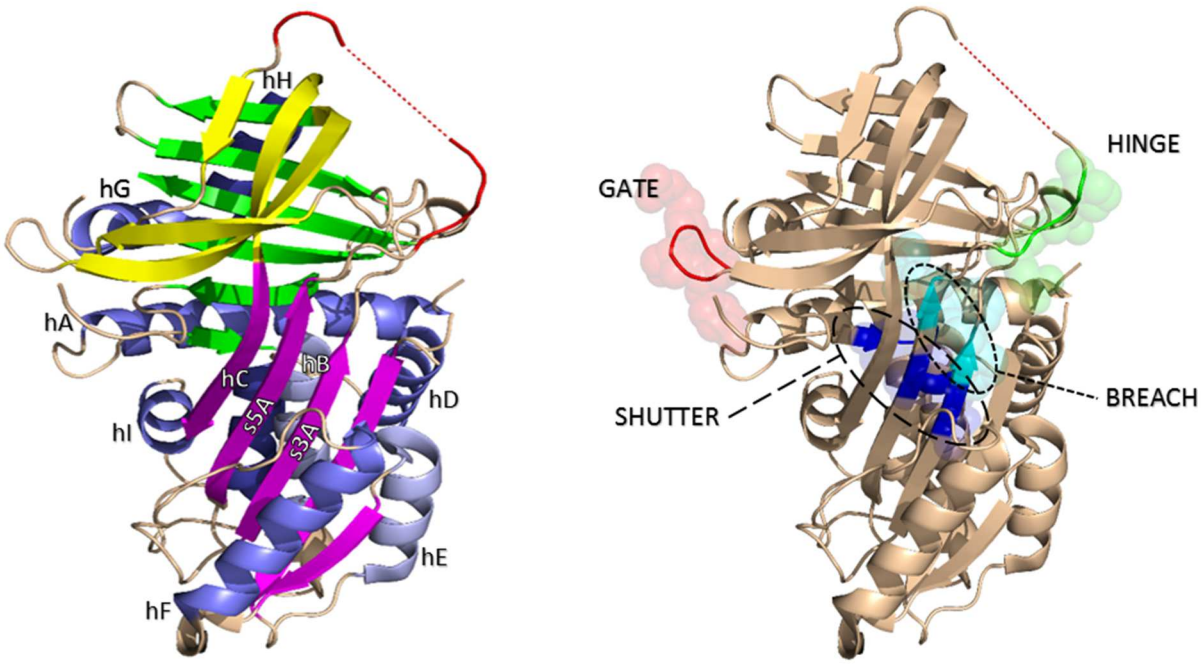


Figure 1.4 | Structural elements of HSP47. ***(Left)*** HSP47 displays a typical Serpin fold consisting of 9 α-helices (hA - hI, blue), 3 β-sheets (βA: magenta, βB: green, βC: yellow) and the reactive center loop (red, missing structural elements are represented by the dotted line). ***(Right)*** Amino acid clusters regulating insertion of the reactive center loop into βA in Serpins. (PDB ID: 4AU3)

proteolytic cleavage within the RCL. The conformer resulting from RCL insertion is thermodynamically much more stable, as evidenced by the large increase in its thermal stability. This has the interesting implication that the native Serpin fold, called the stressed state (S), is intrinsically metastable and capable of spontaneous transition to the relaxed state (R). While the additional anti-parallel β-strand is certainly the most important factor in driving the S → R transition, the native state also displays several non-ideal interactions such as side-chain overpacking, burial of polar groups in hydrophobic pockets, or cavities in the latter[107]. Four clusters of amino acids exert influence over the dynamics of RCL insertion: the breach cluster, consisting of the C-termini of s3A and s5A, as well as the shutter region, consisting of the middle part of βA together with s1B and the N-terminal part of hB, regulate the propensity of s3A and s5A to move apart and allow insertion of the RCL. The hinge region of the RCL is critical for the flexibility of the loop, and in some cases capable of partial insertion into βA. Lastly, the gate region, consisting of a short loop between s4C and s5C, regulates the S → R transition in uncleaved Serpins, since the RCL has to swing its C-terminus around this cluster for full insertion. This also requires dissolution of s1C.

Proteolytic cleavage within the RCL frees its C-terminus, greatly increasing the probability and speed of insertion. This is exploited by inhibitory serpins, which present a bait site close to the C-terminal end of the RCL; upon proteolytic cleavage, the insertion of the freed RCL proceeds fast enough to compete with the final step of proteolysis, the nucleophilic attack of H_2O on the acyl-enzyme intermediate. The

protease is dragged along the outer side of βA and translocated to the opposite end of the serpin. This occurs before the protease is able to release its substrate via hydrolysis; in the resulting complex, distortion of the active site renders it fully inoperable[108].

1.2.3 HSP47 in Collagen Biosynthesis

After initial trimerization of procollagen C-prodomains, aided and stabilized by disulphide bond formation, proline hydroxylation initiates formation of the triple helix. As the folding progresses towards the N-terminus, multiple molecules of HSP47 are recruited to the triple helical regions, showing a slight preference towards more N-terminal sequences[109]. They accompany their client during TANGO1-mediated loading into elongated, COPII-coated vesicles and into the Golgi-body. Experiments in which various stages of Golgi-transport were inhibited have indicated that client release occurs either in the ER - Golgi Intermediary Compartment (ERGIC) or in the *cis*-Golgi[110]. Complex dissociation is presumed to be induced by the mildly acidic environment (pH in Golgi body = 6.2[111]) in these compartments. This presumption is further substantiated by the behaviour of HSP47 in gelatin-affinity chromatography, where elution is observed to occur at a pH of 6.3[112]. After dissociation, retrograde transport of HSP47 to the ER via KDEL receptors occurs[113], while procollagen further continues down the secretory pathway.

The principal functions of HSP47 in collagen biosynthesis are:

(i) Stabilization of the triple helix (i.e. prevention of unwinding). Although it has previously been reported that HSP47 binding did not alter the thermal stability of collagens type I and III[115], it should be noted that the experiment did not account for the temperature dependency of the HSP47 - collagen interaction; the fact that the HSP47 - collagen interface encompasses all three strands strongly suggests that the triple-helical conformation is indeed stabilized.

(ii) Regulation of post-translational modification, which is accomplished on one hand via prevention of local unwinding of the triple-helix, thus denying collagen modifying enzymes access to the single stranded substrate (indirect regulation), and on the other hand via direct binding and modulation of such enzymes. The negative regulatory effect of HSP47 on proline-4 hydroxylation is an example of indirect regulation[114], while a case of regulation via direct interaction has been described above for lysine hydroxylases.

(iii) Marking the collagen as ready for ER-export and serving as a universal collagen-adaptor molecule for TANGO1.

Another role often attributed to HSP47 is the obstruction of premature lateral aggregation of procollagen. While this property of HSP47 is evident from the concentration and pH-dependent inhibition of collagen fibril formation observed *in-vitro*[116], it is somewhat at odds with the fact that a

prerequisite for fibril formation is the removal of the C-propeptides, a process which occurs in post-Golgi compartments while HSP47 already is fully dissociated in the *cis*-Golgi.

1.2.4 Molecular details of the HSP47 - Collagen Interaction

Although it was earlier reported that HSP47 is capable of binding both single-chain and triple-helical procollagen[110], latest evidence indicates that only the latter is recognized[117]. An arginine in the Yaa position of any single α-chain constitutes the primary binding motif for HSP47[118], and the interaction can be further compounded by an additional threonine in position Yaa^{-1}* on the same strand[119]. In addition to these primary binding sites, HSP47 was also shown to interact with CMPs consisting purely of (Gly-Pro-Pro) triplets, albeit very weakly[120]. In the same study it was also noted that successive incorporation of (2S-4R)-Hydroxyproline in position Yaa gradually decreases the strength of the interaction. HSP47 recognition of collagens requires the latter to consist of a minimum of five triplets[118]. Interestingly, crosslinking experiments indicate the presence of an alternative binding mode in which the long axis of HSP47 is oriented parallel to the helical axis of collagens[121]. It is not clear whether this mode of interaction has any biological significance; however, it can be hypothesized that this presumably transient association is utilized by HSP47 to scan the collagen triple-helix for its preferred binding motif Gly-Pro-Arg, or to migrate in direction of the N-terminus as the triple-helix extends.

Widmer *et al.* have previously succeeded in co-crystallizing *Canis lupus* HSP47 with various CMPs and solving the structure via X-Ray diffractometry (Fig. 1.5)[122]. The crystal structure was in satisfactory agreement with previous findings and has shed light on the molecular details of the HSP47 - collagen interaction. The binding interface spans roughly five triplets on the collagen side, which allows for a sufficient twist of the triple-helix for HSP47 to recognize all three α-chains. The interaction has predominantly hydrophobic character; polar interactions are limited to a salt bridge between Asp385 in HSP47 and Arg (arbitrarily assigned position Yaa^{0}) of the CMP, accounting for the binding motif of HSP47, and hydrogen bonding between Arg222 of HSP47 and the backbone carbonyl of Proline in position Yaa^{-1} (Fig. 1.6). The stabilizing contribution of Thr in position Yaa^{-1} stems from the formation of a small H-bond network involving Ala303, Ile304, Ser305 and a water molecule (Fig. 1.6). The largest part of the interaction interface is provided by β-sheet C. While the RCL is positioned such that it could possibly contribute to binding of the CMP, crystal structures reveal that it is highly mobile and does not adopt any fixed position or conformation for interaction with the CMP. The crystal structure shows two molecules of HSP47 bound to one molecule of CMP, each recognizing an arginine residue in the leading and trailing strand, respectively. The long axis of one HSP47 molecule is rotated close to 170° around the CMP helical axis in relation to the other, resulting in a near head-to-head orientation. There are no evident interactions between the HSP47 molecules; it can be assumed that the previously

*Yaa^{-1} indicates the residue in the triplet preceding the arginine, which is defined as Yaa^{0}.

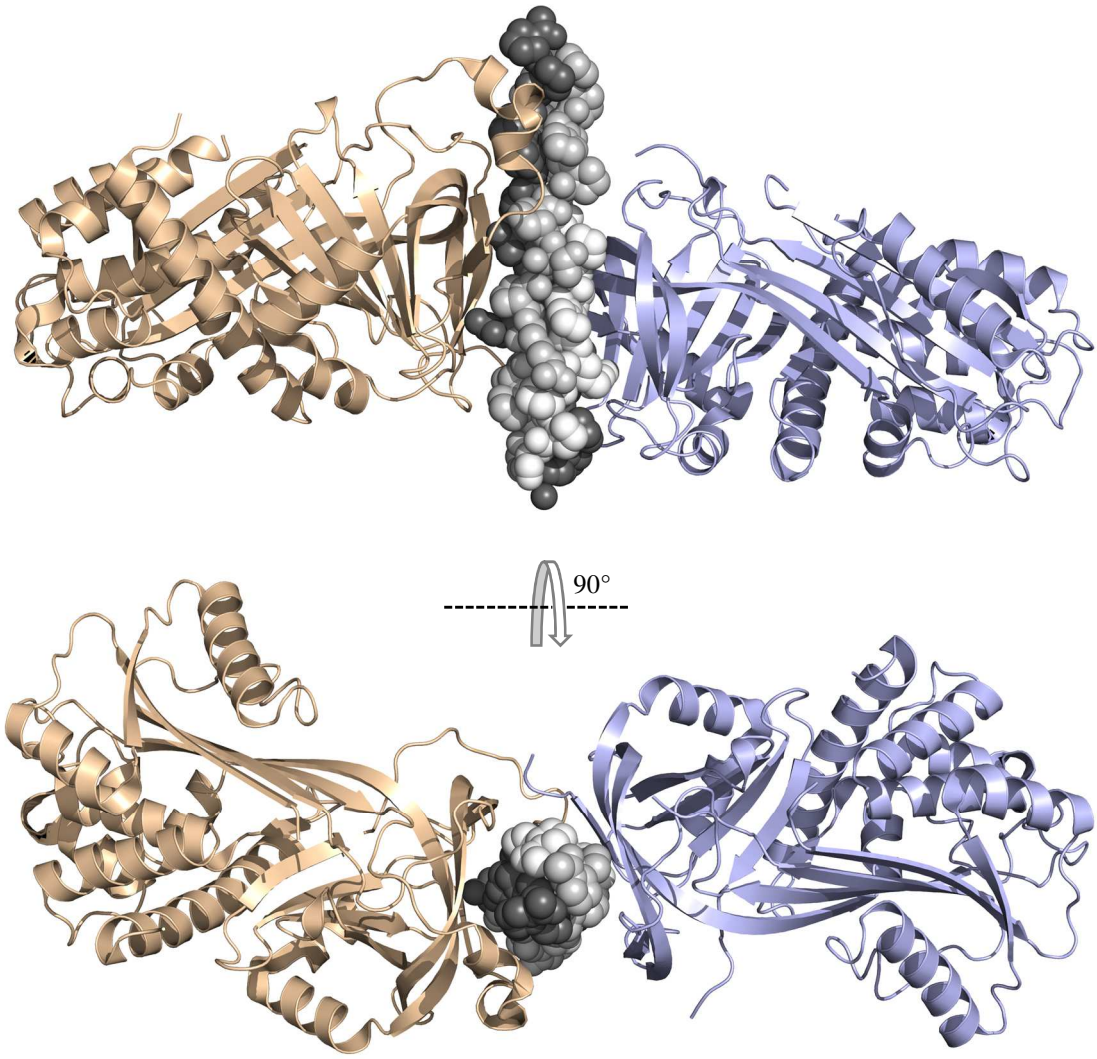


Figure 1.5 | Crystal structure of *cl*HSP47 in complex with a collagen model peptide. The CMP spanning 18 triplets is oriented such that the N-terminus points to the top in the upper panel. The leading, middle and trailing strands are colored white, grey and black, respectively. The homotrimeric CMP can accommodate two molecules of HSP47, which recognize an arginine in the leading *(blue)* and trailing strand *(wheat),* respectively. [PDB ID: 3ZHA].

reported cooperativity in binding[115] stems from the HSP47 induced stabilization of triple-helical collagen, which in turn increases the probability of further HSP47 binding.

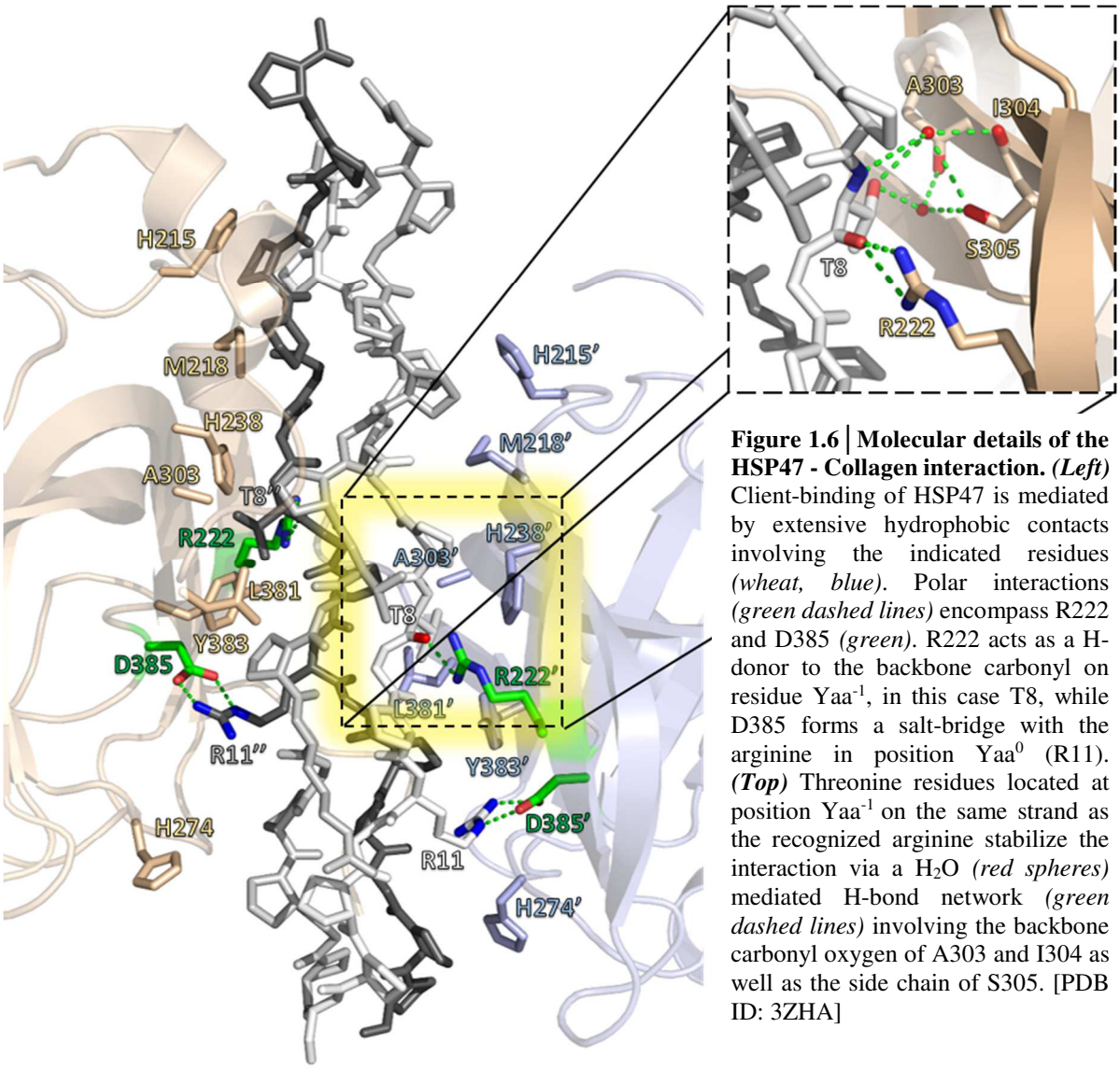


Figure 1.6 | Molecular details of the HSP47 - Collagen interaction. ***(Left)*** Client-binding of HSP47 is mediated by extensive hydrophobic contacts involving the indicated residues *(wheat, blue)*. Polar interactions *(green dashed lines)* encompass R222 and D385 *(green)*. R222 acts as a H-donor to the backbone carbonyl on residue Yaa^{-1}, in this case T8, while D385 forms a salt-bridge with the arginine in position Yaa^{0} (R11). ***(Top)*** Threonine residues located at position Yaa^{-1} on the same strand as the recognized arginine stabilize the interaction via a H_2O *(red spheres)* mediated H-bond network *(green dashed lines)* involving the backbone carbonyl oxygen of A303 and I304 as well as the side chain of S305. [PDB ID: 3ZHA]

1.3 Thesis Aims

The work presented in this thesis revolves around the interaction between the collagen chaperone HSP47 and its client. Two independent aims will be pursued:

(i) Understanding the molecular details of pH-dependent client-release, and

(ii) Identifying small organic molecule inhibitors of the interaction.

Molecular details of the pH-dependent client-release

It has been long postulated that the mildly acidic environment in the Golgi drives dissociation of HSP47 from procollagen. The pH-shift experienced by the complex occurs close to the pKa value of the imidazole side-chain of histidine residues (approximately 6.0, but variable depending on the microenvironment; values ranging from 5.4 to 7.6 have been experimentally determined[123]), suggesting that the protonation of key histidines might trigger client-release. The first part of this work will focus

on validating this hypothesis, identifying the histidine(s) responsible for complex dissociation and understanding the underlying mechanism. This will encompass assessing the impact of histidine mutants* on pH-dependent client-release via interaction studies. These will be supplemented with *in-silico* analysis based on the crystal structure of *cl*HSP47, conducted by Eileen Socher (Research Group Prof. Dr. Sticht, Universität zu Erlangen) in context of a collaborative effort. Furthermore, structural aspects of client-release will be investigated, with the main effort lying on obtaining a low-pH crystal structure of HSP47.

Identifying small organic molecule inhibitors of the HSP47 - collagen interaction

Collagens, being an essential component of the extracellular matrix, are indispensable for the development and homeostasis of tissues and ultimately for the well-being of vertebrates. In humans, aberrant collagen biosynthesis is implicated in a multitude of pathologic conditions, which, due to the ubiquity of collagens often lead to severe diseases that can be lethal. Fibrosis alone, i.e. the excessive deposition of fibrous ECM components, contributes to approximately 45% of mortalities in developed countries[124]. In light of this, HSP47 as a crucial component in collagen biosynthesis has great potential as a therapeutic drug target. The second part of this work will thus focus on developing a high-throughput compatible assay for monitoring HSP47 - collagen complex formation and its application in interrogating a medium-sized (approximately 40,000 compounds) library with the aim of identifying at least one lead structure for small-molecule inhibitors of the HSP47 - collagen complex.

*In the context of this work, the term "histidine mutant" refers to the substitution *of* a histidine, as opposed to substitution *with* one

2. Material and Methods

2.1 Material

2.1.1 Chemicals

Common laboratory chemicals were purchased from AppliChem, Carl-Roth, Hampton Research, Merck KGaA, Promega, Serva, Sigma-Aldrich and Thermo Scientific.

2.1.2 Proteins

Table 2.1 | Enyzmes

Class	Enzyme	Supplier
Restriction endonucleases	DpnI, NdeI, NheI, PstI, SmaI, XbaI, XhoI	New England Biolabs
DNA Polymerases	Pfu DNA Polymerase	Own production
	Q5 High Fidelity DNA Polymerase	New England Biolabs
DNA Ligases	T4 DNA Ligase	Roche
	QuickLigase	New England Biolabs
Proteinases	Porcine Pancreas Elastase I	Sigma-Aldrich

Table 2.2 | Antibody and Protein Conjugates

Class	Conjugate	Supplier
HTRF Reporter	αGST-mAb-Eu^{3+} - Cryptate	Cisbio
	αGST-mAb-Lumi4-Tb^{3+} - Cryptate	Cisbio
	Streptavidin - XL665	Cisbio
HRP-coupled Reporter	Streptavidin - HRP	Thermo Scientific
	α-His_6 mAb - HRP	Thermo Scientific

2.1.3 Buffers and Media

Table 2.3 | Buffers and Media

	Buffer/Medium	Composition
Common Buffers	PBS	2mM KH_2PO_4, 8mM Na_2HPO_4 (pH 7.4), 2.7mM KCl, 137mM NaCl
	HBS	20mM HEPES (pH 7.5), 150 mM NaCl
	TBS	50mM Tris · HCl (pH 7.5), 150mM NaCl
	TBS-T	TBS, 0.05% (v/v) Tween-20
	McIlvaine Buffer	0.1M Citric acid, 0.2M Na_2HPO_4 (Composition defined by cross-titration to yield desired pH)
	TAE	40mM Tris · HCl (pH 7.6), 20mM Acetic acid, 1mM EDTA
Protein Purification	Lysis Buffer	TBS, 0.2mM PMSF, 1mM EDTA, 4mM DTT*
	High-salt Wash Buffer	50mM Tris · HCl (pH 7.5), 500mM NaCl, 4mM DTT*
	Size Exclusion Chromato-graphy Buffer	HBS, 2mM DTT*
SDS-PAGE	Running Buffer	25mM Tris, 192mM Glycine, 1%(w/v) SDS
	Sample Buffer (6 x)	375mM Tris · HCl (pH 6.8), 12% (v/v) β-mercaptoethanol, 60% (v/v) Glycerol, 0.06% (v/v) bromphenolblue
	Staining Solution	20% (v/v) Ethanol, 10% (v/v) Acetic acid. 0.04% (w/v) Coomassie Brilliant Blue G-250
	Destaining Solution	20% (v/v) Ethanol, 10% (v/v) Acetic acid
Immunoblotting	Transfer Buffer	25 mM Tris · HCl (pH 8.3), 192 mM Glycine, 10% (v/v) Methanol
	Developing Buffer	100mM Tris · HCl (pH 8.6), 0.2mM p-Coumaric acid, 1.25mM Luminol, 0.009% (v/v) H_2O_2
HTRF	Standard Assay Buffer	HBS, 0.01% (v/v) Tween-20, 0.1% (w/v) BSA
Agarose Gel Electrophoresis	Running Buffer	40mM Tris · HCl (pH 8.6), 20mM Acetate, 1mM EDTA
Cell Culture	LB-Medium	1% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, % (w/v) NaCl (pH 7.5)
	LB-Agar	LB-Medium, 1.5% (w/v) Agar

*DTT was included only for the purification of HSP47 constructs

2.1.4 Synthetic Peptides

Peptides for use in crystallization experiments were purchased from BioMatik.

2.1.5 Synthetic DNA Oligonucleotides

Primers for PCR-reactions were purchased from Eurofins in "Salt-free" purity and are listed in the appendix (Table A.1).

2.1.6 *E. coli* Strains

Table 2.4 | *E.coli* Strains

Strain	Genotype	Supplier
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Invitrogen
BL21(DE3)	F–*omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3)	Invitrogen

2.1.7 Software

Table 2.5 | Software

Name	Application	Supplier
ACD/Chemsketch	Chemical structures	ACD/Labs
CLC Workbench	DNA, RNA & protein sequence analysis	Qiagen
Excel	Data processing	Microsoft
OriginPro	Data processing, graphical representation	Origin Labs
Powerpoint	Figures	Microsoft
PyMol	Protein structure visualization	Schrödinger LLC
Word	Text editing	Microsoft

2.2 Nucleic Acid Methods

2.2.1 Isolation & Purification of DNA

The plasmid DNA of interest was amplified *in-vivo* in *E. coli* DH5α cells and isolated using the QIAprep Spin Miniprep kit (Qiagen) according to instructions of the manufacturer. The procedure combines a modified version of the alkaline extraction procedure developed by Birnboim & Doly with DNA purification via adsorption on silica in presence of chaotropic salts.

A 5 ml ON culture of cells was pelleted and resuspended in 200 µl hypotonic resuspension buffer supplemented with RNAse A. Lysis was initiated via addition of 200 µl NaOH and SDS containing lysis buffer and terminated within 5 min by addition of 350 µl KOAc containing neutralization buffer. The lysate was centrifuged for 2 min at 10,000 g and the clear supernatant transferred to a silica spin-column. After centrifugation for 1 min at 10,000 g, the column was washed with 700 µl 70% Ethanol and the DNA eluted with 50 µl ddH_2O. After quantification, DNA samples were stored at -20 °C.

2.2.2 Spectrometric Determination of DNA concentration

Since both purine and pyrimidine nucleobases absorb light in the UV range (λ_{max} ranging from 252 nm for pdG to 271 nm for pdC in physiological buffer)[125], the concentration of a DNA solution can be calculated using the Lambert - Beer law after photometric measurement of A_{260nm}. For precise calculations it should be taken into account that the extinction coefficient varies for each nucleobase, between RNA and DNA, and between single-stranded and double-stranded DNA; for the purposes of this thesis, though, the simplified formula [DNA] = $A_{260nm} \cdot 50$ µg/ml was sufficient. Measurements were performed with a NanoDrop (Thermo Scientific) microvolume spectrophotometer.

2.2.3 Agarose Gel Electrophoresis

This method is routinely used for resolving DNA mixtures of varying size. Nucleic acids, being polyanions, can be subjected to electrophoresis in a gel consisting of a three-dimensional mesh of polymer, in this case the oligosaccharide agarose, and buffer solution. Since the charge/mass ratio of DNA and thus the drift velocity is constant independent of size, separation is achieved solely due to increased retention of larger fragments (or less compact forms of plasmid DNA) by the polymeric matrix. The resolving range is determined by the composition of the gel: 1% (w/v) agarose gels can efficiently separate DNA fragments of sizes ranging from 500 bp up to 10,000 bp. For visualization of the DNA, dyes such as ethidium bromide or SYBR Safe are used; these compounds show a strongly increased fluorescence upon intercalation into DNA.

Gels were prepared by dissolving the appropriate amount of agarose in boiling TAE buffer. After cooling down to around 50 °C, SYBR Safe DNA Gel Stain was added according to the supplier's protocol and the solution was cast and left to gelate. DNA samples were mixed with loading dye and loaded onto gels submerged in TAE buffer. Electrophoresis was conducted with a constant voltage of 90 V. The indicator dye bromophenol blue, contained in the loading dye, runs parallel to DNA fragments of approximately 450 bp length and was monitored for timely termination of electrophoresis. The DNA was finally visualized by placing the gel on a 300 nm UV-transilluminator and the size determined via comparison with a reference DNA ladder.

For the extraction of DNA from preparative agarose gels, the DNA bands were cut out, the gel melted at 65 °C and the DNA retained and washed on silica spin-columns as described under 2.2.1.

2.2.4 Polymerase Chain Reaction & Site-directed Mutagenesis

The polymerase chain reaction, first reported 1968 by Kleppe & Khorana and further developed by Kary Mullis in 1983, is a staple technique in molecular biology for the *in-vitro* amplification of DNA. Heat-stable DNA-polymerases from thermophilic organisms are employed to catalyse template guided DNA-synthesis in a cycled reaction. The reaction takes place in alkaline buffer supplied with Mg^{2+} ions, dNTPs and a pair of short (optimally 18 - 22 bp long) oligomers, called primers in analogy to their cellular RNA counterparts. Primers are required by the DNA-polymerase to initiate DNA elongation. They can be designed such that they frame the DNA sequence to be amplified by annealing at the 5' end of the sense (forward primer) as well as the anti-sense strand (reverse primer), respectively.

The PCR cycles between:

(i) Denaturation of double-strand DNA at 95 °C,

(ii) Annealing of the primers at a temperature dependent on their G/C content and the position of these, and

(iii) Extension of the primers in the 3' direction at the optimum temperature of the employed DNA-polymerase.

Starting from the second cycle, the region flanked by the primers is amplified exponentially each round. After selective DpnI digest of the Dam-methylated template DNA, the amplified DNA fragment can be purified via silica spin-columns.

In site-directed mutagenesis, the desired change in DNA sequence is introduced with the primers. Using primers which overlap leads to the amplification of the whole template DNA in a PCR. After DpnI digest of template DNA, the newly synthesized, linear DNA is circularized via ligation using the T4 ligase and transformed into *E. coli* DH5α cells (see next section). The mutagenic primers were designed according to an optimized protocol[126]. The composition of a standard PCR mixture is summarized in table 2.6; table 2.7 shows the parameters of a typical PCR cycle.

Table 2.6 | Composition of PCR Mixtures

Component	Volume / µl
10 x Reaction Buffer	2.5
10mM dNTP Mix	2.5
10µM Fw Primer	1
10µM Rv Primer	1
50ng/µl Template DNA	1
Pfu DNA Polymerase	1
ddH2O	16

Table 2.7 | Parameters of PCR Cycling

Phase	T / °C	t / s	Cycles
Initial Denaturation	95	30	1
Denaturation	95	30	
Annealing	50 - 65	30	30
Elongation	72	120 / kbp	
Final Extension	72	180	1

2.2.5 Restriction Endonuclease Digestion of DNA & Ligation

DNA Fragments which are amplified via PCR can be inserted into an expression vector using bacterial restriction endonucleases and concomitant ligation. The vector is commonly linearized with two endonucleases that create so-called cohesive ends with short single-strand overhangs. After digestion with the same pair of enzymes, the fragment can be ligated into the vector in the correct orientation. If necessary, restriction sites compatible with those present in the expression vector are introduced to the fragment as a 5' extension of the primers.

Template and fragment DNA were incubated for 1 h at 37 °C with the desired endonucleases in the buffer recommended by the manufacturer. After purification via agarose gel electrophoresis, the digested template and fragment DNA were mixed in a stoichiometric ratio of at least 1:3. 9 µl of the DNA mixture were supplemented with 1 µl of the provided 10x buffer and incubated for 2 h at RT following the addition of 0.5 µl T4 ligase. The ligated DNA was finally transformed into competent *E. coli* DH5α cells for amplification (Section 2.3.2).

2.2.6 DNA Sequencing

Named after its developer Fred Sanger, this method of identifying DNA sequences is based on *in-vitro* amplification of the DNA of interest in presence of dideoxynucleotide triphosphates (ddNTPs) that are conjugated to fluorophores with nucleobase-specific emission wavelengths. Incorporation of ddNTPs leads to termination of the replication process due to the lack of the 3'-OH group; after several cycles of the PCR, this chain-termination will have occurred statistically at every single position of the template sequence. This produces fragments of every possible length, with the identity of the 3'-terminal nucleobase being encoded in the fluorophore. The mixture is resolved using capillary electrophoresis, and the sequence assembled via assigning the corresponding nucleobase to fragment size.

Sanger sequencing was performed by GATC Biotech (Konstanz, Germany).

2.3 Cell Culture Methods

2.3.1 Transformation of DNA

Heat-shock transformation of *E. coli* with plasmid DNA is achieved via subjecting the cells to a rapid change in temperature from 0 °C to 42 °C and back to 0 °C. Although the exact mechanism is largely obscure, it has been proposed the heat-pulse initially leads to the release of lipids into the extracellular space, resulting in a decrease of membrane potential as well as in membrane fluidity. The restoration of the latter to normal values after the cold-pulse leads to a further release of lipids from the membrane. This induces formation of transient pores and allows the uptake of DNA, which is facilitated by the decreased membrane potential as well[127]. The heat-shock transformation competency of cells is usually enhanced via prior incubation with ice-cold $CaCl_2$[128]. The inclusion of an antibiotic resistance gene on the expression vector allows the selection of transformed clones.

E. coli DH5α cells were treated with an ice-cold 0.1 M $CaCl_2$ solution for 20 min and stored at -80 °C. For DNA transformation, small aliquots were thawed on ice, 50 ng of plasmid DNA added and the mixture incubated for at least 20 min. Heat-shock was induced via heating the cell suspension to 42 °C for 2 min and subsequent cooling on ice for 30 min. The cells were transferred to fresh LB medium and allowed to recover at 37°C for at least 30 min, after which they were spread on LB-agar plates with the appropriate antibiotic (Ampicillin: 100mg/l, Kanamycin: 50mg/l).

2.3.2 Protein Expression in *E. coli*

Protein expression was conducted in *E. coli* BL21 (DE3) cells using the T7 RNA polymerase/promoter system established by Tabor *et al.*[129]. T7 RNA polymerase displays exceptionally high specificity

towards its promoter and low rates of transcription termination, ensuring overexpression of the desired protein by the cell. The genes for both the polymerase and the protein of interest are under control of the lac operon and as such are inducible via addition of IPTG.

A single colony was picked from a selection plate and transferred to 50 ml sterile LB medium supplemented with the appropriate antibiotic. The cell suspension was kept overnight at 37 °C in Erlenmeyer flasks under constant shaking with 180-200 rpm to ensure proper gas exchange and to prevent sedimentation. The ON culture was diluted 1:20 with antibiotic supplemented LB, and cultivated to an optical density of A_{600nm} = 0.6 - 0.8 (for expression times up to 5 h) or 1.0 - 1.2 (for overnight expression) and equilibrated to the desired temperature. Protein expression was induced via addition of 0.1 - 1 mM IPTG. The cells were harvested after the optimal expression time by centrifugation for 20 min at 4,000 g in a chilled centrifuge. The cells were then re-suspended and washed with PBS, after which they were pelleted and stored at -20 °C. In case of Hsp47, flasks without chicanes were discovered to give a better yield.

2.3.3 Cell lysis

Cell lysis was achieved using sonication. Ultrasonic waves traversing through a liquid medium induce the formation and growth of small vapour cavities. The rapid collapse of these in a process called cavitation generates a liquid jet and shock waves that lead to extreme local temperatures (up to 5000 K) and pressures of (up to 1000 atm). The dissipation of this energy is utilized to rupture cellular membranes.

The cell pellets were thawed on ice, re-suspended in lysis buffer to yield a 5% (w/v) suspension and sonicated on ice in portions of 40 ml using a 50 ml Falcon tube. The sonotrode was immersed ca. 1-2 cm into the cell suspension and a 2 s pulse, followed by a 2 s cooling pause, was applied at 40% amplitude for a total of 15 min pulse time. Cell debris and aggregates were removed via centrifugation at 4 °C for 20 min with a minimum of 50,000 g.

2.4 Protein Biochemistry Methods

2.4.1 Chromatographic Methods

Initial enrichment of the protein of interest from crude cell lysate was achieved using affinity chromatography. The common feature of this ensemble of protein purification methods is the immobilization of the target protein on a resin, which allows for easy washing away of contaminants using batch or continuous flow techniques. The solid phase is functionalized with moieties capable of binding genetically encoded “tags” fused to the protein of interest. These tags can be whole proteins or short amino acid chains and are usually added as N- or C- terminal extensions separated via short linker

regions from the core protein. In this work, immobilized metal affinity chromatography (IMAC), glutathione affinity chromatograpy and the Strep-tactin/Strep-tag system were used.

In its most widespread form, IMAC is based on the chelation of immobilized Ni^{2+} or Co^{2+} cations via polyhistidine tags consisting of 5 - 10 neighboring histidines. Commonly, cross-linked beaded agarose (Sepharose) functionalized with the chelators nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA) serves as the solid phase. Binding of the protein of interest occurs via formation of a coordination complex between the metal ions and imidazole sidechains via either one of their nitrogen atoms. Elution of the protein of interest can be achieved via displacement with high concentrations of imidazole or decreasing pH.

Glutathione affinity chromatography exploits the high affinity enzyme-substrate binding reaction between the 26 kDa protein Glutathione-S-Transferase (GST) and its natural substrate, reduced glutathione. The latter is covalently attached to a solid phase via its sulfhydryl moiety and can be used to immobilize target proteins carrying a GST fusion tag. Elution of the protein of interest is achieved through displacement with an excess of reduced glutathione. Besides serving as an affinity tag, the highly soluble and rapidly folding GST is also commonly utilized to increase the expression yield of target proteins.

The Strep-tag system is based on an eight amino-acid affinity peptide which binds selectively to streptavidin. The affinity tag was initially discovered from a genetic random library and binds reversibly at the same site as (D)-biotin, the natural ligand of Streptavidin[130]. The commercially available system has been further optimized and utilizes Streptactin, an engineered version of Streptavidin, and the Strep[II]-tag WSHPQFEK[131]. Elution of the bound protein is achieved via displacement with an excess of (D)-desthiobiotin.

In a further polishing step, the eluate from the affinity chromatography can be subjected to size exclusion chromatography. In size exclusion chromatography, a mixture of proteins is passed through a matrix consisting of microbeads with pores of variable size. The resin usually consists of cross-linked dextran (Sephadex). Proteins are resolved according to their hydrodynamic radius and shape, with larger proteins eluting earlier since they are restricted from diffusing into smaller pores. This method is a useful complement to affinity chromatography, since aggregates and other contaminants are removed; furthermore, it serves to remove the eluting agent without the need for dialysis or other buffer exchange methods.

Ni-NTA Affinity Chromatography: The crude cell lysate was supplemented with 10 mM Imidazole and passed over 1 - 5 ml pre-equilibrated Ni-NTA Sepharose resin with an approximate rate of 1 ml/min. The resin was washed with at least 10 column volumes (CV) TBS containing 50 mM Imidazole and 10 CV high-salt wash buffer. Elution was performed with 1-2 CV TBS containing 250 mM Imidazole.

Glutathione Affinity Chromatography: The crude cell lysate was passed over 5 ml pre-equilibrated Glutathione Sepharose. The resin was washed with at least 10 CV TBS and 10 CV high-salt wash buffer. The protein of interest was eluted with 1-2 CV TBS pH 8.0 supplemented with 10 mM reduced glutathione.

Strep-tactin Affinity Chromatography: The crude cell lysate was passed over 1-5 ml pre-equilibrated Strep-tactin resin. The resin was washed with at least 10 CV TBS followed by 10 CV high-salt wash buffer and the protein eluted with 1-2 CV TBS containing 2.5 mM (D)-desthiobiotin.

Size-exclusion Chromatography: The eluate from the affinity chromatography was concentrated up to a maximal volume of 2 ml using ultrafiltration units with the corresponding MWCO. The sample was then resolved in Superdex S200 packed 16/60 columns using the AEKTA Purifier chromatograpy system and a flow-rate of 1 ml/min.

2.4.2 SDS-PAGE

In this staple technique developed by Laemmli in 1970, proteins are denatured in presence of the anionic surfactant sodium dodecylsulfate (SDS) as well as the reducing agent β-mercaptoethanol, and separated according to their electrophoretic mobility μ in a polyacrylamide gel. The initial denaturation step produces unstructured, linear peptide chains with a practically uniform negative charge-to-mass ratio resulting from the bound SDS. This serves to negate the influence of protein shape, flexibility and net charge on μ, resulting in a strict dependence of sample retention on protein mass. In discontinuous SDS-PAGE, samples are focused into a thin band using isotachophoresis prior to their separation. This occurs in the stacking gel, which has a less dense polymer mesh and a lower pH (pH 6.8) compared to the resolving gel (pH 8.8). Upon application of an electric field, highly mobile Cl^- anions present in the gel quickly form a front of leading electrolyte (LE). The efflux of charge from the gel is compensated by glycinate ions moving into the stacking gel from the running buffer (pH 8.3); these are protonated to yield zwitterionic glycine, which forms a front of trailing electrolyte (TE) due to its low electrophoretic mobility. The zone between the LE and TE fronts, called the Kohlrausch discontinuity, is characterized by a steep gradient in electric potential stemming from excess counter-ions left behind by the LE. This causes the TE front to migrate at the same speed as the LE front. Proteins with μ values in between those for chlorine and glycine are stacked into thin layers between the ion boundaries as the double front passes through. Upon entering the resolving gel, deprotonation of glycine leads to a collapse of the Kohlrausch discontinuity and standard electrophoretic resolution occurs. The protein is fixed in the gel with an acetic acid and ethanol containing solution and traditionally stained with the protein binding dye Coomassie Brilliant Blue (CBB) R-250 or its colloidal form, CBB G-250. A faster method for visualizing the protein involves an UV-light induced reaction of tryptophan residues with 2,2,2-trichloroethanol (TCE). This results in the formylation of the indole side-chain, generating a

species which fluoresces in the visible light range and thus can be easily detected using a 300 nm-transilluminator[132].

SDS-polyacrylamide gels were prepared from the mixtures summarized in table 2.8. Polymerisation was initialized via addition of the catalyst tetramethylethylenediamine (TEMED) and the free radical initiator ammonium persulfate (APS) as the last two components. The gels were cast using a gel-casting set from Bio-Rad.

Table 2.8 | Composition of SDS-Polyacrylamide Gels

	Volume per Gel / ml		
	12% Resolving Gel	***16% Resolving Gel***	***Stacking Gel***
30% (v/v) Acrylamide / Bis-acrylamide	1.80	2.40	0.42
3M Tris · HCl (pH 8.8)	0.75	0.75	-
1M Tris · HCl (pH 6.8)	-	-	0.32
10% (w/v) SDS	0.05	0.05	0.05
10% (w/v) APS	0.068	0.068	0.05
TEMED	0.003	0.003	0.003
TCE	0.05	0.05	-
H_2O	1.91	1.31	1.75

2.4.3 Western Blotting & Immunodetection

Western blotting involves separation, immobilization and subsequent detection of specific proteins using antibodies coupled to conjugated reporter enzymes. The protein of interest is electrophoretically transferred to a hydrophobic membrane (polyvinylidene difluoride (PVDF) or nitrocellulose) and recognized by monoclonal antibodies that are directed against common affinity tags or the protein itself. This is followed by addition of a secondary (polyclonal) antibody (directed against the primary antibody) which is conjugated to the reporter enzyme, after which the complex can be visualized via the coupled color-reaction. In this work, signal enhancement using the secondary antibody was omitted via utilization of α-His_6 mAb - HRP or Strep-tactin - HRP. For the detection of protein bands, enhanced chemiluminescence (ECL) was used. ECL involves HRP catalysed oxidation of luminol by H_2O_2, yielding 3-aminophtalate in an excited electronic state. Decay of the latter is coupled to the emission of a photon and can thus be quantified.

Protein mixtures resolved via SDS-PAGE were transferred via electrophoresis to either a Nitrocellulose membrane or a PVDP membrane pre-treated with methanol. In this work, the “wet-blot” method was utilized, which requires complete submersion of gel and membrane in transfer buffer (TAE buffer). Blotting was conducted for 1 h at 8 °C with a current of 1.5 mA per cm^2 membrane. The membrane

was then blocked for 1 h with 1% (w/v) milk powder and incubated for 1 h with the appropriate antibody, after which the bands were visualized according to a standard ECL protocol[133].

2.4.4 Biotinylation of Proteins

The selective reaction of primary amines by NHS-activated esters was exploited for chemical coupling of biotin to lysine side-chains and the N-terminus of proteins[134] (It should be noted that the selectivity of this reaction is debated, and acylation of Ser, Thr, Tyr, and even Arg side chains via commonly employed amine reactive esters is reported in literature[135]).

1 mg protein in 1 ml HBS was supplemented with 20 eq NHS-Biotin (Thermo Scientific) from a 20 mM DMSO stock. The Eppendorf tube was vigorously shaken at RT for 30 min, after which the reaction was quenched via addition of 100 mM TRIS. Excess biotin was removed via dialysis or SEC.

2.4.5 Spectrometric Determination of Protein concentration

The concentration of protein samples was determined via measurement of the transmission at $\lambda = 280$ nm using a NanoDrop (Thermo Scientific) microvolume spectrophotometer. At this wavelength, the aromatic side-chains of tryptophan and tyrosine residues absorb strongly, with extinction coefficients close to 5,500 $M^{-1}cm^{-1}$ and 1,490 $M^{-1}cm^{-1}$, respectively[136]. Disulphide bonds contribute weakly to the absorption (ε_{280nm} = 125 $M^{-1}cm^{-1}$)[136], but have to be taken into account when using buffers with dithiothreitol (DTT) as an additive. The extinction coefficients of the used protein constructs were calculated from the sequence using the online tool http://web.expasy.org/protparam.

Since the synthetic CMPs used in this work did not contain any aromatic residues, their concentration was determined via quantification of the peptide bond absorption at $\lambda = 214$ nm. The extinction coefficients of the peptides were calculated from available literature values[137].

2.4.6 Colorimetric Protein Assays

Protein concentration determination via measurement of A_{280nm} was supplemented with colorimetric assays. In these, protein concentration is reported either through proportional association of an environment-sensitive dye to the protein, changing its spectral properties, or through generation of an intermediate species which reacts further to yield a UV-VIS detectable species. The color development is quantified and correlated to protein concentration using a calibration curve prepared with known concentrations of a standard protein, usually bovine serum albumin (BSA).

The *Bradford assay* is based on the dye Coomassie Brilliant Blue, which experiences a large red-shift of its absorption spectrum (λ = 465 nm to 595 nm) upon deprotonation. The anionic form is stabilized upon binding of the dye to basic and hydrophobic side chains of proteins and thus can be utilized to quantify the amount of protein. In the *BCA-Assay*, Cu^{2+} is reduced to Cu^{1+} by the protein peptide bonds;

the monovalent cation is then complexed by two molecules of bicinchoninic acid, forming a complex strongly absorbing at $\lambda = 562$ nm.

2.4.7 Differential Scanning Fluorimetry

This thermal shift assay exploits the change in fluorescence intensity of an environment-sensitive reporter dye upon binding of to the hydrophobic core of proteins exposed by thermal denaturation. The variant used in this work utilizes the dye SYPRO Orange, which can also be used to stain protein gels with a much higher sensitivity than Coomassie dyes[138]. The binding of SYPRO Orange to proteins occurs non-specifically to hydrophobic patches on proteins. Since the fluorescence of the dye is effectively quenched by water, recruitment to hydrophobic patches and the associated shielding of the fluorophore from the solvent leads to an increase in its quantum yield. The exposure of the hydrophobic core of thermally denatured proteins can thus be detected by a rather sharp increase in SYPRO Orange emission intensity at temperatures nearing the T_m.

White 96-Well PCR plates (BioRad) were loaded with 45 µl protein solution diluted in the desired buffer to a final concentration of 0.1 - 0.2 mg/ml. After addition of 5 µl SYPRO Orange from a 10x stock solution, the plate was sealed with a transparent adhesive seal, briefly spun down and placed in a CFX96 real-time PCR Machine (BioRad). Starting from 20 °C, the block was heated at a rate of 1 °C/min and the fluorescence intensity (excitation: 492 nm, emission: 516 nm) recorded in 5 s intervals. The T_m was deduced from a plot of intensity against temperature as the inflection point of the slope.

2.4.8 Circular Dichroism Spectroscopy

The differential absorption of left- and right-circular polarized light by optically active chiral molecules, called circular dichroism and quantified using ellipticity (θ), is a method commonly employed to investigate higher-order structures of chiral macromolecules such as DNA and proteins. Circular polarised light is obtained via passing linear polarized light through a birefringent crystal at an angle of 45°. This splits the light into two equal, perpendicular electric field vectors, one of which is retarded due to the birefringent character of the crystal. Using so called quarter-wave plates, this phase shift can be adjusted to equal $\lambda/4$, resulting in a helical progression of the resulting electric field vector and thus yielding circular polarisation. For studying protein secondary structure, both far UV spectra (180 nm to 240 nm) as well as near UV spectra (> 250 nm) can be recorded. The former involve $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions in the peptide bond, which absorb around wavelengths of 190 nm and 210 nm, respectively. Since transition dipole moments of chiral molecules in an ordered structure influence each other (e.g. via exciton coupling), each secondary structure has a characteristic fingerprint and the far UV spectrum provides information about secondary structure element content in the protein. The near UV spectrum primarily involves absorption by the aromatic sidechains of tryptophan, tyrosine and to a lesser extent phenylalanine, as well as disulphide bonds. These elements are typically interspersed in

the protein and contribute independently to the spectrum, making near UV CD-spectroscopy suitable rather for detection of changes in the tertiary structure than deduction of it.

The protein was dialyzed extensively against phosphate buffer pH 7.0 and diluted to a final concentration of 0.1 mg/ml in phosphate buffer with the desired pH. The sample was transferred to a 1 mm quartz cuvette and spectra were recorded at 20 °C with a Jasco J-715 spectropolarimeter. Instrument parameters were set to a bandwidth of 2 nm, a response of 2 s and a sensitivity of 100 mdeg. The CD-spectra were averaged from 10 accumulations recorded with a data pitch of 0.2nm and a scan speed of 50 nm/min each.

2.4.9 Isothermal Titration Calorimetry

This calorimetric technique is a valuable tool to study biomolecular interactions, revealing the enthalpy (ΔH), entropy (ΔS), affinity (K_D) and stoichiometry of binding in a single experiment using unmodified sample. In a typical ITC experiment, the ligand is titrated into a thermally insulated sample cell containing the interaction partner. The heat released or taken up by the binding is quantified via the difference in electric heating power required to reach temperature equilibrium between the sample cell and a reference cell. Binding isotherms are shown using the Wiseman plot, in which the change in heat per mol injectant is plotted against the molar ratio of the interaction partners.

HSP47 and the CMP were extensively dialysed against the same buffer and calibrated to a concentration of 300 µM and 30 µM, respectively. 70 µl of HSP47 solution were transferred to the syringe unit of a GE Healthcare MicroCal ITC200 calorimeter, while the sample cell was filled with 350 µl of the CMP solution. The cell temperature was equilibrated to the desired temperature and the reference power set to 10 µcal/s. The stirring speed was adjusted to 800 rpm and a total of 40 injections of 2.5 µl over 2.5 s were recorded with a spacing of 150 s.

2.4.10 Dynamic Light Scattering

Also known as quasi-elastic light scattering, DLS allows the determination of the size distribution in solutions or suspensions of small particles. The scattering of monochromatic light by small particles (in context of this work, proteins) in solution leads to an arbitrary pattern of constructive and deconstructive interference (speckle pattern) on a small detection area. As the particles undergo Brownian motion, the interference pattern shifts, and the detected intensities of the spots on the speckle pattern fluctuate. These are sampled after short intervals of time and the spot intensities analysed using the intensity auto-correlation function, with smaller particles showing a fast decorrelation due to the higher velocity of random motion.

100 µl of protein sample were retrieved directly after SEC and centrifuged for 10 min at 10,000 g to remove aggregates and contaminant particles. The solution was then transferred into UVettes (Eppendorf) and the measurements conducted using a DynaPro NanoStar (Wyatt) instrument with a

laser wavelength of 658 nm. For determination of the polydispersity, five measurements with 10 acquisitions each were averaged.

2.4.11 Biolayer Interferometry

Biolayer interferometry (BLI) is a relatively new optical method which, similar to the more widely known surface plasmon resonance spectroscopy, is used in interaction studies to determine the affinity and kinetic parameters of binding reactions. Using either chemical ligation or one of the commonly used tags such as the His_6- or Myc-tag, one of the interaction partners is immobilized on an appropriately functionalized biosensor. The biosensor is then dipped into a solution of the other binding partner and white light passed through. Reflection of the light occuring at an internal reference layer as well as at the periphery of the biolayer leads to an interference pattern which is subject to change as the optical density of the biolayer is altered due to association/dissociation events.

Experiments were performed with streptavidin coated SA- or SAX-biosensors using the BLItz System (Pall ForteBio). The sensors were loaded with 10 µM CMP-R18-biotin to a maximal response of approximately 5 nm, and free streptavidin binding sites were blocked with 10 µM biocytin. HSP47 (C-Strep) was diluted at least 1:10 into the desired McIlvaine buffer to yield a final concentration of 5 µM. The set time parameters were: 10 s baseline, 60 s association, 80 - 140 s dissociation and 10 - 40 s regeneration. All sensorgrams were recorded in triplicate and the measurements performed in blocks of 1 to 3 mutants per day, with wild type HSP47 being included in each block as a reference. The kinetic parameters k_{on} and k_{off} were extracted through non-linear curve fitting using the following equations for the association (Eq. 1) and dissociation (Eq. 2) phases:

$$R = R_{max} \frac{1}{1 + \frac{k_{off}}{k_{on} \cdot [A]}} \left(1 - e^{-(k_{on} \cdot [A] + k_{off}) \cdot t}\right) \qquad \textit{(Eq. 1)}$$

$$R = R_0 \cdot e^{-(k_{off} \cdot t)} \; R_{t \to \infty} \qquad \textit{(Eq. 2)}$$

R = Response, [A] = Analyte concentration, t = Time

The K_D was subsequently calculated from the ratio of the kinetic parameters ($K_D = k_{off} / k_{on}$).

2.4.12 Fluorescence Spectroscopy & Homogenous Time-Resolved FRET

Förster Resonance Energy Transfer (FRET)

The transfer of energy between chromophores via dipole-dipole coupling, FRET, has found many uses in biochemistry as a tool to study molecular interactions and dynamics. Virtually all applications exploit the strong dependency of FRET-efficiency on the distance between the chromophores. The efficiency

is inversely proportional to the 6th power of the distance, which, depending on chromophore properties, translates into a practical effective range of up to 10 nm. This order of magnitude covers the diameter of most proteins (the minimum radius of a 500 kDa globular protein has been calculated to be 5.21 nm[139]), making FRET ideally suited to study protein interactions or conformational rearrangements.

Homogeneous Time-resolved FRET

The HTRF technology extends the basic FRET technique via incorporation of time-resolved fluorimetry. The basic premise is the inclusion of a FRET-donor with a long emission half-life, allowing a delayed read-out, typically after 50-150 µs. This has the advantages of:

(i) Bypassing short-lived fluorescence originating from buffer components, biomolecules or small organic molecules, and

(ii) Providing an internal standard via quantifying the ratio of acceptor fluorescence and residual donor fluorescence.

A key component of the HTRF technique are trivalent lanthanide cations. Two properties render them ideally suited for employment as a FRET-donor: they display a large stokes-shift and possess long excited-state lifetimes in the order of miliseconds. Since lanthanide cations have very small absorption cross-sections (i.e. the probability of an absorption process), their utilization in fluorescence based experiments requires photosensitization via coordination to an organic chromophore, called an "antenna", which typically has a high extinction coefficient and the capacity to transfer energy to the metal ion[140]. This is commonly achieved via insertion of the lanthanide ion as the guest cation into cryptands. As a FRET-acceptor with compatible excitation and emission properties, a modified phycobilliprotein pigment from red algae, XL665, is used.

Assay Establishment

The reporters αGST-mAb-Lumi4®-Tb^{3+}-Cryptate and Streptavidin-XL665 were purchased from Cisbio as lyophilized powder, dissolved in ultrapure water, aliquoted and stored at -80 °C and -20 °C, respectively. Prior to the experiment, the reporters were thawed for at least 20 min on ice, diluted with standard assay buffer to yield a 4x solution based on the manufacturer's recommended final concentration. The resulting solutions were then mixed 1:1 to yield the 2x reporter mixture. The analyte proteins were thawed in hand, diluted with standard assay buffer to yield a 4x solution of the desired concentration and kept at RT.

For the assay, 10 µl of the reporter mixture were dispensed into a white, low-volume, flat-bottom 384-well plate (Greiner) and starting with HSP47, 5 µl of each analyte solution were added to yield a final volume of 20 µl per well. The plate was covered airtight with aluminium foil and incubated for 1 h at RT. In displacement assays, 0.4 µl of a 5x stock solution of the inhibitor were added directly after HSP47 and the plate was incubated for 15 - 60 min before the biotinylated CMP was added. The plate

was read out using a Synergy H4 plate-reader (BioTek) with the following components installed: excitation filter (330/80), emission filter 1 (620/10), emission filter 2 (665/8), dichroic mirror (400 nm cut-off). The optics position was set to top 400 nm and the read height to 7.0 mm. Prior to the measurement, the gain was scaled to a value of 50,000 for the well with the highest fluorescence intensity. Emission intensities at 620 nm and 665 nm were recorded with a 500 µs delay after excitation at 340 nm.

Statistical evaluation of assay robustness

The Z-factor is a screening window coefficient which is commonly utilized to assess the reliability and robustness of an assay[141]. The parameters taken into account are the signal dynamic range, i.e. the window between the mean values for positive and negative controls, as well as the variability within the data itself. The Z-factor is defined as:

$$Z = 1 - \frac{(3\sigma_s - 3\sigma_c)}{|\mu_s - \mu_c|} \qquad \textit{(Eq. 3)}$$

σ_s = Standard deviation of samples
σ_c = Standard deviation of controls
μ_S = Mean of library sample signal
μ_S = Mean of control signal

While the Z-factor reflects the quality of an already conducted HTS assay, the closely related Z'-factor only takes into account the control data and thus quantifies the intrinsic reliability of the assay system without interference by compounds (Eq. 4).

$$Z' = 1 - \frac{(3\sigma_{c+} - 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|} \qquad \textit{(Eq. 4)}$$

σ_{c+} = Standard deviation of positive controls
σ_{c-} = Standard deviation of negative controls
μ_{c+} = Mean of positive control signal
μ_{c-} = Mean of negative control signal

A Z-factor of 1 describes an ideal assay, while Z-factors larger than 0.5 are deemed excellent.

The Hook-Effect

The Hook effect, which is often observed in immunoassays, describes a phenomenon in which a too large analyte concentration leads to a decrease in signal and thus a false negative result. In context of the HTRF-assay, this occurs due to excess analyte saturating each reporter type separately, reducing the

population of signal generating species (i.e. analyte molecules which have recruited one molecule of each reporter type).

2.4.13 Protein Crystallization

Supersaturated solutions are thermodynamically unstable systems. Minimization of the free energy of the system drives the aggregation and removal of excess solute as a solid, which can either be in form of amorphous precipitate or in form of crystals. This principle underlies the strategies to obtain proteins in crystalline form, which is the pre-requisite for X-ray crystallographic structure determination. Achieving supersaturation is commonly aided by additives such as salts, polymers or organic solvents. These crystallants act via modulating the physicochemical properties of the bulk solvent, for instance decreasing the solvation capacity, exerting volume exclusion effects or altering the dielectric constant, and can also promote protein - protein interactions. Other critical factors are the pH of the solution, which influences the protein itself, and temperature, which directly affects the solubility.

In this work, the vapor diffusion technique was utilized for protein crystallization. The protein solution is mixed with precipitant solution, typically in a volumetric ratio of around 1:1, and the drop sealed airtight together with a reservoir of the latter. Due to the difference in precipitant concentration between drop and reservoir, water will diffuse from the former into vapor phase, gradually leading to supersaturation of the protein solution and, if the conditions have been chosen right, to crystal nucleation and growth. Since it is nigh impossible to predict which protein will crystallize under which conditions, obtaining protein crystals usually involves screening against a variety of conditions.

Crystallization of HSP47 was performed according to the protocol established by Widmer *et al*[122]. For initial screening purposes, the protein at a concentration of 20 - 24 mg/ml was transferred to 96-Well sitting drop iQ plates (TTP labtech) and mixed with precipitant solution in ratios of 1:1, 1:2 and 2:1 using a Mosquito pipetting robot (TTP labtech) to yield a total drop volume of 300 nl. For conditions that yielded crystals, the drop volume was up-scaled to 2 µl using CombiClover Junior plates (Jena biosciences). The reservoir volumes were set at 70 µl for TTP plates and 300 µl for Clover plates. The plates were sealed airtight, stored at 20 °C and regularly inspected for crystal growth.

3. Results

3.1 Protein Production and Characterization

3.1.1 Proteins and Peptides used in this Work

HSP47 Constructs

The HSP47 constructs used in this work are all based on *Canis lupus familiaris* HSP47, encompassing residues L^{36} through L^{418}. The residue numbering as it appears in this work is adjusted accordingly (e.g. L^{36} = [M]L^{2}). The omitted residues consist of the ER-localization signal peptide and an unstructured linker region detrimental for crystallization. The original cDNA was purchased from DNA 2.0 as codon-optimized for *E. coli*. HSP47 variants with three different affinity-tags were used: one with a C-terminal Strep[II]-tag (HSP47 C-Strep), one with a C-Terminal His_6-tag (HSP47 C-His_6), and one with a N-Terminal GST fusion (HSP47 N-GST). The first two constructs were provided by Dr. Jan Gebauer at the beginning of this work, while the latter was cloned by the author. For investigation of the pH-dependent client release mechanism, key histidine residues in HSP47 were substituted with asparagine or alanine as described under 2.2.4. The success of mutagenesis was validated via Sanger sequencing performed by GATC Biotech.

Collagen Model Peptides

Initial experiments with CMPs obtained via solid-phase peptide synthesis were abandoned in favour of heterologously expressed CMPs carrying a C-terminal bacteriophage T4-Fibritin Foldon-domain, a trimeric beta-hairpin propeller which bestows the triple-helical CMP species a higher thermostability[142]. The constructs utilized in this work consist of the collagenous host-guest system $(GPP)_5$-(GlyProYaa)-$(GPP)_6$ (Yaa=Asp, Arg), flanked C-terminally by the Foldon-domain and N-terminally by either a twin Strep[II]-tag or by a His_6-tag. Unless otherwise stated, all mentions of CMP in this work refer to this construct, with the amino acid in position Yaa also specified when necessary (e.g. CMP-D18). The plasmids encoding for the Foldon constructs were a generous donation by Dr. Manuel Koch.

For crystallization experiments, synthetic CMPs purchased from Biomatik were utilized. The designation used in this work denotes the total length of the peptide together with the position of the residues required for HSP47 binding (e.g. (18)T8R10 denotes the peptide $(GPP)_2$-GPTGPR-$(GPP)_2$). Incorporation of hydroxyprolines in the Xaa position is indicated by the prefix Hyp (i.e. Hyp(21)R15).

3.1.2 Heterologous Protein Expression in *E. coli*

Optimal expression conditions for HSP47 (C-Strep), HSP47 (C-His_6) and both CMP constructs had been previously established and were only changed slightly over the course of this work. In case of HSP47 (N-GST), these were determined via small-scale test expressions. Of all the constructs and

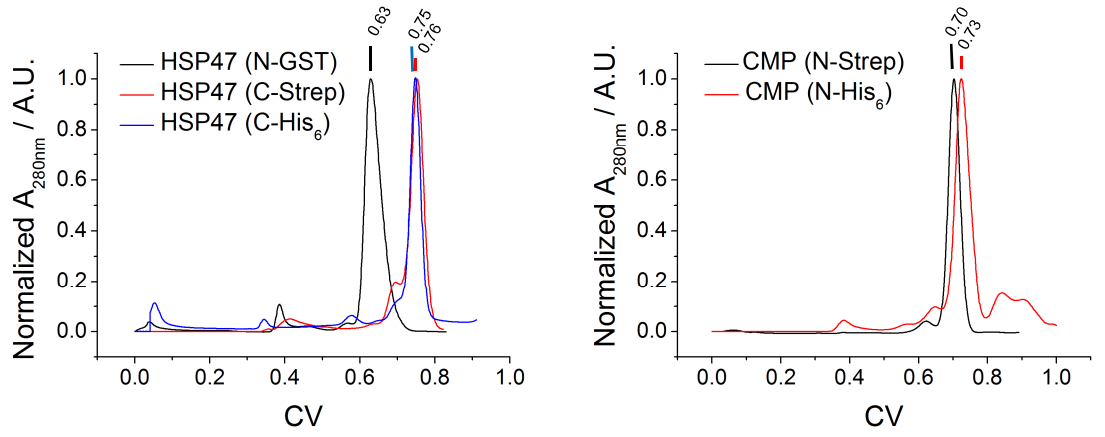
mutants that were attempted to be produced, only the two mutants $HSP47_{H315N}$ and $HSP47_{H386N}$ failed to express. The expression parameters yielding the largest amount of soluble protein are summarized for each construct under table 3.1.

Table 3.1 | Optimal expression conditions for various protein constructs.

Construct	Base Vector	Induction parameters			Expression time
		T / °C	OD_{600}	[IPTG] / mM	t / h
HSP47 (C-Strep)	pJexpress411	37	0.6	0.5	4
HSP47 (C-His_6)	pET-22(b) (+)	37	0.6	0.5	4
HSP47 (N-GST)	pGEX-4T1	30	0.6	0.1	5
CMP (N-Strep)	pET-30a (+)	20	1.0	1	16
CMP (N-His_6)	pET-28a (+)	20	1.0	1	16

3.1.3 Protein Purification

The expressed proteins were enriched from crude cell lysate using affinity chromatography and subsequently subjected to SEC. Typical SEC elution profiles of the various constructs are displayed in Figure 3.1. SDS-PAGE analysis confirms that the two-step purification protocol yields protein of satisfactory purity (Fig. 3.2).



	HSP47 (N-GST)	HSP47 (C-Strep)	HSP47 (C-His_6)	CMP (N-Strep)	CMP (N-His_6)
MW / kDa	69.8	44.6	44.1	31.5	26.7

Figure 3.1 | Size-exclusion chromatograms of protein constructs. Displayed are typical SEC elution profiles of wild-type HSP47 ***(top left)*** and CMP ***(top right)*** constructs. All runs were performed on a HiLoad Superdex S200 16/60 column using a flow rate of 1 ml/min. Due to differences in yield, A_{280nm} was normalized to the highest signal for better visualisation before plotting against the column volume. (***Bottom)*** The corresponding molecular weight of the constructs.

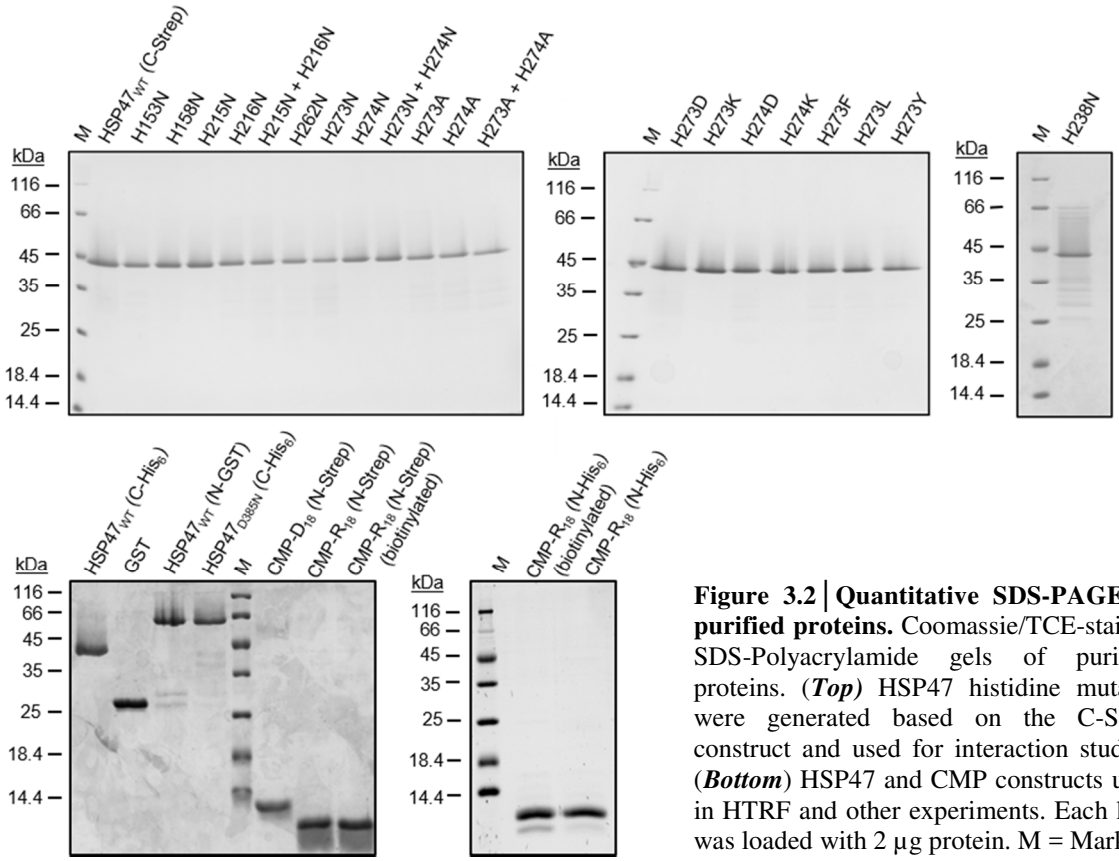


Figure 3.2 | Quantitative SDS-PAGE of purified proteins. Coomassie/TCE-stained SDS-Polyacrylamide gels of purified proteins. (***Top)*** HSP47 histidine mutants were generated based on the C-Strep construct and used for interaction studies. (***Bottom***) HSP47 and CMP constructs used in HTRF and other experiments. Each lane was loaded with 2 µg protein. M = Marker.

3.1.4 Protein Quality Control

Differential Scanning Fluorimetry

The thermal stability of the generated mutants was assessed via DSF. Excluding H238N, all mutants were observed to possess T_m values very close to that of the wild-type (T_m = 57 °C), indicating that the mutants were stably folded (Fig. 3.3). H238N, on the other hand, displayed a T_m 12 °C below wild-type, implying a substantial impairment of the structural integrity by the mutation. This is also reflected in the aggregation-prone character of the mutant. DSF was further used to monitor the stability of $HSP47_{WT}$ in the pH range relevant to this work (Fig. 3.4). The T_m can already be seen to decrease discernibly below pH values of 6.5 and at pH 5.6 plummets to 10 °C below its value at neutral pH.

One peculiarity observed for HSP47 in these experiments was the exceptionally high SYPRO Orange fluorescence intensity measured prior to the thermal unfolding of the protein. In a typical thermofluor experiment, the initial fluorescence intensity is indistinguishable from the background, while in the case of HSP47 it even surpasses the maximum signal generated by full unfolding of the protein. This suggests that the dye is shielded from the solvent by native HSP47, presumably due to the presence hydrophobic binding sites. The gradual decrease of this high initial fluorescence may be attributed to the shift of the

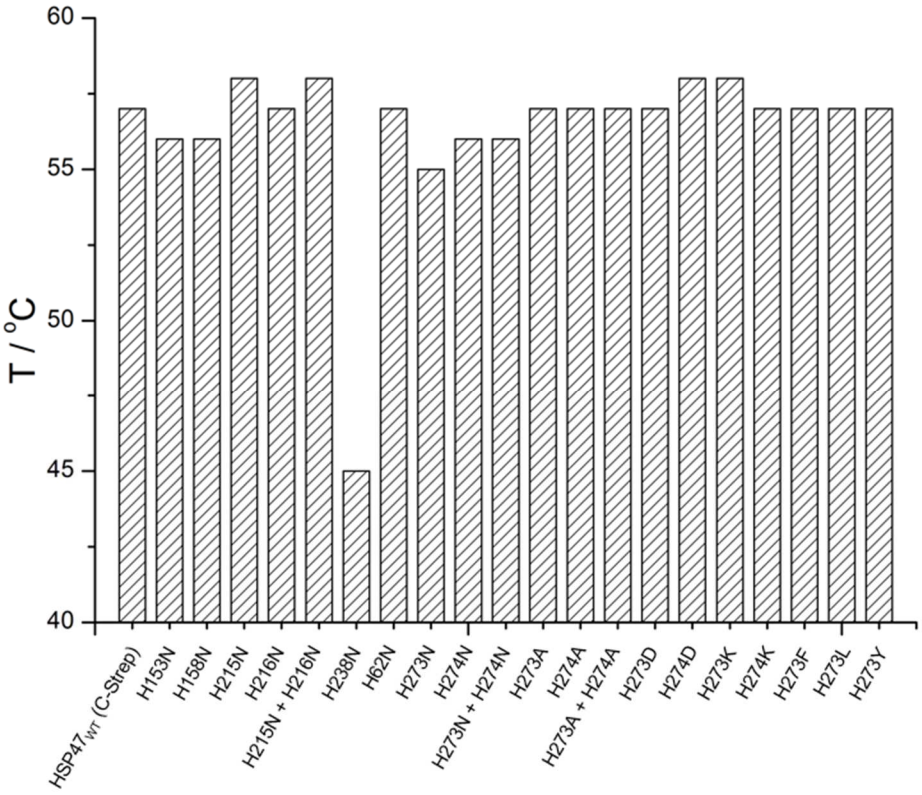
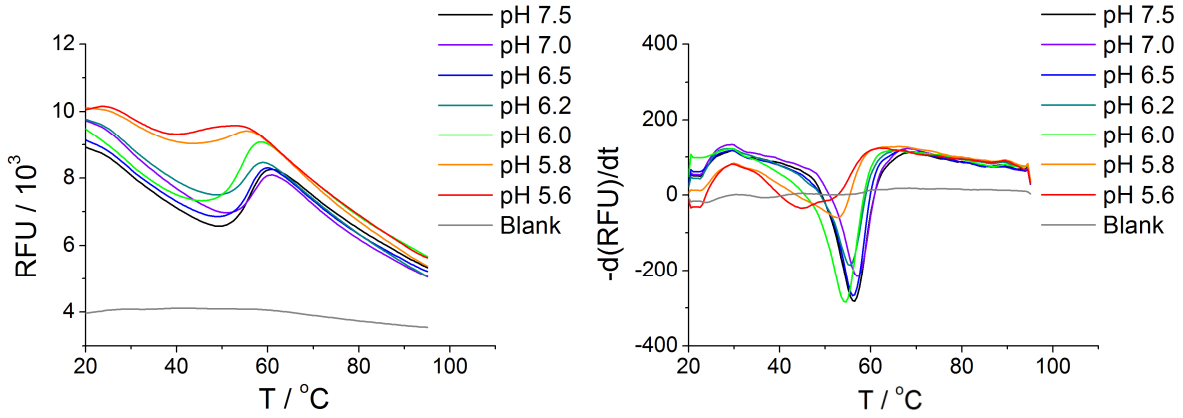


Figure 3.3 | Thermal stability of HSP47 mutants. For the determination of T_m values, DSF was performed using 0.1 mg/ml HSP47 in McIlvaine buffer pH 7.5. All mutants bar H238N show a thermal stability similar to the wild-type; H238N, however, is significantly destabilized and displays a decrease of 12 °C in T_m (45 °C) compared to the average (57 °C).



pH	7.5	7.0	6.5	6.2	6.0	5.8	5.6
T_m / °C	56.5	57.0	56	55.5	54.5	53.0	46.5

Figure 3.4 | pH-dependency of HSP47WT thermal stability. SYPRO Orange fluorescence emission intensity is plotted against temperature for 0.1mg/ml $HSP47_{WT}$ in McIlvaine buffer with the indicated pH ***(top left)***. The T_m corresponds to the inflection point of the first derivative of the thermograms ***(top right)*** and is listed for the indicated pH ***(bottom)***. Blank specifies conditions without protein.

equilibrium constant between bound and unbound dye towards the latter, resulting in increased quenching.

Close inspection of the thermograms at pH 5.6 and 5.8 reveals two peculiarities which could be interpreted such that the protein is already partially or fully denaturing:

(i) The initial fluorescence intensities are typically higher below pH 6.0, suggesting an increase in the total area of exposed hydrophobic surfaces via partial or full unfolding, and

(ii) The increase in fluorescence intensity measured during the denaturation process is less pronounced, implying that the amount of native protein capable of denaturing is lower.

Dynamic Light Scattering

Samples meant for crystallisation experiments were subjected to DLS analysis directly following SEC. A polydispersity higher than 10% has been shown to be detrimental for crystal growth[143]; values below this were unfortunately seldom measured for HSP47 samples. In a typical batch of purified protein, polydispersities ranged from ~15% for factions corresponding to the peak middle to ~35% for peripheral factions. An arbitrary limit of 25% polydispersity was chosen and factions with a higher polydispersity discarded for subsequent crystallization experiments.

3.2 HSP47 Lacks the Serpin-Typical Hyperthermostable Conformation

The expression vector for a HSP47 (C-strep) construct containing the RCL of α1-antitrypsin ($HSP47_{\alpha1AT\text{-}RCL}$) (Fig. 3.5) was provided by Dr. Jan Gebauer. The protein was successfully expressed and purified using the standard protocol for HSP47 (C-Strep). Fig. 3.6 compares the elution profiles of $HSP47_{\alpha1AT\text{-}RCL}$ and wild-type HSP47 in preparative SEC; the larger loop of the mutant leads to a slightly earlier elution. The ratio of the main peak to the characteristic shoulder, or in other words of monomeric HSP47 to the trimeric species, is marginally decreased in case of the mutant. This need not be a meaningful observation, though, since the formation of the latter is sensitive to relatively small differences in the age and concentration of DTT as experienced by the author and reported elsewhere[121].

$HSP47_{WT}$: G^{329}-NPFDQDIYGREELRSP-K^{346}

$HSP47_{\alpha1AT\text{-}RCL}$: G^{329}-TEAAGAIQEEAIPM┆SIPPKL-K^{350}

Figure 3.5 | RCL sequences of wild-type HSP47 and the α1AT-RCL substituted mutant. The dashed line carboxyterminal to the "bait" methionine in the α1AT-RCL represents the scissile bond targeted by Elastase. The native RCL does not contain any cleavage sites for elastase.

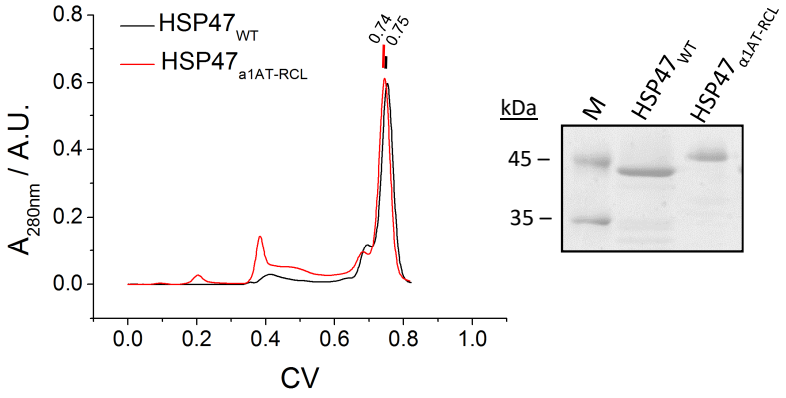


Figure 3.6 | Size-exclusion chromatography and quantitative SDS-PAGE of HSP47$_{WT}$ and HSP47$_{\alpha 1AT\text{-}RCL}$. Comparison of size exclusion profiles of HSP47$_{WT}$ and HSP47$_{\alpha 1AT}$ shows that the mutant elutes marginally earlier than wild-type HSP47 in SEC ***(left)***. Coomassie stained SDS polyacrylamide gel of 2µg protein. The mutant can be seen to run at a higher molecular weight ***(right)***.

The thermostability of the RCL mutant at neutral pH was assessed using DSF. HSP47$_{\alpha 1AT\text{-}RCL}$ was found to have a slightly lower T_m (51.5 °C) compared to the wild-type (54 °C) (Fig 3.7); this, together with the SEC results, suggests that the protein is stably folded. It should be noted that the T_m of wild-type HSP47 as determined here is lower than the T_m reported above (57 °C, see 3.1.4); this stems from the use of different buffer systems (HBS in this case, as opposed to McIlvaine buffer).

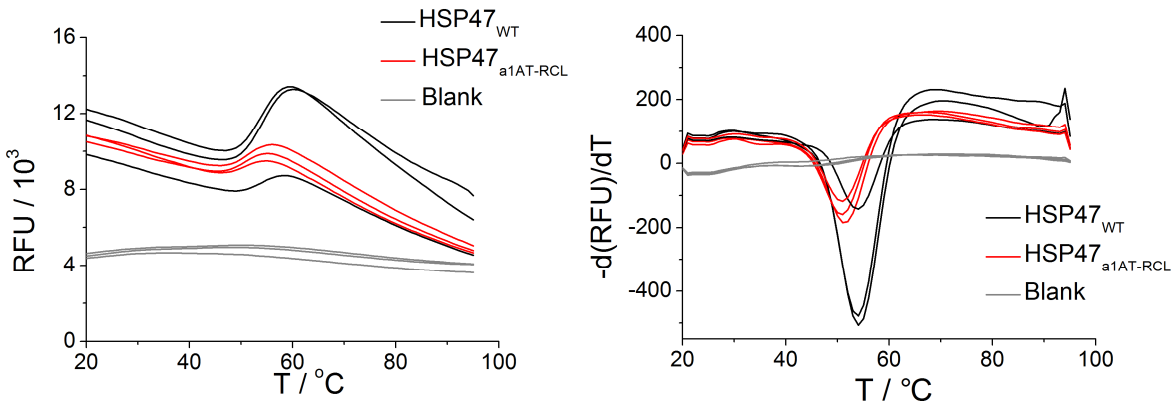


Figure 3.7 | T_m values of wild-type HSP47 and the α1AT-RCL substituted mutant. DSF was used to compare the thermal stability of the mutant with that of the wild-type. ***(Left)*** SYPRO Orange fluorescence intensity is plotted in triplicate against temperature for 0.1mg/ml HSP47 in HBS. ***(Right)*** T_m corresponds to the inflection point of the first derivative of the thermograms. Blank specifies conditions without protein.

Both wild-type HSP47 and the RCL mutant were then subjected to limited proteolysis with porcine pancreatic elastase type I (PPE, Sigma-Aldrich). The protein concentration was adjusted to 0.2 mg/ml in 25 mM NH_4HCO_3 (pH 8.2) buffer and PPE was added in a 1:1 stoichiometry. The mixture was incubated at 37 °C and the reaction monitored via SDS-PAGE. Samples retrieved for SDS-PAGE were immediately mixed with Laemmli-buffer and heated for 5 min at 95 °C in order to denature proteins and stop the reaction. As can be seen in Fig. 3.8, PPE digest of HSP47$_{WT}$ and HSP47$_{\alpha 1AT\text{-}RCL}$ shows differences in progression and products. In the case of wild-type HSP47, a band some 3 kDa smaller

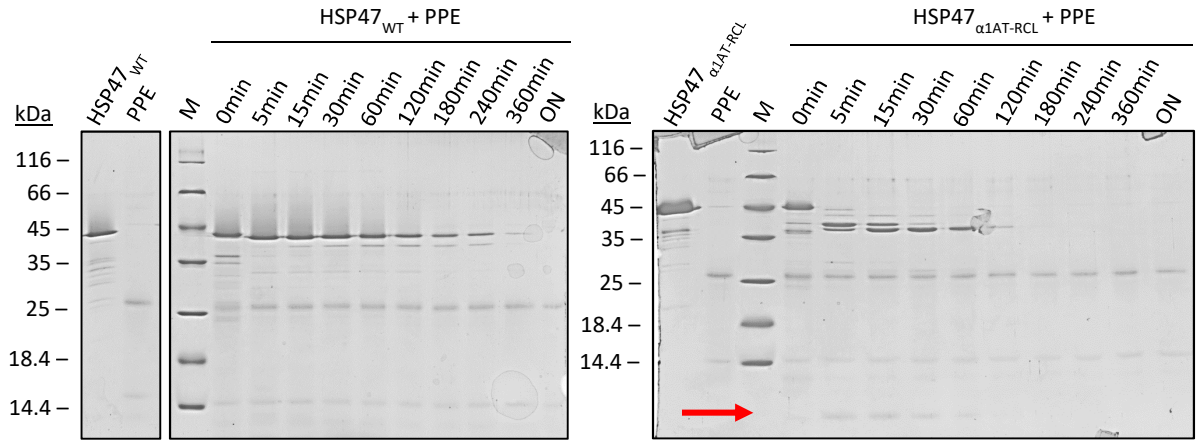


Figure 3.8 | PPE digest of wild-type HSP47 and the α1AT-RCL substituted mutant. Samples were retrieved at the indicated time points and the progress of proteolysis analysed via SDS-PAGE and Coomassie staining. ***(Left)*** Digestion of $HSP47_{WT}$; ***(Right)*** Digestion of $HSP47_{\alpha1AT\text{-}RCL}$. M = Marker. The red arrow highlights a transiently appearing band which might possibly be the cleaved-off C-terminal part of the protein.

appears after around 15 min. A possible cleavage site which would yield this species can be localized in a flexible, solvent exposed loop just C-terminal to α-helix A ($V^{64}E^{65}$). However, HSP47 is not fully converted to this species and both are almost fully degraded after 6h of incubation.

In the case of $HSP47_{\alpha1AT\text{-}RCL}$, almost all of the protein is converted into two species with a molecular weight of ~38 - 40 kDa within 5 min. Concurrently, the appearance of a distinct protein band of small molecular weight (red arrow) can be observed. After one hour of incubation, the double band disappears and solely the lower weight species can be detected, which then is degraded completely after 3h. It can be postulated that the higher weight species of the double band corresponds to $HSP47_{\alpha1AT\text{-}RCL}$ after cleavage in the RCL and removal of the resulting C-terminal fragment (expected molecular weight: 37.9 kDa) and that the lower weight species occurs due to further cleavage elsewhere in the protein. The low molecular weight band indicated with the red arrow can be assigned to the above-mentioned C-terminal fragment generated by cleavage in the RCL (expected molecular weight: 6.7 kDa).

In order to further investigate the biochemical characteristics of $HSP47_{\alpha1AT\text{-}RCL}$, the PPE digestion was up-scaled and repeated. Proteolysis was terminated after 1 h at 37 °C via addition of 0.2 mM PMSF, and the mixture was subjected to analytical size-exclusion chromatography (Fig 3.9). The chromatogram shows that digested $HSP47_{\alpha1AT\text{-}RCL}$ runs at the same volume as the undigested protein, indicating that in the C-terminal fragment generated via cleavage in the RCL is still incorporated into the native fold, presumably because it harbors β-strands s4B, s5B and s1C, which are still embedded into the tertiary structure.

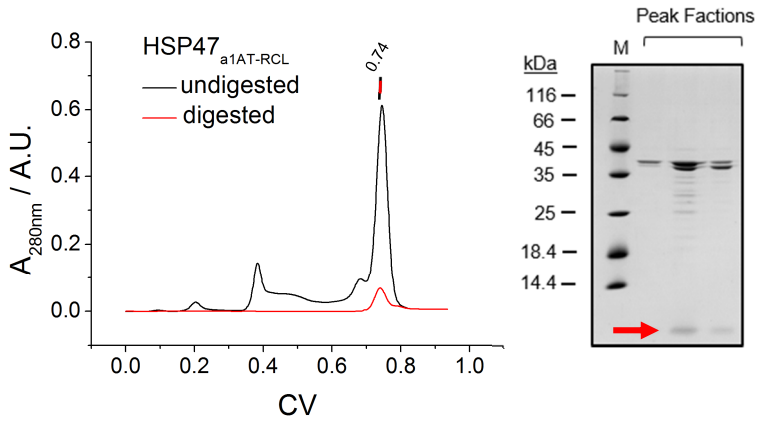
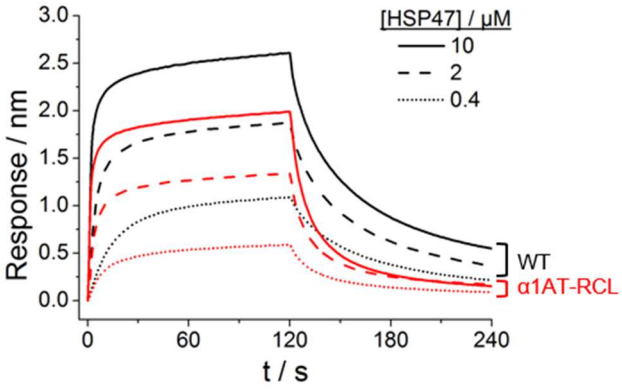


Figure 3.9 | Size-exclusion chromatography of Elastase digested α1AT-RCL mutant. The retention volume of the HSP47$_{\alpha 1AT\text{-}RCL}$ mutant in SEC is not affected by cleavage in the RCL ***(left)***. In SDS-PAGE analysis, the cleaved off fragment (*red arrow*) runs separately ***(right)***.

The influence of the exogenous RCL on the collagen binding properties of HSP47 was investigated using BLI (For details on BLI-Assay establishment please refer to the next section). In comparison to the wild-type, HSP47$_{\alpha 1AT\text{-}RCL}$ was seen to generate a lower maximal response upon association to immobilized CMP-R18 at neutral pH (Fig. 3.10). Extraction of the k_{off} values revealed a two-fold increased dissociation rate for the mutant: the differences in k_{on}, on the other hand, were marginal. In crystal structures of HSP47 in complex with CMP, the RCL is seen to be disordered, suggesting that it is not involved in interactions with the client. However, it is possible that the increased space requirement of the larger α1AT RCL is deleterious for client binding.



	k_{on} / $mM^{-1}s^{-1}$	k_{off} / s^{-1}	K_D / µM
WT	43.03 ± 0.68	0.029 ± 0.001	0.67 ± 0.07
α1AT-RCL	38.46 ± 0.86	0.058 ± 0.001	1.51 ± 0.09

Figure 3.10 | CMP affinity of the α1AT-RCL mutant. ***(Left)*** Comparison of the concentration dependent interaction of HSP47$_{WT}$ (*black)* and HSP47$_{\alpha 1AT\text{-}RCL}$ (*red)* with immobilized CMP-R18 in McIlvaine buffer pH 7.5. ***(Right)*** The k_{on} and k_{off} values were extracted from the sensorgram corresponding to 2 µM protein *(dashed line)*.

3.3 Characterization of the HSP47 - Collagen Interaction

3.3.1 Establishing Biolayer Interferometry Experiments

Initial considerations focused on assigning the role of ligand and analyte to the interaction partners (i.e. whether HSP47 or the CMP should be immobilized onto the sensor) as well as selecting the method of immobilization. Owing to the fact that the obtained signal is directly proportional to the relative change of the density and thickness of the biolayer, it was obvious that loading the smaller CMP onto the sensor and detecting the binding of the larger HSP47 would provide a better signal to noise ratio. As for the method of immobilization, loading His_6-tagged ligand onto Ni-NTA-functionalized sensors or loading biotinylated ligand onto streptavidin-functionalized sensors was considered. While the former foregoes potentially problematic chemical modifications of the ligand with biotinylating reagents, there were concerns that performing experiments in the pH-range of the pKa of histidine sidechains would lead to increased dissociation of ligand from the sensor especially in more acidic conditions, which would require careful control experiments to account for. Biotinylation of CMPs was found to be straight forward, and thus the latter immobilization strategy was chosen.

One problem which arose in using streptavidin sensors was that HSP47 is capable of unspecific binding to the streptavidin coated biosensors. However, this unspecific binding can be supressed via the inclusion of 0.01% (v/v) Tween-20 and/or 0.1% (w/v) BSA in the buffer. Alternatively, unspecific binding sites on the biosensor can be saturated via repeated exposure of the sensor to HSP47 in a blocking step prior to the experiment (Fig 3.11). The latter strategy was chosen for all subsequent biolayer-interferometry experiments.

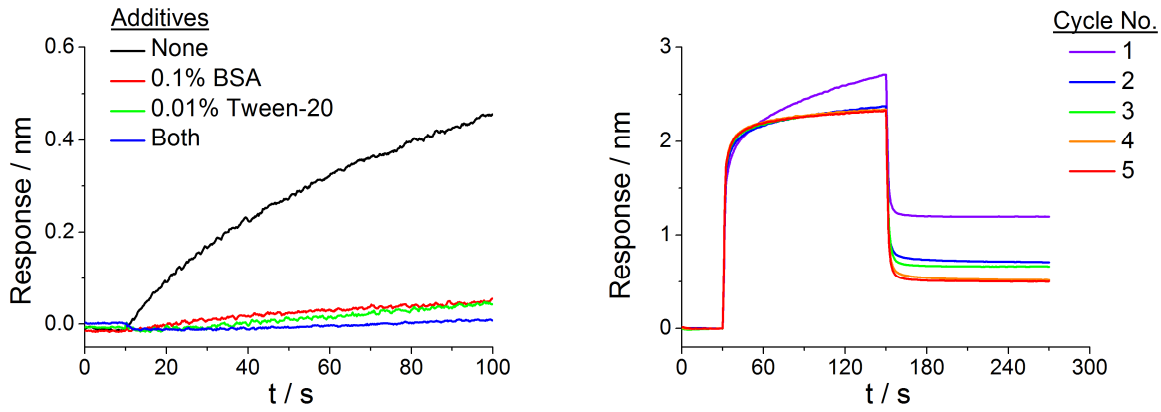


Figure 3.11 | HSP47 shows unspecific binding to SA biosensors. ***(Left)*** HSP47 (1.5 µM) associates to a blank SA biosensor in standard assay buffer *(black)*. This unspecific binding can be largely suppressed with buffer additives such as 0.1% (w/v) BSA *(red)* or 0.01% Tween-20 *(green)* and practically eliminated by simultaneously employing both *(blue)*. ***(Right)*** Repeated association/dissociation cycles of HSP47 (1.5 µM) to a CMP-R18-biotin loaded SA biosensor saturate the unspecific binding sites and give reproducible curves after 3 cycles *(orange and red)*. The unspecific binding is noticeable in the first cycle *(violet)* through the extended time required for the association curve to reach a plateau and the higher residual response after the dissociation, indicating some protein has irreversibly bound to the biosensor.

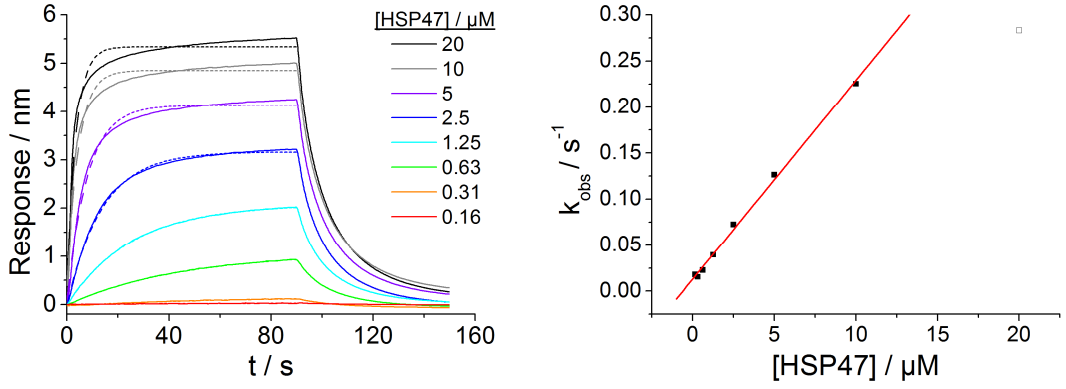
3.3.2 Characterization of the HSP47 - Collagen Interaction Using BLI

Kinetic parameters and K_D of the HSP47 - collagen interaction

The established assay was used to quantify the kinetic parameters as well as the K_D of the HSP47 - CMP-R18 interaction. The binding of a serial dilution of HSP47 (Analyte, A) to immobilized CMP-R18 (Ligand, L) is displayed in Figure 3.12. The observed reaction A + L ↔ AL* follows pseudo first-order kinetics, since under standard assay conditions the analyte concentration [A] is in great excess while the total ligand concentration [L] stays constant throughout. The "true" association rate (k_{on}) as well as the dissociation rate (k_{off}) can thus be determined via graphing the apparent association rate (k_{obs}) against analyte concentration [A] according to the equation of a first-order reaction

$$k_{obs} = k_{on}\,[A] + k_{off} \qquad \textit{(Eq. 5)}$$

The apparent association rates corresponding to different HSP47 concentrations were acquired via fitting a monophasic exponential association function to the association phase of the sensorgrams. The values were then plotted against the HSP47 concentration to yield k_{on} and k_{off} as the slope and y-intercept of the applied linear fit, respectively. The apparent association rate for 20 µM HSP47 was not included in the fit, since the data loses linearity at high concentrations, presumably due to saturation of the biosensor.



k_{on} / $mM^{-1} \cdot s^{-1}$	k_{off} / s^{-1}	K_D / µM
21.6	0.013	0.60

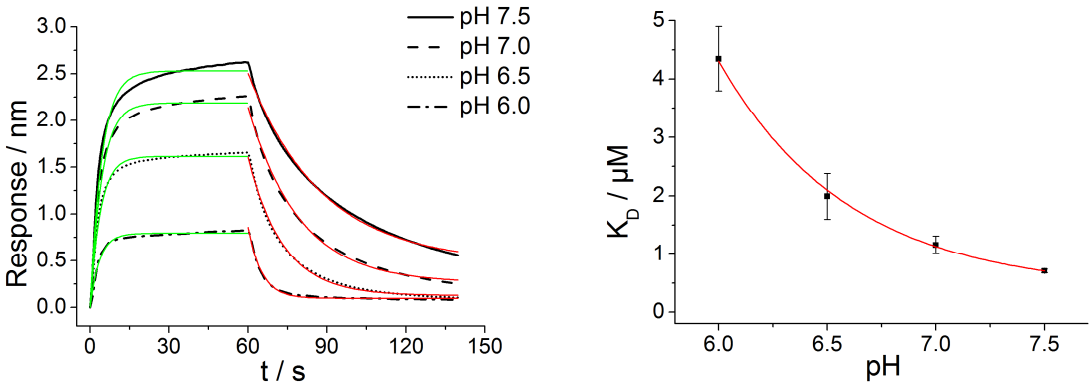
Figure 3.12 | BLI response in dependency on analyte concentration. Binding of a serial dilution of $HSP47_{WT}$ (C-Strep) to a biosensor loaded with biotinylated CMP-R18 ***(left)***. The k_{obs} values were extracted via fitting a monophasic exponential association function *(dashed lines)* to the association curves and graphed against HSP47 concentration ***(right)***. The increase in k_{obs} with concentration loses linearity at higher concentrations; thus the data for 20 µM HSP47 *(hollow point)* was omitted for the linear fit *(red line)*. According to Eq. 5, the association rate and dissociation rate are given by the slope and y-intercept of the line, respectively. The dissociation constant can be calculated via $K_D = k_{off} / k_{on}$ ***(bottom)***.

* It is assumed that HSP47 binds CMP in a 1:1 stoichiometry (please refer to 3.3.3)

It is notable that the association curves never quite reach an equilibrium; this was still the case in experiments in which the association time was extended up to the recommended maximum of 4 minutes (not shown). This limitation is imposed by evaporation of solvent in the small analyte reservoir (4 µl), which increases sample concentration and also changes the optical density of the liquid medium and thus increasingly interferes with the measurement. As discussed under 4.2.1, in case of the HSP47 - collagen complex the expected time to reach equilibrium is well below the limit imposed by the BLI method, indicating that either the unimolecular binding model is erroneous, or that the observed "drift" is due to unspecific binding.

HSP47 client-affinity decreases with pH

The binding of 2.5 µM $HSP47_{WT}$ (C-Strep) to biotinylated CMP-R18 immobilized on a SA-biosensor was monitored in a pH range of 6.0 to 7.5. The kinetic parameters of the individual binding isotherms, k_{on} and k_{off}, were extracted as described under 2.4.11 and allowed the calculation of the K_D, which was subsequently plotted against the pH (Fig. 3.13) Analysis of the experimentally derived parameters leads to the conclusion that the increase in K_D can be largely attributed to an increase in k_{off}.



pH	$R_{max,t=60s}$	k_{on} / $mM^{-1}·s^{-1}$	k_{off} / s^{-1}	K_D / µM
7.5	2.62	39.6 ± 1.1	0.028 ± 0.001	0.71 ± 0.03
7.0	2.26	36.6 ± 1.7	0.042 ± 0.004	1.16 ± 0.15
6.5	1.66	38.5 ± 5.9	0.074 ± 0.004	1.99 ± 0.39
6.0	0.83	22.4 ± 6.8	0.197 ± 0.004	4.49 ± 0.55

Figure 3.13 | pH-dependency of the HSP47 - Collagen interaction visualized with BLI. (***Left)*** A decrease in pH leads to less binding of HSP47 to immobilized CMP-R18-biotin, as indicated by the decrease in maximal response. The k_{on} and k_{off} values were extracted from a monophasic exponential function fitted to the data of the association *(green)* and dissociation *(red)* phases, respectively. ***(Right)*** The calculated K_D values are shown plotted against pH; at pH 6, the K_D has increased more than 6-fold compared to pH 7.5. (***Bottom)*** Summary of experimentally determined parameters for indicated pH values. All sensorgrams were recorded in triplicate.

Looking at the individual kinetic parameters, it is immediately evident that the determination of k_{off} is much more precise than the determination of k_{on}. The higher error observed for the latter is primarily induced by discrepancies between the mathematical fit and the recorded association curve, as can be seen in the left panel of Fig. 3.13. The error is furthermore seen to be more substantial at low pH, where the total signal to background ratio is low and presumably increased denaturation or aggregation of HSP47 occurs.

3.3.3 Stoichiometry of the HSP47 - Collagen Interaction

Triple-helical, homotrimeric collagen can accommodate more than one HSP47 molecule per stagger unit containing the GPR motif - in fact, 2 HSP47 molecules per trimeric CMP is the only stoichiometry ever observed in crystal structures up to date. In order to investigate whether this is also the case in aqueous solution and to verify that the mono-phasic association model used for the evaluation of BLI data is valid, ITC experiments were performed as described under 2.4.9. In experiments conducted in HBS pH 7.5 at 20 °C, the binding was revealed to be endothermic (ΔH = 11.67 kJ·mol^{-1}), indicating that the driving force of the interaction is of entropic origin (ΔS = 159.79 J·K^{-1}·mol^{-1}) (Fig.3.14). The comparison of crystal structures of HSP47 alone and in complex with CMPs reveals that HSP47 does not undergo any significant conformational changes upon client-binding, ruling out any contribution of protein rearrangement to the increase in entropy. It can be assumed that the largest contributing factor is the release of water molecules from clathrates at the relatively hydrophobic interaction interfaces.

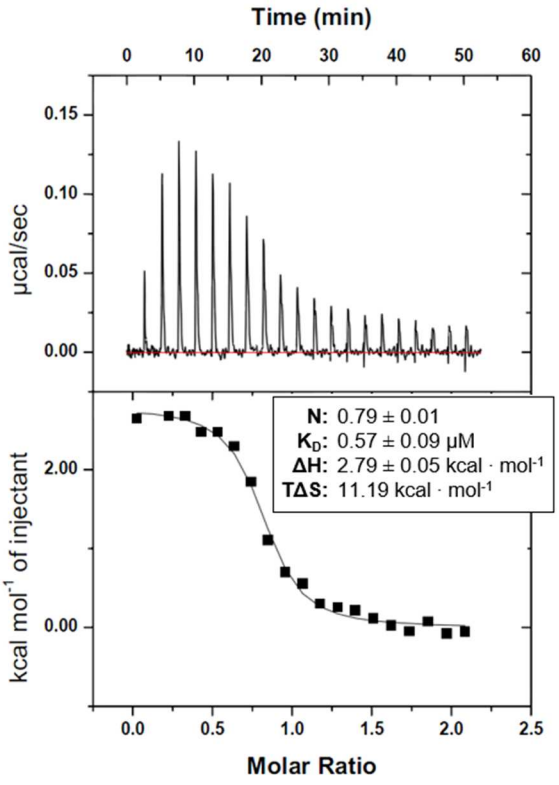


Figure 3.14 | Thermodynamic parameters of the HSP47 - Collagen interaction. (***Top)*** 300µM HSP47 in PBS (pH 7.5) was injected in 2.5µl steps to the reservoir containing 350µl 30µM CMP-R18 in PBS (pH 7.5) at 20°C. The difference in heating power between sample and reference cells was recorded; the positive sign indicates that the association is endothermic**.** (***Bottom)*** The stoichiometry and thermodynamic parameters of the interaction can be extracted via the Wiseman plot, in which the heat taken up per injection is plotted against the ratio between total HSP47 concentration and total CMP concentration.

The calculated K_D value of 0.57 µM is in accordance with the ones obtained via BLI (0.60 µM, 0.71 µM), while the calculated stoichiometry of 0.79 (HSP47 : CMP) can be assumed to deviate from 1 due to inaccurate determination of protein concentration or the presence of binding-incompetent HSP47.

3.4 Molecular details of the pH-dependent client release

3.4.1 Systematic Analysis of Histidine Residues in HSP47

As mentioned under the aims of this thesis, histidine residues are the prime candidates for triggering the pH-induced client release of HSP47. In order to pinpoint the histidine residues involved in this process, the 14 histidines present in HSP47 were subjected to structural and phylogenetic analysis and their protonation behaviour assessed with molecular dynamics simulations, which subsequently guided the selection of mutants to be generated and analysed *in-vitro*.

Phylogenetic analysis

The phylogenetic sequence alignment was prepared by Dr. Jan Gebauer and visualized by the author (Fig. 3.15). Within the class of mammalia, all histidines are conserved; variations are primarily observed in more distant ancestors such as fish, amphibians, reptiles and birds. His108 and His353 are found in other serpins clades as well (PEDF, Antitrypsin, Antithrombin), suggesting that they are a common feature of the serpin fold. His108 is located near the N-terminal end of hE and serves as a H-bond acceptor for Tyr154 on hD, a residue also conserved in other serpins. His353, which is part of β-sheet A and located at the shutter region, is buried in the hydrophobic core of the protein. In orthologs of HSP47, His238, His315 and His386 are seen to be strictly conserved. The imidazole side chain of His315, located at the C-terminus of s6A, is buried in the same hydrophobic region as the side chain of His353, which in the tertiary structure is immediately adjacent to it. His238 and His386 are constituents of β-sheet C, the scaffold of the interaction interface. The conservation and immediate vicinity to bound collagen suggests a crucial role for these residues in client binding and release. His153, His158, His209, His215, His274 and His320 are largely conserved, with few exceptions found in Actinopterygii, while in the case of His216, His262 and His273, there are substantial variations.

The localization of histidine residues in the three dimensional structure of HSP47 is displayed in Fig. 3.16.

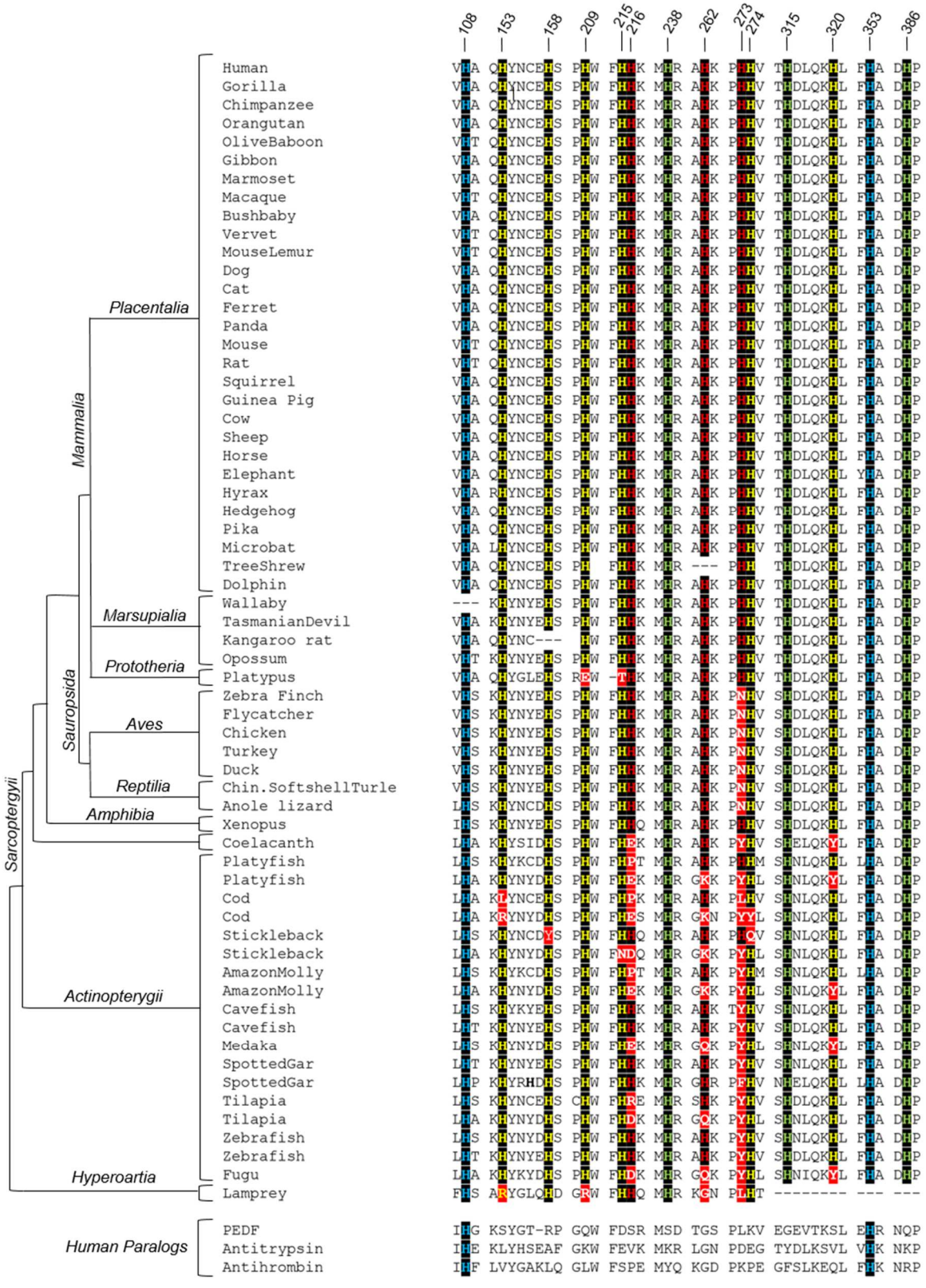


Figure 3.15 | Phylogenetic analysis of histidine residues in HSP47. Histidines are highlighted with black and color coded according to their degree of conservation: strictly conserved in HSP47 as well as in the listed Serpins *(blue)*, strictly conserved only in HSP47 *(green)*, conserved with few exceptions *(yellow)*, least degree of conservation *(red)*. The amino acids replacing histidines are highlighted with red.

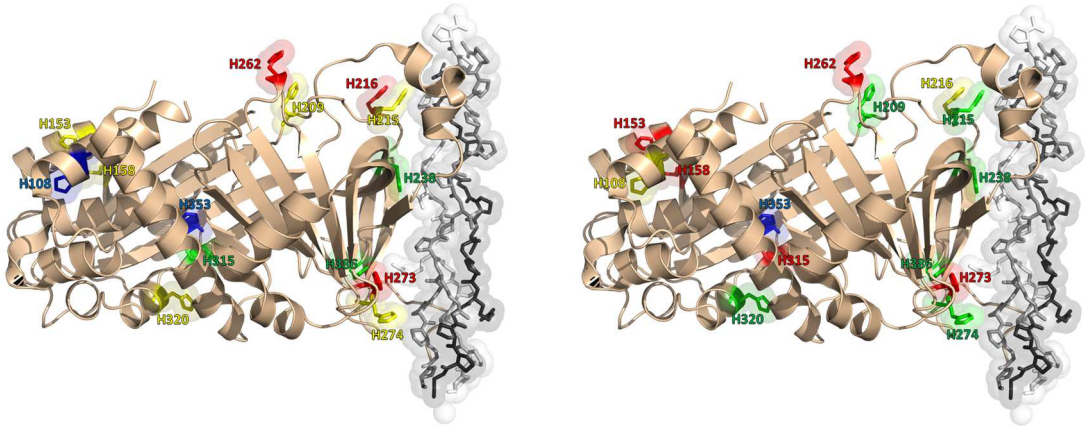


Figure 3.16 | Conservation and basicity of histidine residues in HSP47. Crystal structure of HSP47 with CMP (18)T8R11 (*white:* leading strand, *grey*: middle strand, *black:* trailing strand). ***(Left)*** The 14 histidine residues of HSP47 are color-coded according to their degree of conservation. *Blue:* Strictly conserved in HSP47 and Serpins, *green:* Strictly conserved only in HSP47, *yellow:* Conserved in HSP47 with few exceptions, *red:* Least degree of conservation. ***(Right)*** The same residues are color-coded according to the calculated percentile change in the doubly protonated faction during the transition from pH 7.0 to 6.0. *Blue:* < 0% *green:* ≈ 0%, *yellow:* < 25%, *red:* > 25% (PDB ID: 3ZHA)

Molecular dynamics simulations

The *in-silico* experiments were performed by Eileen Socher under supervision of Prof. Dr. Heinrich Sticht at the University of Erlangen. The crystal structure of HSP47 in complex with a 18 residues long CMP harbouring an arginine per α-chain at position Yaa (PDB entry code: 3ZHA) served as the template for constant pH molecular dynamics (CpHMD) simulations run at a temperature of 37 °C for pH 4.0, 5.0, 6.0, 7.0 and 8.0. For each of the 14 histidines, the fraction of doubly protonated residues was determined at the indicated pH and enabled the calculation of the pK_a values (table 3.2). It is notable that His215, His238 and His274 remain singly protonated over the whole tested pH range. While the influence of the microenvironment on amino acid basicity is hard to predict, this decrease can possibly be attributed to the proximal presence of positively charged residues (Lys213 + Lys217, Arg222 and Lys252, respectively). His386, on the other hand, is already doubly protonated at pH values as high as 8, possibly due to the proximity to the carboxylate moiety of the neighboring Asp385 as well as the H-bond formed with the backbone carbonyl of Pro272. The histidines located at positions 153, 158, 262, 273, and 315 show a significant increase in the doubly protonated fraction (>25%) going from pH 7.0 to 6.0; these were judged to be potential triggers for client release and were investigated further along with all other interface histidines (His215, His216, His238, His274 and His386). Intriguingly, His353 was calculated to show a decrease in the doubly protonated faction under more acidic conditions. This

Table 3.2 | Basicity of histidine side chains in HSP47 as predicted by CpHMD. Asterisks indicate that the histidine is located at the collagen binding interface. The residues in bold were selected for further *in-vitro* studies.

Residue No.	% change in doubly protonated fraction (pH 7.0 → pH 6.0)	Theoretical pK_a
108	+ 9.1	5.59
153	+ 60.9	6.21
158	+ 31.1	5.50
209	+ 0.7	5.08
215*	- 0.1	< 4
216*	+ 15.3	5.40
238*	+ 0.3	< 4
262	+ 63.7	6.51
273*	+ 29.4	5.23
274*	+ 0.4	< 4
315	+ 58.2	5.63
320	- 1.7	4.47
353	- 19.4	5.18
386*	- 0.1	> 8

behaviour is hard to explain without pronounced alterations of the microenvironment, which in itself seems rather unlikely, considering the histidine in question is buried in the core of the protein

3.4.2 Role of HSP47 Histidines in pH-Dependent Client Release

Histidine Scanning Mutagenesis

HSP47 (C-Strep) constructs with a single His → Asn mutation at positions 153, 158, 215, 216, 262, 273, 274, 315 and 386 as well as tandem His → Asn mutations at positions 215 + 216 and 273 + 274 were generated as described under 2.2.4. The rationale for choosing asparagine as a substitute was that its side chain, among all natural amino acids, displays the largest structural and stereoelectronic similarity to the imidazole moiety of histidines and does not undergo protonation at the relevant pH. Following initial experiments, where the H273N + H274N double mutant was identified to have attenuated response to decreasing pH (*vide infra*), His → Ala mutations were introduced at these positions with the aim of excluding asparagine-specific effects. In order to simulate the protonated form of the histidines and investigate the overall impact of charge, lysine and aspartic acid were also

incorporated at said positions. Finally, His273 was substituted by the amino acids encountered in Actinoptergyii and Hyperoartia (Leu, Phe and Tyr) in order to further understand the influence of the amino acid at this position on client-release. The generated mutants were tested for their affinity to CMP-R18 at pH 6.0, 6.5, 7.0 and 7.5 using biolayer interferometry as described under 2.4.11. The sensorgrams and extracted parameters k_{off}, k_{on} and K_D are shown in the appendix (Fig. A.1, tables A.1-3). For reasons that will be discussed under section 4.2.1, determination of k_{on} and thus also the calculation of K_D was accompanied with substantial error, and the interpretation of the data will be primarily through the dissociation rate k_{off}.

Non-interface histidines are not involved in client-release

According to MD simulations, the non-interface residues His153, His158, His262 and His315 are subject to the largest increase in the fraction of doubly protonated side chains during the transition from pH 7 to 6. This suggests that these residues are potential triggers, ostensibly relaying information about their protonation states to the collagen binding interface via conformational rearrangements. However, BLI analysis of H153N and H158N shows that the collagen binding behaviour of the mutants is not altered compared to wild-type, while the in case of H262N the interaction with the CMP is destabilized at increasingly acidic conditions (Fig. 3.17). The mutant H315N failed to express, which is unsurprising considering that the histidine is located in the protein core and strictly conserved in the HSP47 family.

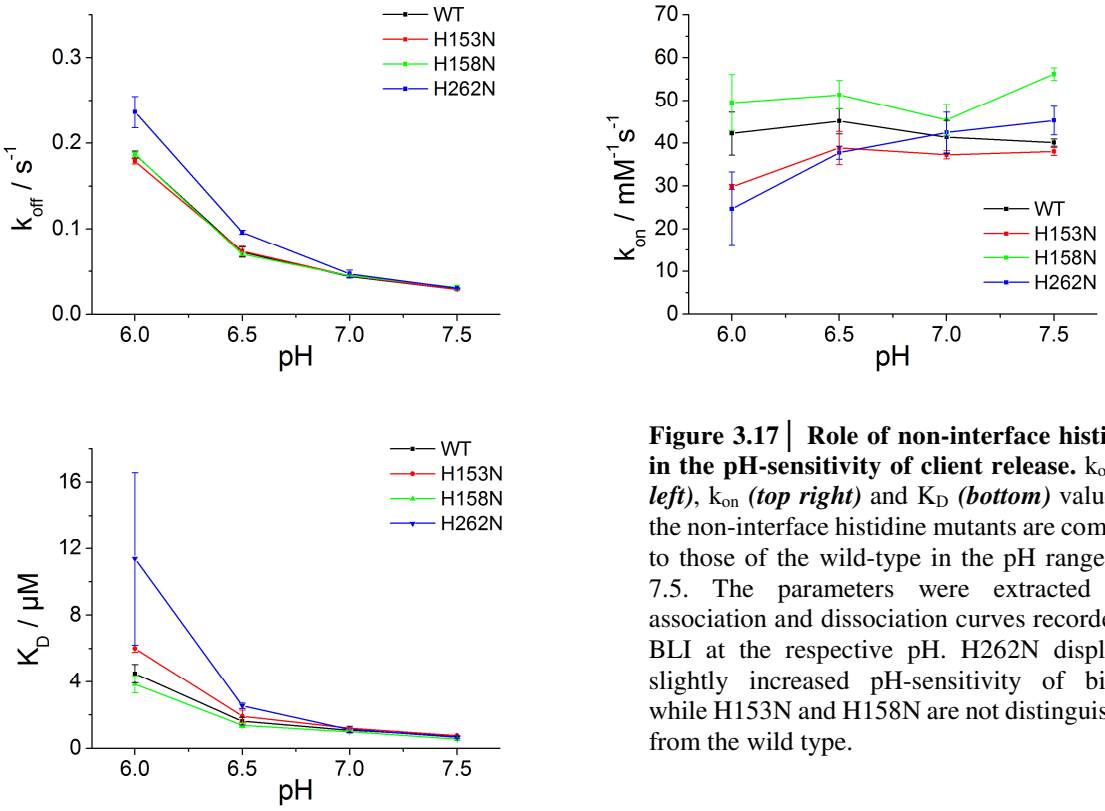


Figure 3.17 | Role of non-interface histidines in the pH-sensitivity of client release. k_{off} ***(top left)***, k_{on} ***(top right)*** and K_D ***(bottom)*** values for the non-interface histidine mutants are compared to those of the wild-type in the pH range 6.0 - 7.5. The parameters were extracted from association and dissociation curves recorded via BLI at the respective pH. H262N displays a slightly increased pH-sensitivity of binding while H153N and H158N are not distinguishable from the wild type.

His215 is important for collagen binding

Histidines 215 and 216 are located at the fringe of the binding interface, with the side-chain of His215 being situated in close proximity (roughly 3.9 Å) to the Yaa^{-3} residue in the CMP strand with a -1 stagger compared to the one harbouring the arginine recognized by HSP47. The calculated pKa of His215 is abnormally low (<4), making it unlikely that this residue is one of the pH triggers, since the increase in the doubly protonated faction as the pH approaches 6 can be expected to be negligible. As stated before, this rather drastic decrease in the pKa might be attributed to the proximity of two lysine residues (Lys213 and Lys217), although why this effect is attenuated in case of the neighbouring His216 is unclear.

Of all the tested His → Asn mutants, H215N and the double mutant H215N + H216N were the only ones to display significantly increased k_{off} (Fig. 3.18) and K_D values as well as a decreased R_{max} at all tested pH values. Since H216N is seen to behave similar to the wild-type, the effect observed in the double mutant can be attributed solely to His215, implying that this residue is important for the HSP47 - collagen interaction.

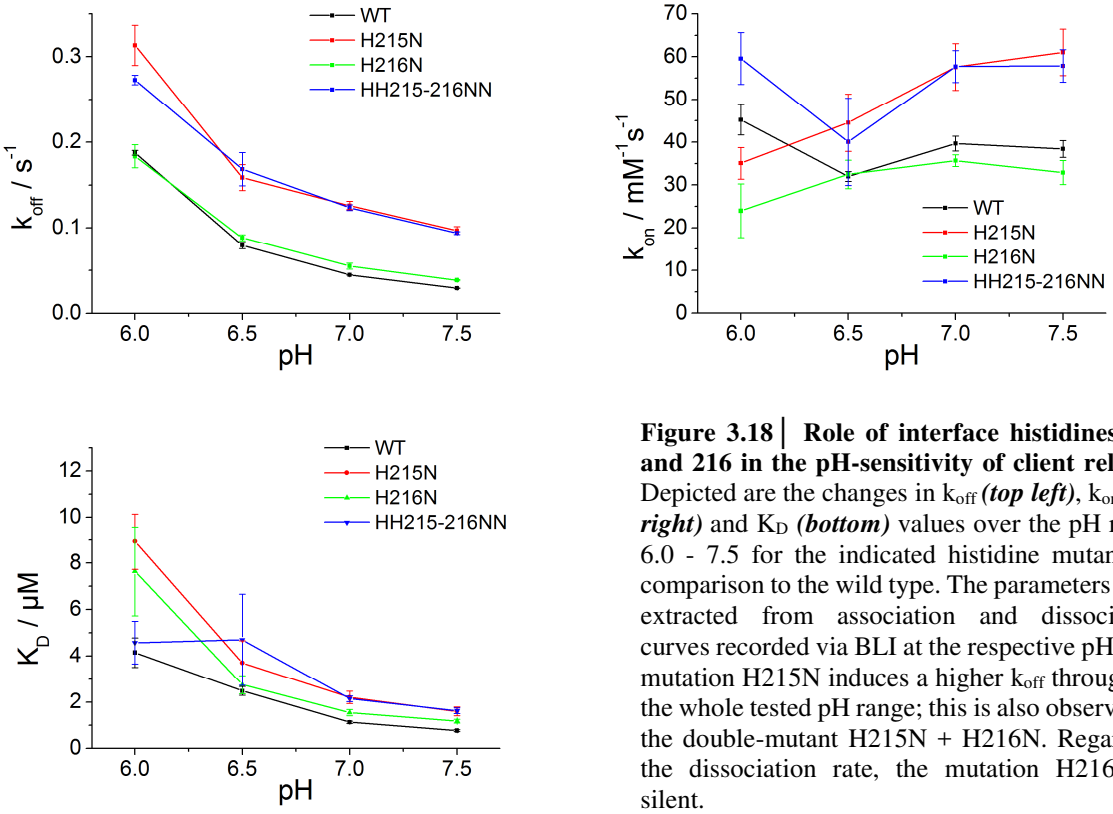


Figure 3.18 | Role of interface histidines 215 and 216 in the pH-sensitivity of client release. Depicted are the changes in k_{off} ***(top left)***, k_{on} ***(top right)*** and K_D ***(bottom)*** values over the pH range 6.0 - 7.5 for the indicated histidine mutants in comparison to the wild type. The parameters were extracted from association and dissociation curves recorded via BLI at the respective pH. The mutation H215N induces a higher k_{off} throughout the whole tested pH range; this is also observed in the double-mutant H215N + H216N. Regarding the dissociation rate, the mutation H216N is silent.

The mutation H238N abolishes collagen binding

His238, located on s1C close to the centre of the binding interface, is one of the three histidines that are strictly conserved in the HSP47 family but not in other Serpins. Although the H238N mutant expressed well, the protein was observed to be rather unstable and prone to precipitation over time, particularly at mildly acidic pH. Strikingly, this single mutation lowers the T_m of the protein by up to 12 °C (please refer to section 3.1.3). In order to verify the structural integrity of the mutant, far-UV CD spectra were recorded, which indicated that at neutral pH, the mutant is stably folded and similar in secondary structure composition to the wild-type (Fig. 3.19, *left*). The binding of $HSP47_{H238N}$ to CMP-R18, however, was not detectable via BLI, even at a pH of 7.5 (Fig. 3.19, *right*).

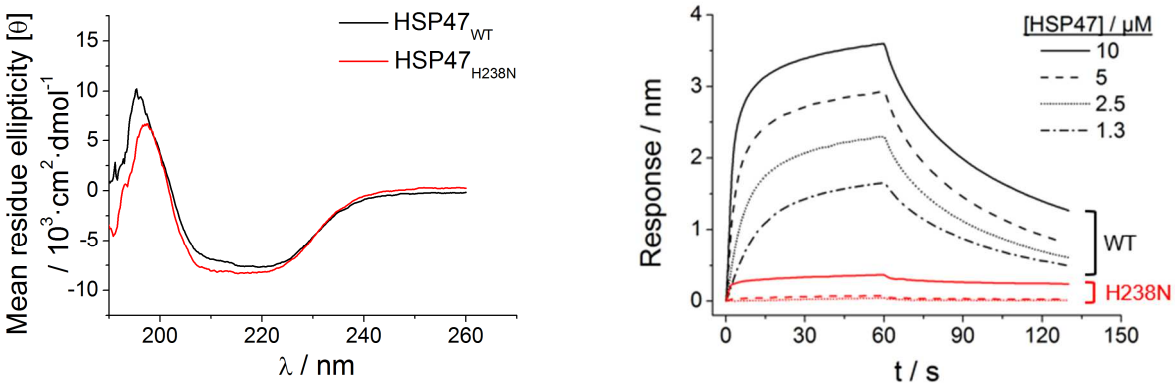


Figure 3.19 | The H238N mutation abolishes CMP binding. **(*Left)*** Comparison of CD spectra recorded at pH 7.5 for wild-type HSP47 as well as the H238N mutant confirms that the latter is stably folded. **(*Right)*** Serial dilutions of wild-type *(black curves)* as well as $HSP47_{H238N}$ *(red curves)* were tested towards their affinity to immobilized CMP-R18 using BLI. Binding of $HSP47_{H238N}$ was not detected.

This finding was further investigated via MD simulations. Figure 3.20 displays a comparison of the wild-type crystal structure with a snapshot of $HSP47_{H238N}$ modelled without the collagen client. It can be seen that the side chain of Arg222, a critical residue for collagen binding, is in the wild-type prepositioned by a network of H-bonds involving Asp220, His238, Ser305 and Tyr383. His238 is wedged in between Asp220 and Ser305, serving as a H-donor to the former via its *tele*-nitrogen atom and as a H-acceptor to the latter via the *pros*-nitrogen atom. In the H238N mutant, however, Asp220 cannot maintain the salt bridge to Arg222, which instead displaces the phenol side chain of Tyr383 to interact with Asp385. This short circuits the two residues crucial for client binding, resulting a total loss of affinity towards collagen.

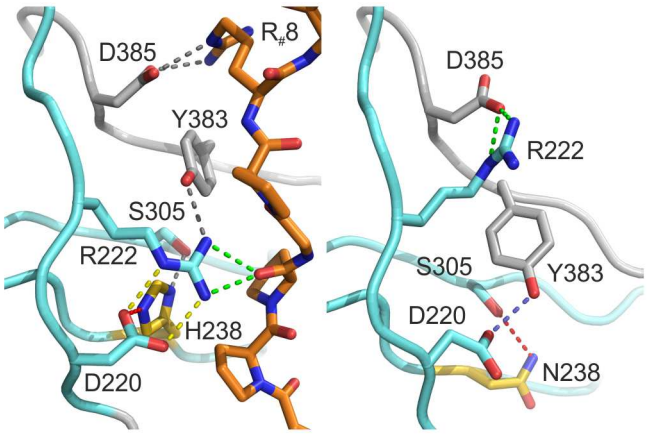


Figure 3.20 | The H-bond network centered on R222 is reorganized in the H238N mutant. ***(Left)*** Crystal structure of wild-type HSP47 in complex with CMP (18)T8R11 (PDB ID: 3ZHA). ***(Right)*** Snapshot of uncomplexed $HSP47_{H238N}$ obtained via CpHMD simulations. In the wild type, H238 interacts with D220 *(red dashed lines)* which in turn forms a salt-bridge with R222 *(yellow dashed lines)*. Additionally stabilized by a H-bond with Y383 *(grey dashed line),* R222 is thus oriented such that it can interact with a backbone carbonyl of the collagen chain. This interaction is essential for collagen recognition, along with the salt-bridge between D385 and R8 *(grey dashed line)*. The mutation H238N leads to a rearrangement of the residues, resulting in an intramolecular salt-bridge between D385 and R222 *(green dashed lines)*, disabling two critical contacts for collagen binding at once. (Figure taken from Oecal *et al.*[184])

In additional MD simulations, distances between key amino acid side-chains were monitored over time as an indicator for the stability of the interaction (Fig. 3.21). In the wild-type, the distance between Arg222-Cζ and Asp220-Cγ is locked at around 4 Å independent of collagen binding, indicating a preorganization of the arginine side-chain. In case of the H238N mutant, however, the determined interatomic distance starts to fluctuate, and a conformer where the indicated atoms are roughly twice as far apart becomes predominant, suggesting a dynamic equilibrium between the binding-competent and short-circuited forms. The same pattern is reflected in the distance between Arg222-Cζ and Asp385-Cγ: the stable separation by approximately 7 Å yields to a dynamic equilibrium in $HSP47_{H238N}$, with the predominant conformer placing the indicated atoms roughly 4.5 Å of each other. Over a large period of the simulated timespan, the distance between the hydroxyl-O of Ser305 and the *tele*-N of His238 is too large for H-bonding, and client binding noticeably stabilizes the transiently forming H-bonds. A specific effect of the H238N mutation is the rather static distance of 2 Å between the hydroxyl-O of Ser305 and the carboxamide-N of Asn238, indicating a stable interaction between these residues.

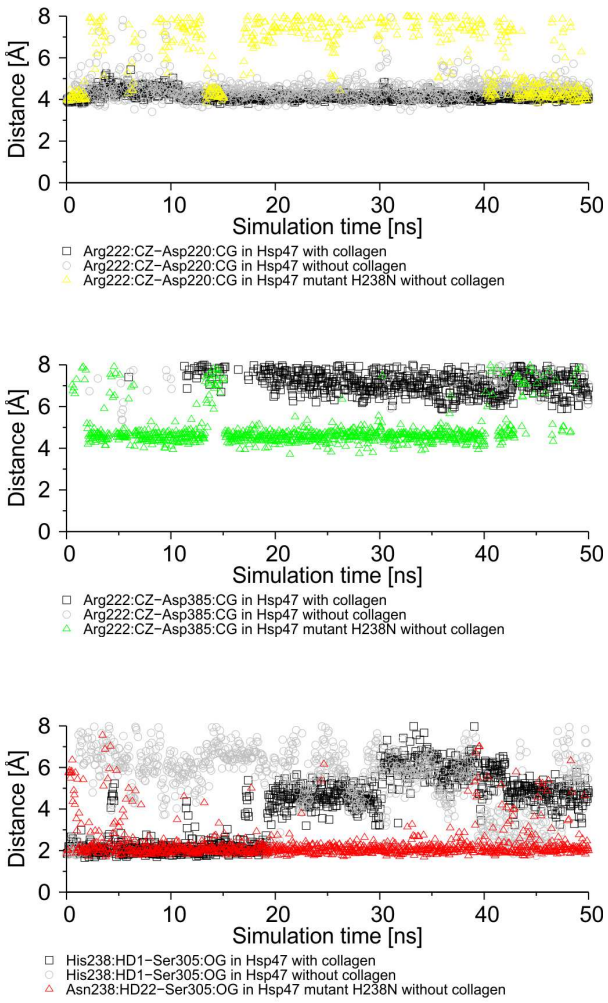


Figure 3.21 | Conformational stability of selected side-chain interactions. The distances between indicated atoms were monitored over simulation time for collagen-bound $HSP47_{WT}$ *(black squares)*, free $HSP47_{WT}$ *(grey circles)* and free $HSP47_{H238N}$ *(colored triangles)*. The interaction between R222 and D220 is stable and not influenced by the presence of collagen in the wild type, but disrupted in the mutant ***(top)***. The formation of the intramolecular salt bridge between R222 and D385 in the mutant is reflected by the decreased interatomic distances ***(middle)***. Client binding slightly stabilizes the H-bond between H238 and S305; the H-bond between N238 and S305 in the mutant, however, is stable regardless of the presence of collagen ***(bottom)***. (Figure taken from Oecal *et al.*[184])

The collagen binding of mutant H273N + H274N is less sensitive to a decrease in pH

His273 and His274 are located at the edge of the binding interface, close to the salt bridge between D385 of HSP47 and the arginine of the collagen. While His274 is rather well conserved across the HSP47 family, His273 is often found to be substituted with asparagine (Sauropsida) or tyrosine (Actinoptergyii) as well as with phenylalanine or leucine in individual cases. According to MD simulations, the fraction of doubly protonated His273 residues increases from 3.3% to 32.7% upon reduction of the pH from 7 to 6, while His274 is virtually not protonated. This difference in basicity is dictated by the microenvironment: the imidazolium cation is stabilized in case of His273 through the H-bond formed with Asp247, and destabilized in case of His274 by the close proximity to the positively charged Lys252. Interestingly, while singly replacing either of the histidines with asparagine did not discernibly influence the HSP47 - CMP interaction, concurrent substitution of both histidines led to a markedly attenuated increase of the k_{off} with increasing acidity (Fig. 3.22). This would imply that protonation of either one of the histidines is sufficient to induce client release. The experiments were repeated using His → Ala mutants; these behaved in the same pattern, albeit with the inconspicuous but

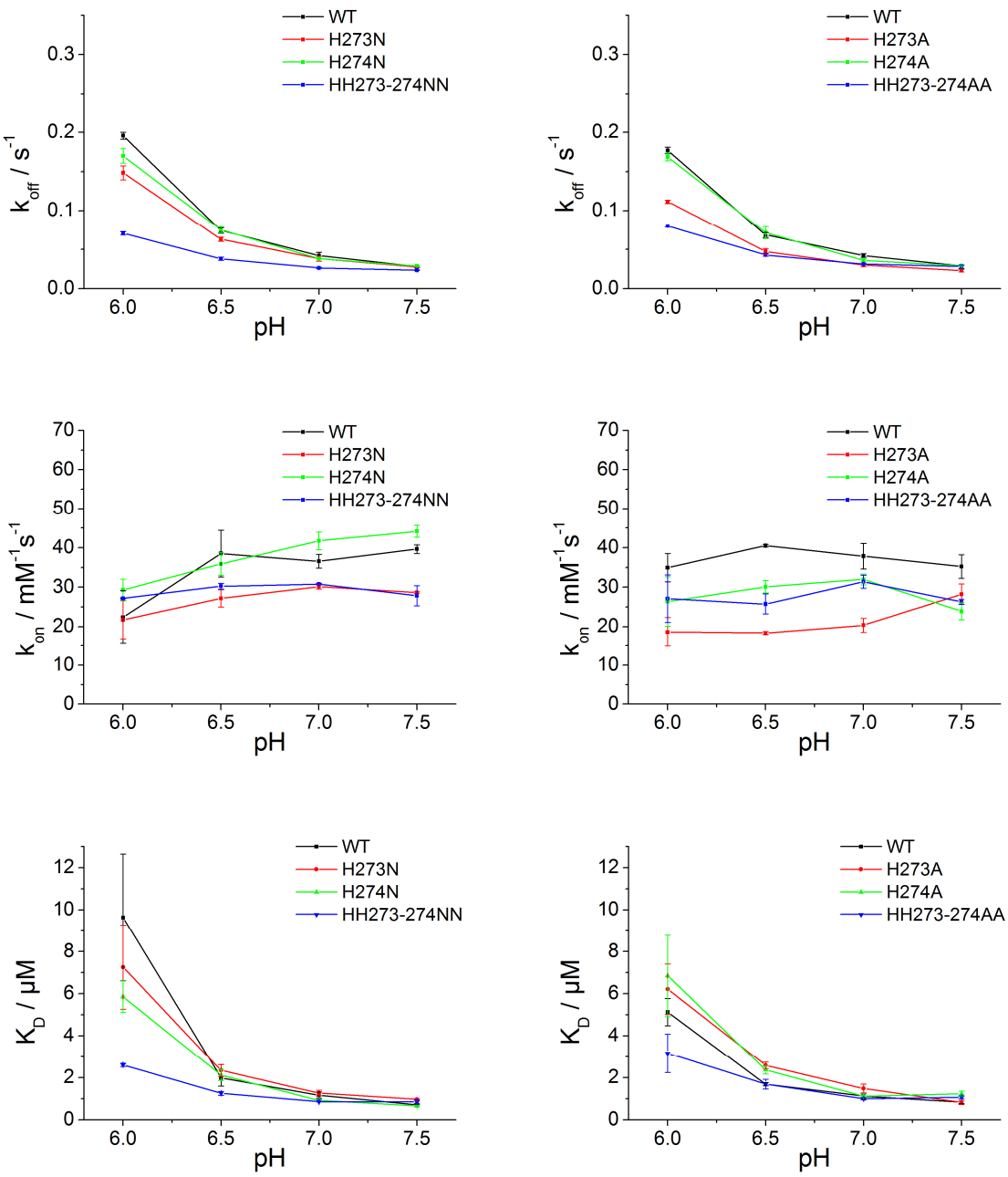


Figure 3.22 | Role of interface histidines 273 and 274 in the pH-sensitivity of client release. | Histidines 273 and 274 were substituted with asparagine ***(left column)*** and alanine ***(right column)*** and the pH-dependency of their dissociation rate ***(top row)***, association rate ***(middle row)*** and dissociation constant ***(bottom row)*** determined via BLI. While mutation of His274 did not affect binding behaviour, the increase in koff with lower pH is attenuated in case of His273 mutants as well as the double mutants. The impact of the mutation increases in the order H273N < H273A < H273N + H274N ≈ H273A + H274A.

nonetheless important detail that H273A responds less to a change in pH than H273N, which in turn is slightly less pH-sensitive than the wild-type.

Influence of charged residues at positions 273 and 274

The observation that a positive charge, introduced via protonation at either position 273 or 274, is required to drive client release prompted further studies using amino acids that are uniformly charged throughout the tested pH range. For this, both histidines were replaced by either aspartic acid or lysine. Concurrent with the idea that a positive charge is required for inducing client-release, H273D and H274D were seen to possess slower dissociation rates, especially under the more acidic conditions (Fig. 3.23). Although the anionic aspartic acid will increase the basicity of the neighboring histidine, the positive charge brought in via protonation will be locked in place and neutralized, resulting in impairment of client release. Following the same line of thought, incorporation of lysine residues at these sites was expected to result in an increased k_{off} and thus a decreased K_D at neutral pH values. Furthermore, an attenuated or completely abolished pH-dependency of the binding was anticipated. However, experimental data revealed that H273K has similar dissociation rates as the wild-type, even displaying a slight decrease in the pH-dependency of its binding. This decrease could stem from the detrimental influence of the positively charged lysine on the protonation of His274, which in this case would seem to be the requirement for client release. The mutant H274K, on the other hand, does display a markedly weakened binding compared to the wild-type, with higher k_{off} rates over the whole tested pH range. Whether this is the direct result of the positive charge or rather of the increased steric demand of the lysine side-chain is unclear.

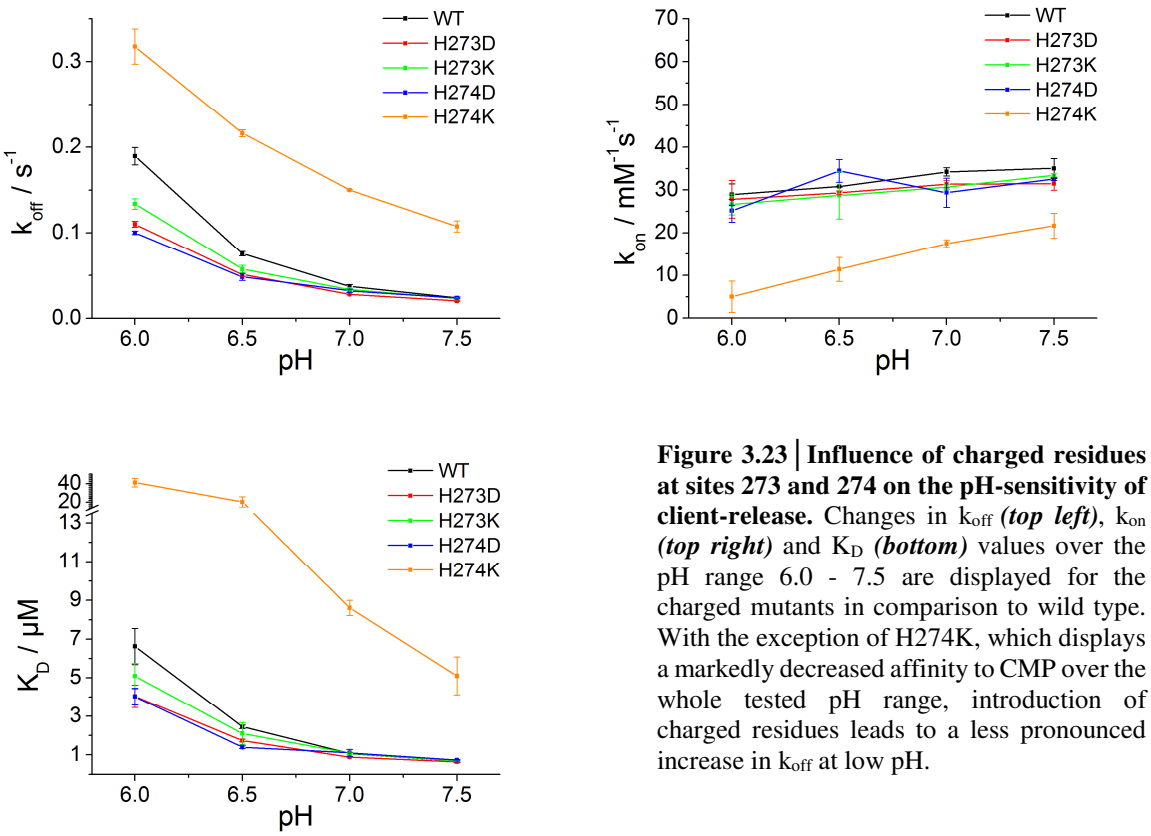


Figure 3.23 | Influence of charged residues at sites 273 and 274 on the pH-sensitivity of client-release. Changes in k_{off} ***(top left)***, k_{on} ***(top right)*** and K_D ***(bottom)*** values over the pH range 6.0 - 7.5 are displayed for the charged mutants in comparison to wild type. With the exception of H274K, which displays a markedly decreased affinity to CMP over the whole tested pH range, introduction of charged residues leads to a less pronounced increase in k_{off} at low pH.

The residue in position 273 fine-tunes pH-dependency of client-release in many organisms

As previously mentioned, mutation of His274 did not have any significant impact on the pH-dependency of client release. In case of H273A and H273N, however, the response of the dissociation rate to more acidic conditions was slightly attenuated and furthermore discernibly different between the mutants. While it is certainly unexpected that substitution of the less well conserved histidine causes a larger perturbation on client-release, the amino acid in position 273 might be involved in the fine-tuning of the basicity of His274 in SERPINH1 orthologs. Coincidentally, initial experiments had already covered the NH motif found in Sauropsida; these were supplemented with analysis of the motif YH, common to Actinopterygii, as well as the infrequently encountered motifs LH and FH. (Fig. 3.24). While the binding behaviour of the mutant with the YH motif was shown to be indistinguishable from that of the HH motif, the pH-sensitivity of FH and LH was shown to be less pronounced, similar to the previously tested NH and AH motifs. Combining all the acquired data, dissociation rates at pH 6.0 were observed to decrease in the order YH (0.215 s^{-1}) ≈ HH (0.192 s^{-1}) > NH (0.148 s^{-1}) ≈ FH (0.139 s^{-1}) > LH (0.112 s^{-1}) ≈ AH (0.111 s^{-1}) > AA (0.080 s^{-1}) ≈ NN (0.071 s^{-1}). If there is any pattern at all, it would seem to suggest that that the basicity of H274 decreases the smaller and less polar the neighbouring residue is.

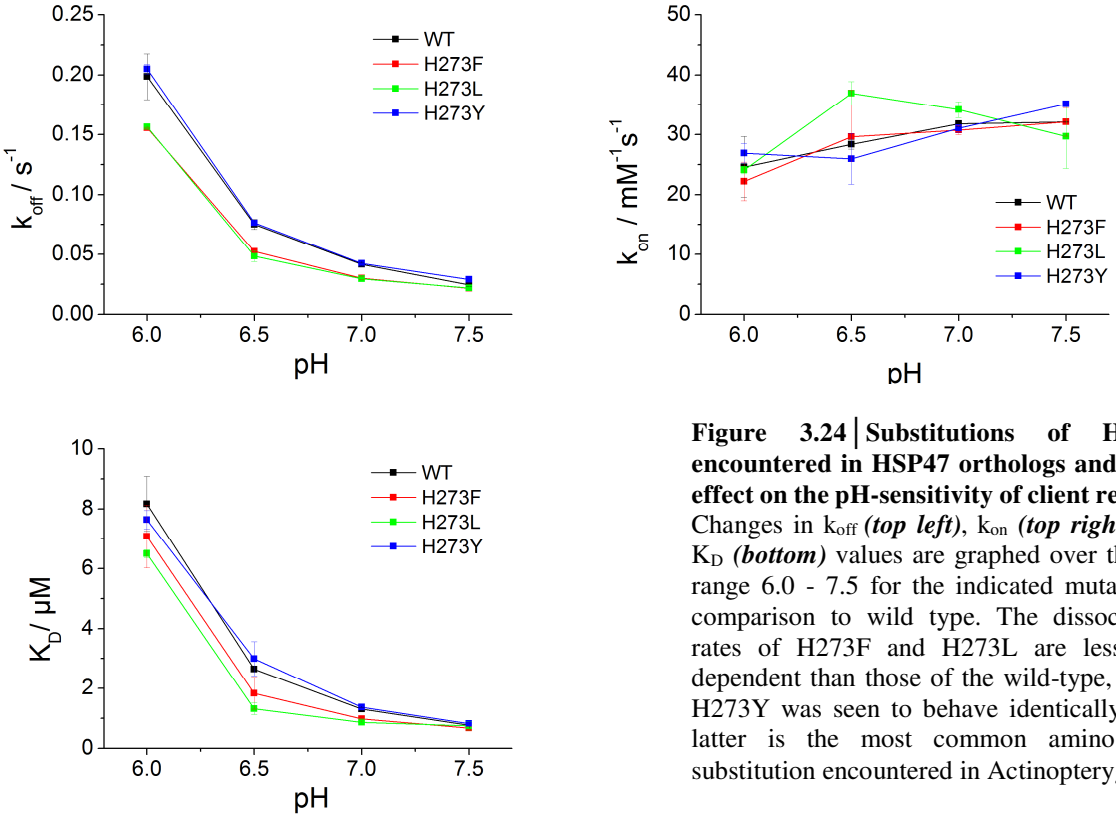


Figure 3.24 | Substitutions of His273 encountered in HSP47 orthologs and their effect on the pH-sensitivity of client release. Changes in k_{off} ***(top left)***, k_{on} ***(top right)*** and K_D ***(bottom)*** values are graphed over the pH range 6.0 - 7.5 for the indicated mutants in comparison to wild type. The dissociation rates of H273F and H273L are less pH-dependent than those of the wild-type, while H273Y was seen to behave identically. The latter is the most common amino acid substitution encountered in Actinopterygii.

There is no evidence for conformational changes accompanying client-release

In order to determine whether HSP47 undergoes any pH-induced conformational changes that might be responsible for client-release, far-UV CD spectra of the wild-type as well as the pH-insensitive double mutant H273N + H274N were recorded at pH 7.5 and pH 6.0. The spectra of both proteins display features characteristic for folded proteins with high α-helical content: a broad saddle of negative ellipticity above 200 nm, with minima observed at 214 nm and 222 nm (Fig. 3.25). The difference in the magnitude of ellipticity between both proteins can be attributed to errors in concentration determination, since the overall shape of the curves is identical. The experiment further revealed that there are no significant differences in secondary structure between pH 7.5 and 6.0 and suggests that conformational arrangements have no effect on the client-binding capacity of HSP47.

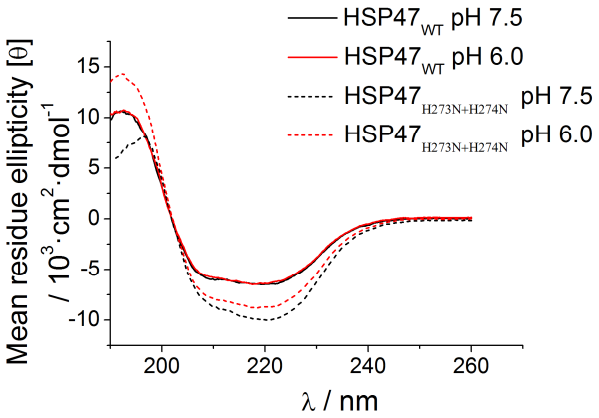


Figure 3.25 | Changes in secondary structure of HSP47$_{WT}$ and HSP47$_{H273N+H274N}$ between pH 7.5 and 6.0. CD spectra of wild type as well as mutant HSP47 reveal no changes in secondary structure between pH 7.5 and 6.0; in case of the latter a slight decrease in the amplitude can be observed. This might be attributed to a decrease in the concentration of natively folded protein at pH 6.0.

3.5 High-Throughput Screening for Inhibitors of the HSP47 – Collagen Interaction

3.5.1 Design principles

Ono *et al.* have previously utilized HTRF to prove that HSP47 preferentially binds to triple-helical collagens[117]. The basic setup of their assay has been judged to be quite compatible for high-throughput screening, and was adopted and optimized for this purpose (Fig. 3.26). One modification introduced over the course of this work was the choice of Tb^{3+}-Cryptate as the fluorescence donor over Eu^{3+}-Cryptate. The latter is, in its excited state, very effectively quenched by CH, NH and OH oscillators, necessitating its shielding from bulk water and other buffer components such as TRIS or HEPES[144]. This is usually achieved via inclusion of up to 500 mM KF; the resulting high ionic strength buffer, though, is hardly physiological, and the toxicity of fluoride requires special handling and disposal which may be complicated to incorporate into the highly standardized and automated high-throughput screening process. Fortunately, Tb^{3+}-Cryptate does not show the same quenching properties as its Eu^{3+} counterpart, and allows the circumvention of these problems.

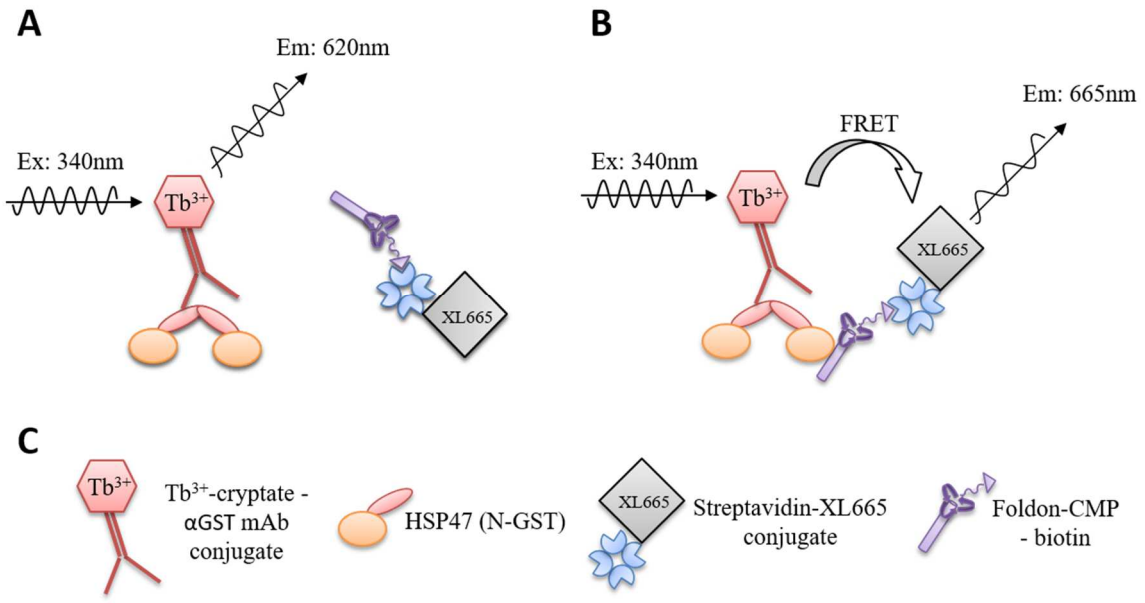


Figure 3.26 | Schematic depiction of the HT-FRET assay. ***(A)*** In case of no interaction between the analytes, only the emission of Tb^{3+}-cryptate at 620nm is observed. ***(B)*** Interacting analytes bring XL665 into proximity of Tb^{3+}-cryptate upon which FRET occurs and emission from XL665 at 665nm is observed. The ratio of the emission intensities I_{665nm}/I_{620nm} can be used to quantify the fraction of interacting analytes. ***(C)*** Legend.

The choice of the right tag and corresponding binding partner to mediate the interaction between analytes and their corresponding FRET donor or acceptor, from now on referred to as reporters, is a central aspect of any FRET based binding assay. Ideally, the affinity between the analytes and their corresponding reporters would be significantly higher than the affinity between the analytes themselves; furthermore, the affinity of each analyte to their corresponding reporter should be of similar magnitude. This ensures that both reporters show a comparable level of occupancy, or, in the ideal case, are saturated with analyte, increasing the probability of an interaction to be reported. The composition of the assay as described by Ono *et al.* was not modified in this regard: reporter binding was mediated via an N-terminal GST fusion to HSP47, which is recognized by αGST-mAb conjugated to the FRET donor Lumi4-Tb^{3+}-Cryptate, and biotinylation in the case of the CMP, which in turn is recognized by streptavidin conjugated to the FRET acceptor XL665. The corresponding binding affinities (K_D values in the range of low pM to low nM[145] and fM[146], respectively) are much higher than that of the HSP47 - CMP interaction ($K_D = 0.74 \pm 0.02$, as determined with BLI).

3.5.2 Characterization of the Assay

Overview

The utilized proteins, summarized in table 3.3, were produced as described under section 3.1.

Table 3.3 | Proteins used for Assay establishment

Protein	Remarks
$HSP47_{WT}$ (C-His_6)	No reporter binding
$HSP47_{WT}$ (N-GST)	Standard analyte
$HSP47_{D385N}$ (N-GST)	No CMP binding
CMP-R18	Displacement agent, no reporter binding
CMP-R18-biotin	Standard analyte
CMP-D18-biotin	No HSP47 binding
GST	-
GST-biotin	Counter-screen analyte

Determination of optimal analyte concentration

A grid-screen with varying concentrations of both HSP47 (N-GST) and CMP-R18-biotin served as a starting point and allowed determination of the Hook-point. The ratio of acceptor emission intensity to donor emission intensity I_{665nm}/I_{620nm} (from now on referred to as signal) was measured in standard assay buffer for varying analyte concentrations ranging from 2 nM to 128 nM. As can be seen in Figure 3.27, maximum signal to background ratio is observed for the condition containing 16 nM of each analyte, and any further increase of analyte concentration leads to signal decrease due to the Hook-effect. For following experiments, a working concentration of 10 nM for both analytes was chosen as the standard condition.

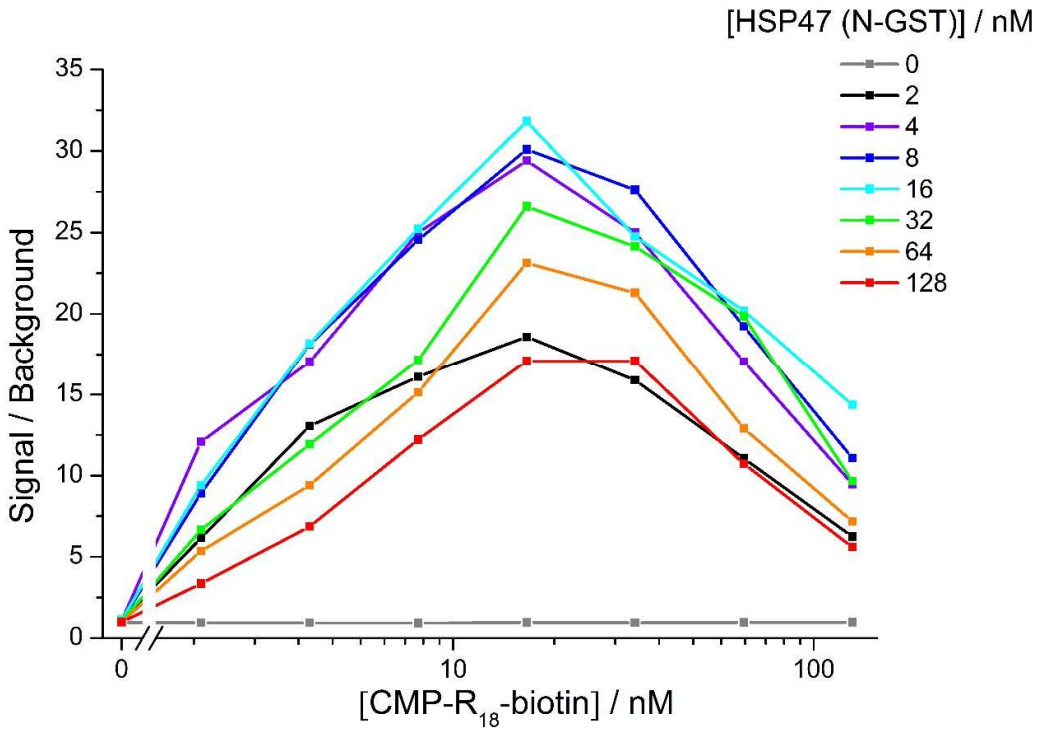


Figure 3.27 | Optimal analyte concentration and Hook-point of the HTRF assay. Dependency of the signal to background ratio on analyte concentrations. The signal to background ratio was calculated by dividing the ratio of acceptor/donor emission intensity (I_{665nm}/I_{620nm}) for each condition by the averaged ratio of acceptor/donor emission intensity of the background. The Hook-effect can be observed under conditions where the concentration of either either analyte surpasses 16nM *(cyan line)*.

Signal specificity

With the optimal working concentrations for the analytes established, the specificity of the observed signal was examined. Unspecific interactions of HSP47 (N-GST) or the CMP with the reporters were investigated via assaying analytes lacking the tags for reporter binding, and could not be detected (Fig 3.28). Whether the GST-tag undergoes any unspecific interactions with the reporters or the CMP was tested via conducting the assay with GST instead of HSP47 (N-GST) in absence as well as presence of the CMP, respectively, where no unspecific signal was observed Experiments with loss-of-interaction mutations ($HSP47_{D385N}$ and CMP-D18), however, revealed that a small part of the signal cannot be attributed to the canonical HSP47 - CMP interaction, which relies largely on the salt bridge between

D385 in HSP47 and the Arginine in the Yaa position of collagen. A possible source for this signal, besides unspecific binding between the analytes, might be the previously mentioned (Section 1.2.4) interaction of HSP47 with the GPP motif.

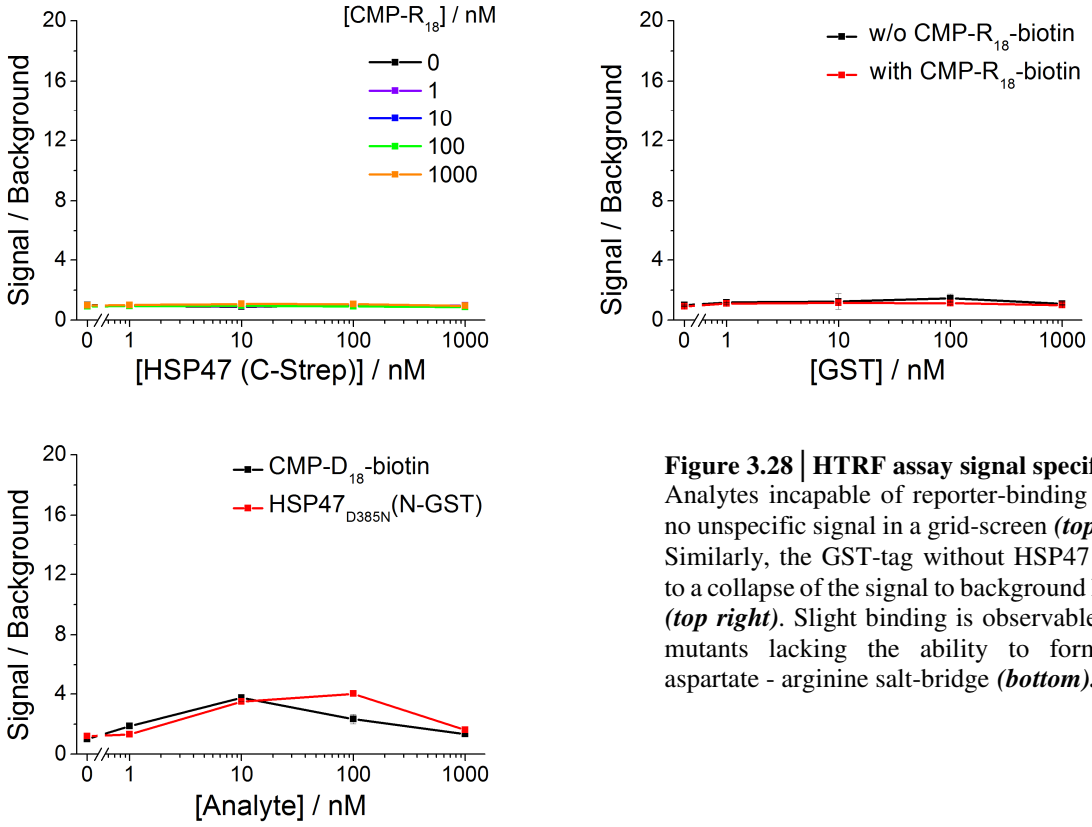


Figure 3.28 | HTRF assay signal specificity. Analytes incapable of reporter-binding show no unspecific signal in a grid-screen ***(top left)***. Similarly, the GST-tag without HSP47 leads to a collapse of the signal to background levels ***(top right)***. Slight binding is observable with mutants lacking the ability to form the aspartate - arginine salt-bridge ***(bottom).***

Signal stability

The stability of the signal was measured over the timespan which the screening was anticipated to require. Figure 3.29 shows that the signal to background ratio of the established standard condition develops over time as the thermodynamic equilibrium is approached. In the control experiments lacking either one of the analytes, no significant increase in signal was observed, which would have indicated unspecific binding attributable to proteins denaturing over time.

The DMSO tolerance of the assay has also been shown to be suitable for screening purposes, since no significant change in the magnitude of the signal was observed in conditions containing as high as 10% (v/v) DMSO (Fig. 3.29, *right panel*).

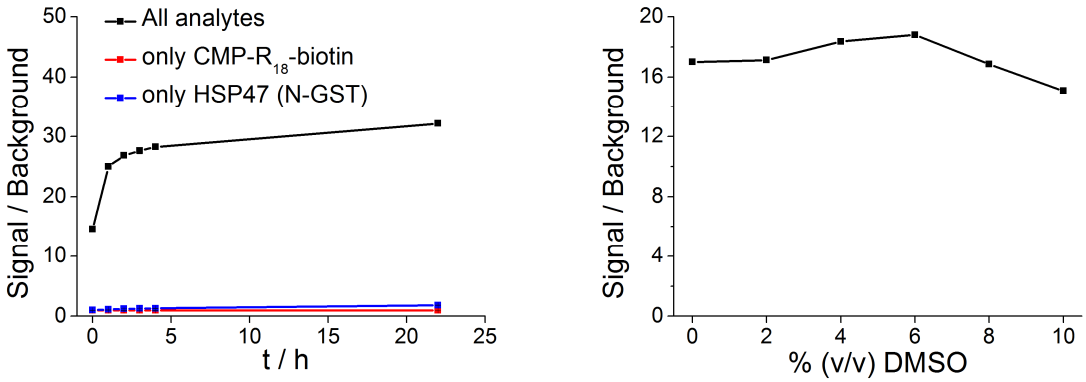


Figure 3.29 | HTRF assay signal stability. The signal doubles over the first hour of incubation and further increases by roughly 20% over the measured timespan ***(left).*** Inclusion of DMSO did not significantly affect the signal, with a roughly 15% decrease compared to the standard condition observed in presence of 10% (v/v) DMSO ***(right).***

Assay response to inhibition

After having established that the observed signal is specific and robust, a competition assay was prepared in order to test the response of the assay to inhibitors of the HSP47 – CMP interaction. Due to the lack of known inhibitors of the interaction, competition experiments with unbiotinylated CMP were conducted. Given that the dynamics of the complex formation allow it (i.e. a sufficiently high k_{off}), the untagged CMP should mimic an inhibitor and displace or prevent the binding of a tagged CMP to HSP47, leading to a decrease in the signal. This has been observed to be the case, as shown in Fig. 3.30.

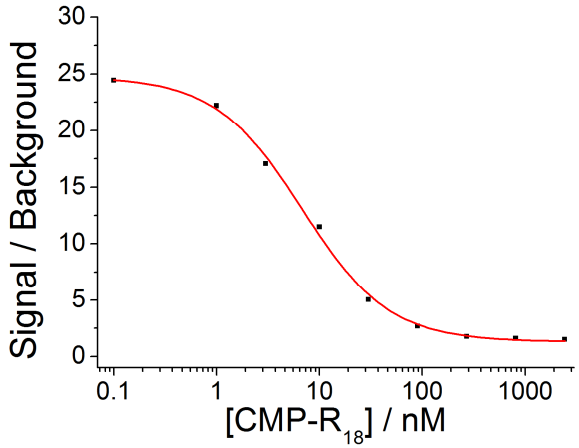


Figure 3.30 | Competition assay. The addition of unbiotinylated CMP displaces biotinylated CMP and leads to a concentration dependent decrease in signal to background ratio. The red line represents the logistic fit of the data points.

3.5.4 High-throughput Screening

(The experiments in this section were performed by Andreas Oder at the screening unit of the Leibnitz Institute for Molecular Pharmacology in Berlin.)

After internal testing at the screening facility, the standard assay condition was slightly altered to include 3.5 nM HSP47 (N-GST), 7 nM CMP-R18 -biotin, 1.25 ng/well αGST-mAb-Lumi4®-Tb^{3+}-Cryptate and 25 ng/well SA-XL665 in standard assay buffer, which was further supplemented with 0.1% (w/v) BSA (*vide infra*). The reagents were dispensed into the wells of a 384-well plats in the order 10 µl reporter mixture, 5 µl HSP47 (N-GST), 0.5 µl compound solution and finally 5 µl CMP-R18-biotin to yield the above stated final concentrations. The plates were incubated for 1 h and read out with a Tecan Genios Pro Platereader.

Statistics based assessment of high-throughput suitability

The donor and acceptor fluorescence intensities of a 384-well plate with 192 positive and negative controls each (i.e. with both analytes present and with HSP47 (N-GST) only, respectively) were read out and the Z'-factor for their ratio determined as described under 2.4.12. The obtained values, Z'_{ROBUST}: 0.76, and $Z'_{STANDARD}$: 0.73, indicated that the assay was suitable for high-throughput screening. The relevant statistics are displayed in table 3.4.

Table 3.4 | Statistical parameters for determination of the Z'-factor

	Median	Median absolute Deviation	Mean Value	Standard Deviation
Positive Controls	1.215	0.061	1.196	0.099
Negative Controls	0.067	0.001	0.066	0.002

Pilot Screen

The robustness of the assay towards measurement artefacts stemming from the up-scaling and automation as well as from the compounds themselves was tested in a pilot screen. One often encountered complication is that assay components adsorb to the tubing or metal tips of the dispenser system. This was revealed to be the case for both HSP47 and the CMP in preliminary tests, where incubation of either protein in the presence of metal dispenser tips was observed to lead to the collapse of signal. The inclusion of 0.1% (w/v) BSA, however, alleviated this problem (Fig 3.31).

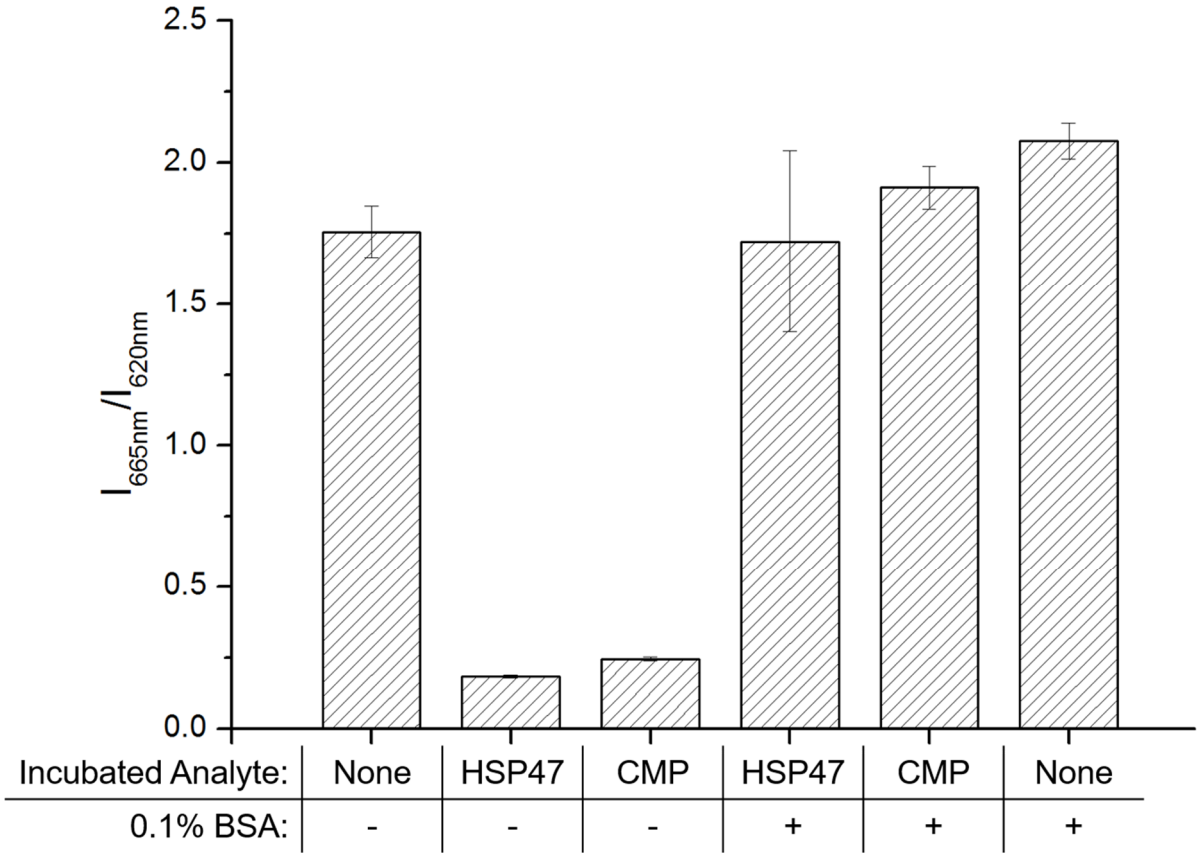


Figure 3.31 | Unspecific interactions of assay components with dispenser system. The donor to acceptor ratio of a standard assay condition was determined with and without pre-incubation of each analyte in presence of metal dispenser tips. Both HSP47 and the CMP adsorb to the dispenser tips, as is evidenced by the collapse of the signal. Utilizing 0.1% BSA as a buffer supplement prevents adsorbtion of the analytes and rescues the signal. Inclusion of BSA also leads to a slight increase in signal when neither analyte is subjected to preincubation with dispenser tips.

The pilot screen was conducted with ten 384-well plates, each with a randomly drawn collection of 368 compounds as well as 16 wells reserved for positive and negative controls. The compounds were administered at a final concentration of 10 µM. The Z'-factor for the individual plates ranged from 0.84 to 0.93 (Z'_{ROBUST}) and from 0.66 to 0.90 ($Z'_{STANDARD}$).

Primary Screen and Hit Selection

Table 3.5 gives an overview of the compound libraries used in the primary screen. A total of 40,480 compounds were tested. The following selection criteria were applied for filtering out artefacts:

(i) A reduction of at least 15% for the acceptor fluorescence intensity (I_{665nm}) compared to the controls

(ii) $-4 \leq$ Z-factor ≤ 4 for the donor fluorescence (I_{620nm})

Table 3.5 | Compound libraries included in the screening

Library	# of Compounds	Description
CBB1	16,544	World Drug Index derived scaffolds based on diversity
CBB2	7,744	See CBB1; emphasis on solubility
CBB3	4,576	Carboxylate, ketone and amine fragments
CBB4	6,688	Selleck library (FDA approved drugs), LOPAC (Aldrich), Artchem
CBB5	4,928	Donations of academic research labs

336 compounds were observed to increase donor fluorescence, artificially decreasing the signal (I_{665nm}/I_{620nm}), and 38,288 compounds failed to satisfy the set criteria. The remaining 1,471 compounds were sorted by ascending Z-factor for I_{665nm}/I_{620nm} and the top 352 compounds were selected for IC_{50}-determination and counter-screening. The average Z'-factor of the primary screen was 0.90 ± 0.04.

IC_{50}-Determination and Counter-screening

The determination of the IC_{50}-value was conducted in parallel to the counter-screen, in which the analytes were substituted for biotinylated GST which bridges the two reporters. The counter-screen serves to eliminate any false-positive hits that may arise from alterations of reporter fluorescence properties by the compounds. For each of the top 352 compounds, a 1:1 serial dilution ranging from 50 µM to 0.2 µM was assayed under normal as well as counter-screen conditions. Compounds which satisfied the following criteria were considered suitable for further validation experiments:

(i) An activity difference of at least 25% (i.e. a difference of at least 25% in I_{665nm}/I_{620nm} between the highest and lowest concentrations)

(ii) A concomittant decrease in I_{665nm}

(iii) No or a less pronounced activity difference in the counter-screen.

3.5.5 Experimental Hit-validation

Hit validation with Biolayer Interferometry

The high-throughput screening delivered 31 compounds as potential inhibitors, 29 of which were provided by the FMP Berlin for further study. The validation was conducted with bio-layer interferometry using the Octet system from ForteBio. For this, the association of HSP47 (C-His_6) to immobilized CMP- R18 -biotin was monitored in presence of 50 µM compound. Figure 3.32 displays the maximal response achieved during the experiment as well as the extracted k_{obs} values. 4 compounds which induced a notable decrease in either the maximal response (control mean value: 0.99 nm) or k_{obs} (control mean value: 0.22 s^{-1}) while at the same time leading to no increase of the other were selected

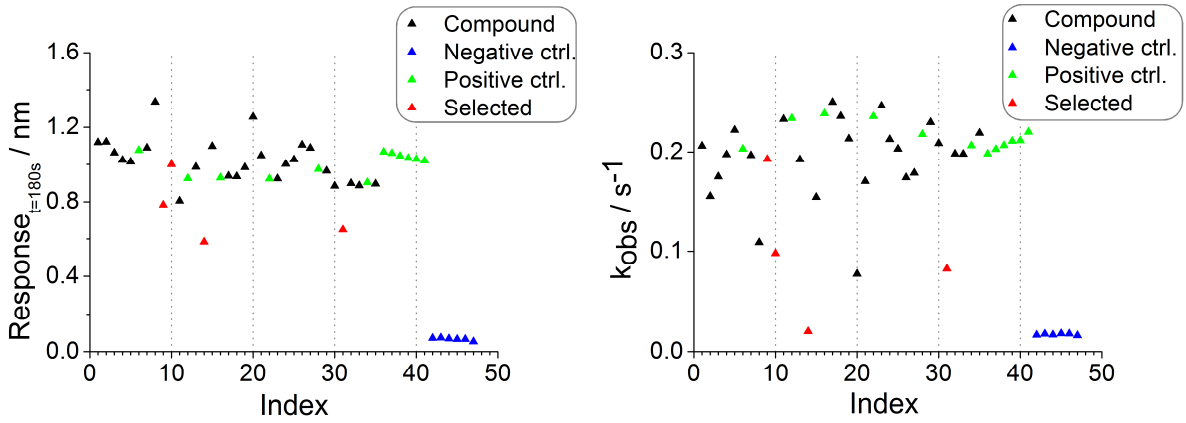


Figure 3.32 | Hit validation via biolayer interferometry. The maximal response after 180s ***(left)*** and corresponding k_{obs} ***(right)*** of the association of HSP47 to immobilized CMP-R18-biotin was determined in presence of 29 compounds *(black and red)*. In the positive controls, DMSO was added to a final concentration of 0.5% (w/v) *(green)* while for negative controls, loading of CMP-R18-biotin on the biosensor was omitted *(blue)*. The compounds with the indices 9, 10, 14 and 31 were selected for further verification experiments *(red)*.

for further studies. The data for some compounds were discarded due to unsatisfactory curve fitting or the detection of precipitates in the stock solution.

The inhibitor candidates with the index numbers 14 and 31, from now on referred to as Compound 1 and 2, respectively, are well-known pharmacologically active molecules and thus commercially available; as such, they were the first choice for more in-depth investigation. The efficacy of both compounds was scrutinized with HTRF (Fig 3.33). Both compounds seem to have an inhibitory effect, with the estimated IC_{50} values lying in the low micro-molar range for Compound 1 and in the high micro-molar range for Compound 2. Analysis of the reporter fluorescence, however, shows that for Compound 1 the signal decrease in the concentration range from 1 to 10 µM can be attributed to a decrease in donor fluorescence. It should be noted that under the established assay conditions, the decrease in donor intensity due to FRET is masked by the presence of excess reporter; this implies that Compound 1 is capable of quenching donor fluorescence via an unknown mechanism.

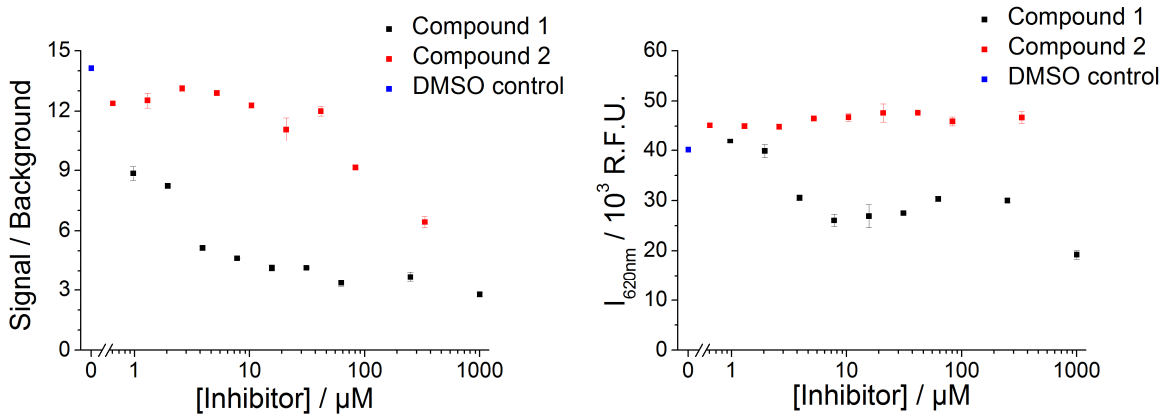


Figure 3.33 | Assessment of the efficacy of inhibitor candidates with HTRF. The signal to background ratio for the standard assay condition was plotted against different concentrations of compound 1 (1000nM - 1nM, *black*) and compound 2 (333nM - 0.7nM, *red*) ***(left)***. Both compounds are capable of disrupting the HSP47 - collagen complex, albeit with rather high IC_{50} values (low µM for compound 1, high µM for compound 2). However, analysis of donor fluorescence intensity I_{620nm} reveals that compound 1 acts as a fluorescence quencher and thus is possibly a false positive hit ***(right)***.

Compound 2 was subjected to a more detailed analysis via BLI, in which the interaction of HSP47 with immobilized CMP-R18 was monitored in presence of a serial dilution of the inhibitor candidate (Fig. 3.34). An estimate of the IC_{50} value can be given via graphing the maximal response reached in the experiment against compound concentration; however, it should be noted that since steady-state was not reached, this estimate only serves qualitative purposes.

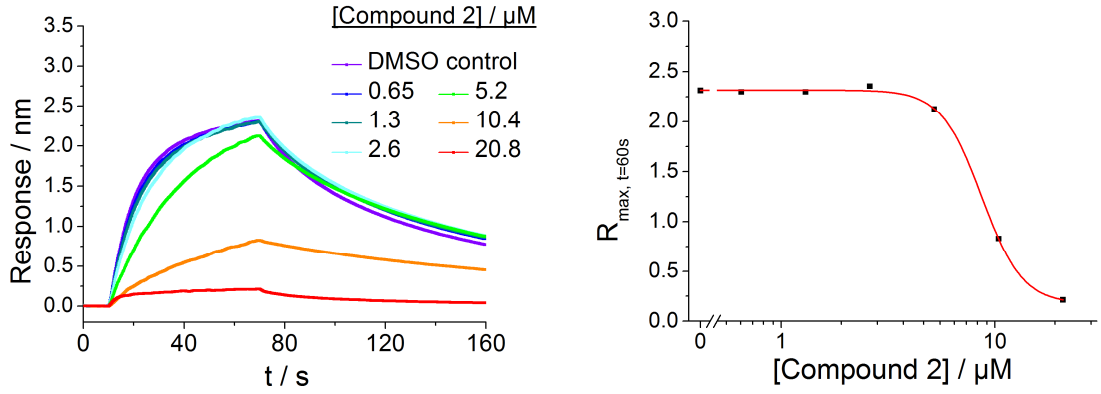


Figure 3.34 | Assessment of the efficacy of inhibitor candidates with BLI. 2.5µM HSP47 were pre-incubated in presence of a serial dilution of compound 2 and subsequently the binding curves to immobilized CMP-R18 recorded via BLI ***(left)***. The compound is seen to interfere with complex formation in a concentration dependent manner. The maximum response after 60s was plotted against compound concentration to give an estimated IC_{50} between 5 - 11µM ***(right)***.

3.6 Crystallization of HSP47

The initial focus of crystallization experiments lay on obtaining a low-pH structure of HSP47, with the aim of illuminating the molecular details leading to a loss in affinity towards collagen. Unfortunately, the intrinsic instability of HSP47 at mildly acidic pH, as previously shown using DSF, has been detrimental to these efforts, and it was observed that the protein precipitates overnight in conditions with a pH below 7. Similarly, attempts in crystallizing the binding-incompetent mutant H238N, which is markedly less stable than the wild type already at pH 7.5, were also not met with success. The focus was then shifted on the pH-insensitive mutant H273N + H274N, which did yield crystals, albeit at a pH of 7.5 (Fig. 3.35) At this pH, the binding characteristics of the double-mutant are indistinguishable from those of wild type HSP47.

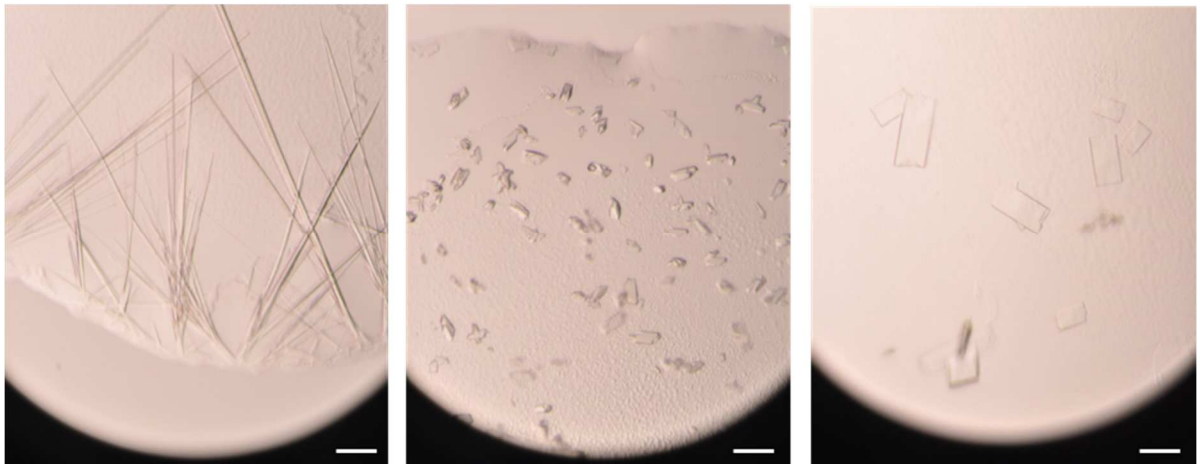


Figure 3.35 | Crystals of pH-insensitive mutant H273N + H274N. ***Left:*** Crystals of the apo form *(*Precipitant solution: 4% Tacsimate pH 7.0, 18% PEG-3350, 0.1M HEPES (pH 7.5). ***Middle:*** Crystals of HSP47 in complex with (15)R8 (Precipitant solution: 8% Tacsimate pH 8.0, 19% PEG-3350, 0.1M HEPES (pH 7.5) ***Right:*** Crystals of HSP47 in complex with (18)T8R11 (Precipitant solution: 5% Tacsimate pH 7.0, 18% PEG-3350, 0.1M HEPES (pH 7.5).

4. Discussion

4.1 HSP47 function does not require conformational rearrangements

The perhaps most intriguing aspect of HSP47 is that this molecular chaperone for collagen spawned from a family of protease inhibitors. In this regard, HSP47 is not unique among Serpins, since collagen-binding properties are also attributed PEDF[147] and Maspin[103]. It would be intuitive to suggest that the structural and dynamic peculiarities of Serpins, i.e. the meta-stable native structure, the highly flexible RCL and the conformational transitions and oligomerizations involving its partial or full insertion into β-sheet A, might somehow contribute to the function of HSP47. In line with this assumption, Dafforn *et al.* have previously proposed a "flying-capstan" model for HSP47, in which the protein trimerizes via insertion of the RCL of one into the β-sheet of another[95]. This model originated from the analysis of trimeric HSP47, a species routinely detected during purification, which attested it a high thermostability with an increase in T_m of up to 20 °C. This remarkable shift in thermal stability is a well-known consequence of RCL insertion, prompting the authors to hypothesize that the observed trimeric species is similar to loop-sheet polymers reported for other Serpins, such as α1 anti-trypsin[148]. Koide *et al.*, however, have clearly shown that the trimeric species is not formed in presence of thiol-alkylating reagents, implying the involvement disulphide-bridges[121], and the crystal structure has conclusively disproven the flying-capstan model. The conundrum here is that HSP47 does not have the capacity to form disulphide-bridged trimers since it harbors only one cysteine. The residue, located adjacent to the breach-region in the hinge between β-sheet A and helix E, is furthermore inacessible without disruption or translocation of these structural elements.

As shown in this work, substitution of the native RCL of HSP47 with the larger RCL of α1-antitrypsin did not result in a spontaneous transition to the latent conformation; RCL insertion was also not triggered upon proteolytic digest. While these were by no means exhaustive studies, with the available data it seems improbable that the Serpin-typical conformational transitions are relevant for the function of HSP47; a more plausible role for the RCL might be one as an element of interaction with collagens, as suggested by its localization close to the binding interface and the finding that the RCL substituted mutant has a decreased affinity towards CMPs. The fact that client-binding in HSP47 occurs at the same site where protease binding occurs in other serpins evokes the question whether binding to collagen serves as protection against proteolytic cleavage. It is quite possible that the emergence of HSP47 from Serpins, which are ancient proteins stretching back to bacteria, can be solely attributed to the fact that the latter were simply available in the pool of secretory proteins when the need for a collagen chaperone arose.

Structural analysis shows that HSP47 does not undergo any conformational rearrangements upon client binding, since the structure of apo-HSP47 is identical to that of HSP47 in complex with various CMPs.

This observation does not exclude pH-induced conformational changes resulting in a binding-incompetent form, though, since all crystal structures reported in literature[122] as well as those obtained by the author grew at pH values greater or equal to 7.5, granting insight at the binding competent form of HSP47. There might yet exist a low-pH conformation of HSP47; unfortunately, crystallization of HSP47 at pH values close to 6.0 has not been successful, presumably due to the decreased stability of the protein under such mildly acidic conditions. The presence of such an "acidic" conformation is hinted at by spectropolarimetric analysis, which indicates a pronounced change in secondary structure below pH 6, as reported by several studies[85,149,150]. At a physiological pH of 7.2, the circular dichroism spectrum of HSP47 is dominated by the broad, negative ellipticity in the region between 208 nm and 222 nm which is characteristic for folded proteins with a high α-helical content. Lowering the pH gradually decreases magnitude of the ellipticity as the pH approaches 6.0; once this threshold is passed, a characteristic negative peak emerges at roughly 205 nm. While this process is reversible and adjusting the pH back to 7.2 restores the original curve (albeit only after overnight incubation), it remains to be proven whether this observed "acidic" state of HSP47 corresponds to a distinct conformation which can be associated with client release. The lower magnitude of the ellipticity above 210 nm as well as the trend towards negative ellipticity at 195 nm are usually associated with disordered proteins[151]; furthermore, the resemblance to spectra of partially acid-denatured proteins[152] is conspicuous. It is quite possible that the observed "acidic" conformation is in fact HSP47 which is partially unfolded or in a molten globule state. In this work, spectropolarimetry did not reveal any difference in the secondary structure of canine HSP47 between pH 7.5 and pH 6.0. Since pH values below 6.0 are not encountered by HSP47 *in vivo* and BLI indicates that a pH of 6.0 is sufficient to drive client dissociation, it is rather unlikely that the proposed "acidic" state is of relevancy for the function of HSP47.

4.2 Molecular Details of Client Release

4.2.1 On Data Quality in BLI Experiments

As previously mentioned, determination of the association rate from BLI experiments generally suffered from a low precision. This is immediately apparent from Figure 4.1, where the kinetic parameters k_{on} and $k_{off,}$ calculated for $HSP47_{WT}$ in different experiments are contrasted. The increasingly large average error observed for k_{on} values at low pH presumably reflects the decrease in the thermal stability of HSP47 going from pH 7.5 to 6.0: Aggregated or partially unfolded HSP47 would lead to a decrease in the concentration of binding-competent protein and presumably increase the probability of unspecific interactions with the biosensor. These phenomena are hard to account for and possibly add a significant random component to the generated signal. However, the same trend is notably absent in case of the k_{off}, where the calculated errors are small and do not discernibly increase at lower pH. Dissociation is a first-order reaction and thus independent of free analyte concentration; furthermore, it

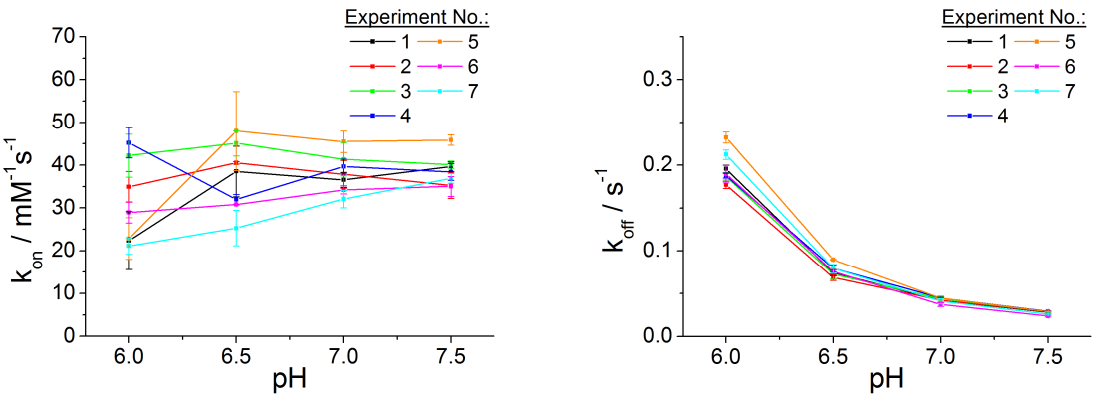


Figure 4.1 | Data variability in BLI experiments. Comparison of the pH dependent k_{on} ***(left)*** and k_{off} ***(right)*** values determined for $HSP47_{WT}$ over seven different experiments.

can be presumed that aggregated, "unhealthy" HSP47 will remain immobilized on the sensor, meaning that only the dissociation of intact HSP47 is picked up.

Another source for the large error accompanying extraction of k_{on} from BLI sensorgrams was the often poor fit of the applied monophasic exponential association fit to the data (Fig. A.1). This initially prompted the validation of the assumed stoichiometry using ITC, which confirmed the 1:1 Langmuir model and suggested that the reasons for the poor fit did not originate from an unsuitable mathematical model but rather from the experiment itself. As mentioned in section 3.3.1, the association of HSP47 to the CMP loaded biosensor does not reach steady-state in the time frame allowed by the method. Due to this, the maximum response is not a defined value, and needs to be approximated during curve fitting. The time required to reach equilibrium is correlated to the kinetic parameters as shown in the following equations[153]:

$$t_{1/2} = \ln(2) / k_{obs} \qquad \textit{(Eq. 6)}$$

$t_{1/2}$ = half-time of equilibration reaction

$$t_{97\%} = 5 \cdot t_{1/2} \qquad \textit{(Eq. 7)}$$

$t_{97\%}$ = time required to achieve 97% of the final equilibrium

Since $k_{obs} = k_{on} \cdot [A] + k_{off}$, the time required for reaching steady state can be calculated from the kinetic parameters. For 5 µM $HSP47_{WT}$ at pH 7.5, this yields:

$$t_{97\%} = 5 \cdot 0.693 / [(35.04 \cdot 10^3\,M^{-1}s^{-1}) \cdot (5 \cdot 10^{-6}\,M) + 0.028\,s^{-1}] = 16.5\,s$$

Clearly, the time limit imposed by the BLI technique (a maximum of 4 minutes for the association phase) should be more than sufficient to reach steady state in case of the HSP47 - CMP interaction. While the asymptotic behaviour of the association curve is usually an indication of unspecific binding in BLI (as stated by the manufacturer), regeneration of the sensor and repetition of the experiment

produces nearly identical association curves, which rather defies the definition of unspecific interactions.

Despite these uncertainties, it is clear that the association rates either stay relatively constant over the tested pH range or decrease along with the pH. While the accuracy of the determined values for k_{on} and K_D are certainly disputable, the overall trend does indicate that the increase in k_{off} at lower pH is not offset by an increase in k_{on}.

4.2.2 Characterization of the Histidine Mutants

The Role of His215 in Collagen Binding

Abdul-Wahab *et al.* have previously reported that mutation of either His215 or His216 in murine HSP47 affects the pH-sensitivity of the collagen interaction[150]. Corresponding alanine mutants were shown to elute early (i.e. at a higher pH) during collagen-affinity chromatography, which was argued to reflect a perturbation in pH-dependent, controlled client-release, leading the authors to suggest that these residues play a critical role in the pH-switch mechanism. This is partially contradicted by the results presented in this work: while H215N and H215N + H216N indeed display a decreased affinity to CMP-R18, which explains the observed early elution from the gelatin-sepharose column, H216N was indistinguishable from the wild-type. Moreover, the weaker binding does not necessarily stand in relation to a change in the pH-sensitivity, and accordingly no significant difference in the pH-dependency of client-release was observed in BLI experiments.

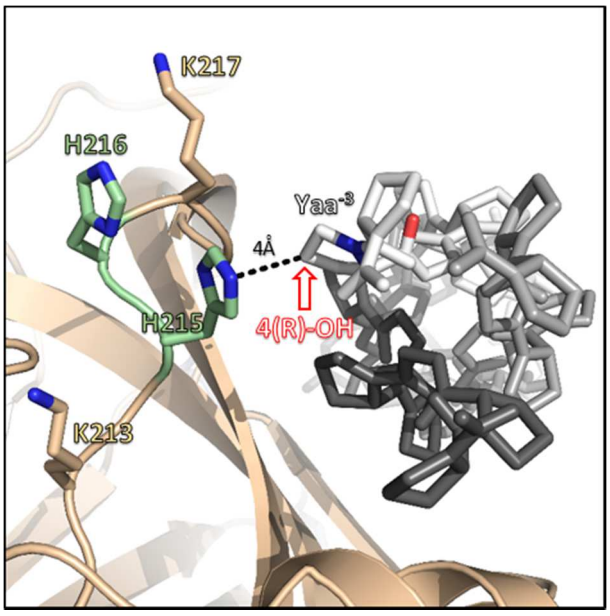


Figure 4.2 | Close-up view of interface histidines 215 and 216. The histidines *(green)* are framed by a pair of lysine residues (K213, K217) which possibly decrease the pKa values of their imidazole side-chains. His215 comes into close proximity of Yaa^{-1} (in this case proline), with a distance of 4Å between them *(black dashed line)*. Proline hydroxylation overwhelmingly occurs at Cγ in the indicated stereochemical configuration *(red arrow)*. (PDB ID: 3ZHA)

The side-chain His215 comes into close proximity to the Yaa^{-3} residue on the chain with a register of +1 in relation to the chain harbouring the recognized arginine. In cases where this residue is a proline, the distance between the *tele*-nitrogen of the imidazole ring and Cγ of the pyrrolidine ring amounts to approximately 4 Å. The interaction lacks a discernible polar component and seems to be hydrophobic

of nature; it is thus unclear why the incorporation of a smaller amino acid such as asparagine disrupts client binding in such a magnitude. The sensitivity of the HSP47 - Collagen interaction towards disruptions at this site evokes the question whether His215 also partly responsible for the decreased affinity of HSP47 to collagens with hydroxyproline in this position, since the 4-hydroxyl group in the Cγ-exo pucker would experience increased steric clash with the imidazole moiety.

The Role of His273 and His274 in Collagen Binding

In this work it was shown that although His273 and His274 can interchangeably act as the pH-trigger, substitutions at position 273 have a larger impact on the pH-sensitivity of the HSP47 - collagen interaction, as was the case with H273N and H273A. In agreement with the CpHMD simulations, which predict that His273 is substantially more basic (pKa = 5.23) than His274 (pKa < 4), this suggests that the pH-trigger role should be attributed to His273, and that His274 merely compensates for this role in case His273 is substituted. However, this assumption is called into question by the fact that His274 is almost universally conserved in SERPINH1 orthologs, while His273 is commonly found to be replaced by asparagine or tyrosine. Taking this into account, a more reasonable model would be that His274 is the actual pH-trigger, and that the changes in pH-sensitivity observed for His273 mutants do not arise from the loss of protonation at this site but are rather a consequence of an altered microenvironment influencing the pKa of His274. A more universal formulation of these findings would be that the pH-sensitivity of the HSP47 - collagen interaction is by and large determined by the motif Xaa273His274. The molecular details, however, are still very much unclear. A closer inspection of the motifs location reveals that it is embedded in a 6 residues long loop connecting s4B and hG. With the exception of Xaa273, the amino acids comprising this short region are highly conserved in HSP47 orthologs. Since the peripheral and solvent-exposed location of this loops renders large contributions to structural integrity unlikely, the selection pressure can rather be attributed to functional relevancy. The proline residues making up both termini of the loop are strictly conserved and presumably function as secondary structure breakers, a feature commonly attributed to prolines due to the absence of a nitrogen-bound hydrogen as well as their intrinsic geometric limitations. The XH motif is flanked C-terminally by a hydrophobic residue (predominantly valine; methionine and leucine are also encountered in Actinopterygii), followed by a glutamic acid (Glu276) residue which forms a salt bridge with an arginine located on hG (Arg280). Both of these residues are strictly conserved in HSP47 orthologs as well. Together with a hydrogen bond between the backbone carbonyl of Pro272 and the side-chain of His386, this interaction seems to lock the loop in place and limit its conformational flexibility. In cases where Xaa273 is a histidine, an additional H-bond interaction between the imidazole sidec-chain and Asp247 is observed. Comparison of the loop sequences between HSP47, α1-Antitrypsin as well as the collagen binding Serpins PEDF and Maspin reveals that the salt-bridge between Glu276 and Arg280 is unique to HSP47 (Table 3.6).

Table 4.1 | Sequence comparison of the XH-motif containing loop of HSP47 with selected Serpins

Protein:	HSP47	α1-Antitrypsin	PEDF	Maspin
Sequence:	^{272}PHHVEP	^{255}PDEGK	^{259}PLKVTQN	^{233}PKDVTG

How exactly the positive charge introduced via protonation of His274 (or His273) perturbs the interaction remains subject to speculation. CD spectroscopy did not detect pronounced changes in secondary structure between pH 7.5 and 6, nor did the pH-insensitive mutant behave differently than the wild-type, suggesting that at most small-scale rearrangements, leading to a distortion of the interaction interface, are involved in the pH-switch mechanism. One possible model would be that upon protonation, the imidazolium side chain of His274 (or His273) can interact with Glu276, weakening the salt-bridge to Arg280 and liberating the loop from its tethering to hG. This could enable the loop to "flick" away the collagen client. Alternatively, the loop may directly interfere with the order around the nearby intermolecular salt-bridge.

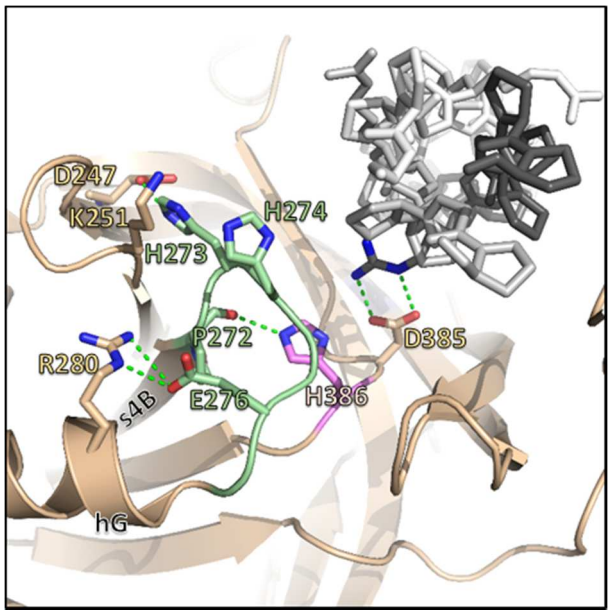


Figure 4.3 | Close-up view of interface histidines 273 and 274. The histidines are part of a 6-residue long loop connecting s4B to hG *(green)*. The flexibility of the loop is limited N-terminally by H-bonds between P272 and H386 as well as between H273 and D247; C-terminally, a salt-bridge between E276 and R280 keeps the loop fixed *(green dashed lines)*. The critical salt-bridge between D385 and R8 is in proximity of this loop.

Evolution of the XH-motif

Phylogenetic analysis shows that the identity of the N-terminal amino acid in the XH motif is clade-specific. In lamprey, the only representative of Hyperoartia for which sequence data is available, this residue is a leucine. The most common amino acid encountered in ray-finned fish (Actinopterygii) is tyrosine, which is also found in Coalecanth, representative for lobe-finned fish (Sarcopterygii). In Tetrapoda, the tyrosine is replaced by either a histidine (Amphibia and Mammalia) or an asparagine (Sauropsida), both amino acids being strictly conserved within their respective clades according to available data.

Since lamprey are the most ancient vertebrates included in the analysis, it is plausible to assume that the LH motif is ancestral to all others. The codon for leucine displays a twofold and fourfold degeneracy in the first and third position, respectively; still, only the codon for histidine is accessible via substitution of a single nucleobase. Tyrosine and asparagine, on the other hand, can be derived from histidine with a single point mutation. Thus, the YH motif of Actinopterygii either emerged from the LH motif via an intermediary HH motif, or the HH motif itself is ancestral to all others, the LH motif being a derivative of it (Fig. 4.4).

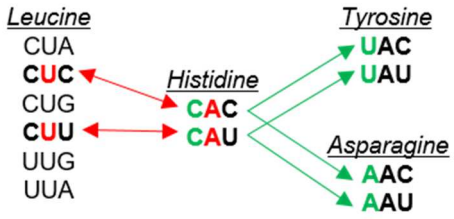


Figure 4.4 | Clade-specific residues encountered in the XH motif are accessible via singular missense mutations in the histidine codon. Nucleotide substitution in the first position of the histidine codon gives rise to tyrosine or asparagine (*green*) while mutations in the second position exclusively lead to leucine (*red*).

4.3 Open Questions regarding the HSP47 - Collagen Interaction

4.3.1 The Influence of Chain-register on HSP47 Binding is Unknown

Limitations of Homotrimeric CMPs

The foldon-domain of bacteriophage T4 fibritin has proven itself an excellent tool for the assembly of thermostable collagen model peptides. The stability of the trimer is quite extraordinary: it routinely survives the standard protocol for SDS-PAGE sample preparation, which includes 5 minutes of boiling in SDS-containing Laemmli buffer. The 27mer appendix further boasts excellent solubility in water and introduces a tryptophane residue to the CMP construct, rendering it detectable via UV-spectroscopy. One of its innate disadvantages, though, is that foldon-domains do not allow the steered assembly of heterotrimeric CMPs with defined composition: the heterotrimeric CMP with the desired register would have to be identified and extracted from a mixture of CMPs with nearly identical physical and chemical properties, which can be a daunting task. The structural characterization of the HSP47 - collagen interaction by Widmer *et al.* as well as the biophysical characterisation presented in this work were thus conducted using homotrimeric CMPs. Since HSP47 interacts with all three α-chains, access to heterotrimeric CMPs would allow many open questions regarding structural aspects of the HSP47 - collagen interaction to be addressed. For one, it is peculiar that in all available crystal structures, HSP47 is only seen interacting with arginine residues either in the leading or in the trailing strand. The details of this preference could be illuminated with access to heterotrimeric CMPs of the type PRR, RPR, RRP,

with one α-chain lacking the arginine in the leading, middle and trailing strand, respectively. Furthermore, the register-specific influence of amino-acid substitutions or modifications such as proline hydroxylation would be open to investigation.

Synthetic Heterotrimeric CMPs

Several strategies for the synthesis of heterotrimeric CMPs have been developed in the past. In cases where the α-chain composition or register markedly influences CMP stability, the thermodynamically favourable trimer can be enriched via simple cooling of a monomeric (i.e. heated) α-chain mixture[154]. Gauba *et al.* for instance have steered the self-assembly of a heterotrimeric (α1,α2,α3) CMPs by assigning the different α-chains a neutral, positive or negative net charge each, yielding the electrostatically neutral species α1α2α3[155]. More specific synthesis strategies employ orthogonal protecting groups to selectively couple α-chains to a trimeric template. This is exemplified by the synthesis approach utilized by Li *et al.*, who coupled the trimeric tris(2-aminoethyl)amine (TREN) to α1-chains immobilized on a resin at a very low density, ensuring the coupling of only one α1-chain to the template. Subsequent attachment of α2 chains in liquid phase yielded a mixture of $\alpha1(\alpha2)_2$ trimers which was resolved using HPLC[156]. In yet another approach, cystine knots are (in analogy to natural collagens) introduced between each pair of α-chains via sequential deprotection and oxidation of cysteines, allowing the synthesis of heterotrimeric CMPs with a defined composition and register[157].

Heterologous Expression of Heterotrimeric CMPs

The C-terminal trimerization domains of collagens have been subject to research for over a decade, and crystal structures have been obtained for the NC1 domains of collagens I (only homotrimeric $[\alpha1(I)]_3$)[158], III[159], IV[160], VIII[161], X[162], XV[163] and XVIII[164]. Of particular interest among these are the trimerization domains of collagens I and IV, since they are able to organize into heterotrimeric isoforms. (Collagen I: $[\alpha1(I)]_2[\alpha2(I)]$, Collagen IV: $[\alpha1(IV)]_2[\alpha2(IV)]$, $[\alpha3(IV)][\alpha4(IV)][\alpha5(IV)]$, $[\alpha5(IV)]_2[\alpha6(IV)]$). These domains, however, are rather large and complex, consisting of ~250 residues in collagen type I and ~230 residues in collagen type IV. Furthermore, they are subjected to several post-translational modifications, including a very unusual crosslink in form of a sulfilimine bond between a methionine and a hydroxylysine discovered in the latter type of collagen[165]. These proteins are thus either directly extracted from tissues or expressed in mammalian or insect cells, and an easier to produce alternative has been long sought after. Unexpectedly, the ideal candidate for a straightforward assembly of heterotrimeric CMPs has not been found among NC1 domains, but rather in the NC2 domain of FACIT collagen type IX. This domain not only initiates trimerization and triple-helix formation independent of the C-Propeptide (NC1)[166,167], but is also able to determine chain register and grants easy access to heterotrimeric CMPs via heterologous expression in *E. coli*[167]. The domain consists of three distinct chains, each with a length of 36 or 37 amino acids, and is stabilized by a single disulphide bridge between chains α1 and α3.

4.4 HSP47 as a Therapeutic Drug Target

Owing to their unusual size and shape as well as their structural and chemical peculiarities, such as the staggered triple-helix or incorporation of hydroxyprolines, collagens impose out of the ordinary challenges regarding their production, processing and trafficking. Their biosynthesis is a complex, multi-step endeavour which is prone to perturbation at many levels. Of particular importance for collagens is the concept of post-translational over-modification, which describes aberrant (i.e. excessive) hydroxylation, glycosylation and cross-linking of collagens, resulting in defective proteins which are associated with a variety of diseases such as Osteogenesis Imperfecta or the Ehlers-Danlos syndrome. HSP47 is a key component for regulating post-translational modifications of collagens, and deficiencies in its function typically result in over-modified collagens and the associated pathological conditions. Unfortunately, the factors resulting in loss of function are primarily of genetic origin, and as such cannot be addressed via conventional drugs. This renders HSP47 an unlikely target (at the protein level) for intervention against diseases characterized by collagen overmodification, despite its pivotal role in regulating collagen processing.

The true potential of HSP47 as a therapeutic drug target arises from its involvement in fibrosis. Fibrosis is defined as the excessive deposition of fibrous matrix proteins such as fibronectin, laminin and, predominantly, of interstitial collagens type I and III[168]. While the fibrogenic response is an important part of benign processes such as scar tissue formation, it can also develop into a pathogenic condition when triggered by acute or chronic inflammation[169], oxidative stress[170] or epithelial injury[124]. Fibrosis is capable of disrupting tissue function and can lead to organ failure and ultimately death[171]. A common hallmark of fibrosis is the upregulation of HSP47 expression levels. This has not only been observed in fibrotic diseases such as keloid[172], oral submucous fibrosis[173] and pulmanory fibrosis[174], to name a few, but also in complex diseases with fibrotic manifestations, such as systemic sclerosis of skin fibroblasts[175]. Over the last decade, several studies have focused on the silencing of the HSP47 gene in order to resolve fibrosis, and siRNA knockdown of the collagen chaperone has been demonstrated to resolve fibrosis in cases of liver cirrhosis[176] and pancreatic fibrosis[177], among others.

Despite these promising results, the development of small organic molecule inhibitors of the HSP47 - Collagen interaction has been left rather orphaned. In an article published 2005, Thomson *et al.* have presented the results of a small-scale (2080 compounds) screening, conducted using their previously established assay based on turbidimetric measurement of collagen fibril formation[116]. This resulted in the identification of 4 compounds as potential inhibitors of the HSP47 - Collagen interaction[178]. Pillai *et al.* utilized *in-silico* docking studies to identify HSP47 binders among a library of known natural products. Of each study, one compound was acquired and tested for its inhibitory potential using BLI; however, in both cases no inhibitory activity was observed (Fig. A.2).

The following features of HSP47 make it stand out as an attractive target for anti-fibrotic treatment:

(i) It fulfils a highly specialized role, interacting primarily with collagen and, to a lesser extent, with other collagen-modifying or recognizing proteins such as LH2 or TANGO1. This limits the effects of HSP47 inhibition to collagen biosynthesis.

(ii) Although the pathogenesis and progression of fibrosis is influenced by numerous different factors, all fibrotic diseases ultimately converge on excessive collagen deposition. Thus, intervening at this point would ensure a broad range of application for the treatment.

Notwithstanding the latter point, coverage of the whole spectrum of diseases with fibrotic manifestations would be well beyond the scope of this work, and the discussion will thus rather focus on skin fibrosis, since the research was conducted in context of investigating molecular mechanisms regulating skin homeostasis. Currently available treatment options for dermal fibrosis are surgical excision, targeting of the inflammatory response with corticosteroids, or anti-sense oligonucleotide treatment directed against various key elements in the fibrotic cascade such as TGF-β or LOX[179]. An endostatin-derived peptide (E4) has been shown to ameliorate TGF-β or bleomycin induced skin fibrosis[180] and its efficacy as an orally bioavailable, potent anti-fibrotic agent in animal models is well established[181]. According to the clinical trial registry of the US National Institute of Health (www.clinicaltrials.gov), the efficacy of Pirfenidone (5-methyl-1-phenyl-2-(1*H*)-pyridone), an anti-fibrotic drug indicated for the treatment of idiopathic pulmanory fibrosis, is currently under evaluation for the treatment skin fibrosis associated to systemic sclerosis. Interestingly, one of the observed effects of Pirfenidone treatment is the inhibition of increased HSP47 expression induced by TGF-β1[182]; however, it is unclear whether this is tied to the concurrently observed reduction of collagen type I production, the expression of which is known to be tied to that of HSP47.

To summarize, small organic molecules which allow the modulation or inhibition of the HSP47 - collagen interaction are not only of interest as research tools, but are also promising candidates for anti-fibrotic drugs. As such, the author hopes that the work laid out in this thesis may provide a small but nevertheless worthwhile initial step (i.e. a promising lead structure) in the development of therapeutic means to combat fibrosis.

5. Literature

1 Heinemeier, K. M. *et al.* Radiocarbon dating reveals minimal collagen turnover in both healthy and osteoarthritic human cartilage. *Sci Transl Med* **8**, 346ra390, doi:10.1126/scitranslmed.aad8335 (2016).

2 Wenger, M. P., Bozec, L., Horton, M. A. & Mesquida, P. Mechanical properties of collagen fibrils. *Biophys J* **93**, 1255-1263, doi:10.1529/biophysj.106.103192 (2007).

3 O'Reilly, M. S. *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277-285 (1997).

4 Sorensen, D. R. & Read, T. A. Delivery of endostatin in experimental cancer therapy. *Int J Exp Pathol* **83**, 265-274 (2002).

5 Daviesz, D., Calderwood, D. A., Weston, S. A., Takigawa, M. & Humphries, M. J. Molecular Characterisation of Integrin–Procollagen C-Propeptide Interactions. *European Journal of Biochemistry* **246**, 274-282, doi:10.1111/j.1432-1033.1997.t01-1-00274.x (1997).

6 Abdillahi, S. M., Balvanović, S., Baumgarten, M. & Mörgelin, M. Collagen VI encodes antimicrobial activity: novel innate host defense properties of the extracellular matrix. *J Innate Immun* **4**, 371-376, doi:10.1159/000335239 (2012).

7 Ricard-Blum, S. & Ruggiero, F. The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol (Paris)* **53**, 430-442, doi:10.1016/j.patbio.2004.12.024 (2005).

8 Colombatti, A. *et al.* The EMILIN protein family. *Matrix Biology* **19**, 289-301, doi:http://dx.doi.org/10.1016/S0945-053X(00)00074-3 (2000).

9 Exposito, J. Y., Cluzel, C., Garrone, R. & Lethias, C. Evolution of collagens. *Anat Rec* **268**, 302-316, doi:10.1002/ar.10162 (2002).

10 Deutzmann, R. *et al.* Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in hydra. *Development* **127**, 4669-4680 (2000).

11 Sicot, F. X. *et al.* Molecular adaptation to an extreme environment: origin of the thermal stability of the pompeii worm collagen. *J Mol Biol* **302**, 811-820, doi:10.1006/jmbi.2000.4505 (2000).

12 Yamada, Y. *et al.* The collagen gene: evidence for its evolutinary assembly by amplification of a DNA segment containing an exon of 54 bp. *Cell* **22**, 887-892 (1980).

13 Rasmussen, M., Jacobsson, M. & Björck, L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* **278**, 32313-32316, doi:10.1074/jbc.M304709200 (2003).

14 Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M. & Lukomski, S. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318, doi:10.1074/jbc.M201163200 (2002).

15 Sylvestre, P., Couture-Tosi, E. & Mock, M. A collagen-like surface glycoprotein is a structural component of the Bacillus anthracis exosporium. *Mol Microbiol* **45**, 169-178 (2002).

16 Yang, F. *et al.* Complete genome sequence of the shrimp white spot bacilliform virus. *J Virol* **75**, 11811-11820, doi:10.1128/JVI.75.23.11811-11820.2001 (2001).

17 Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu Rev Biochem* **78**, 929-958, doi:10.1146/annurev.biochem.77.032207.120833 (2009).

18 Ricard-Blum, S. The collagen family. *Cold Spring Harb Perspect Biol* **3**, a004978, doi:10.1101/cshperspect.a004978 (2011).

19 Boot-Handford, R. P., Tuckwell, D. S., Plumb, D. A., Rock, C. F. & Poulsom, R. A novel and highly conserved collagen (pro(alpha)1(XXVII)) with a unique expression pattern and unusual

molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. *J Biol Chem* **278**, 31067-31077, doi:10.1074/jbc.M212889200 (2003).
20 Koch, M. *et al.* Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. *J Biol Chem* **278**, 43236-43244, doi:10.1074/jbc.M302112200 (2003).
21 Adzhubei, A. A. & Sternberg, M. J. Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* **229**, 472-493, doi:10.1006/jmbi.1993.1047 (1993).
22 Adzhubei, A. A., Sternberg, M. J. & Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J Mol Biol* **425**, 2100-2132, doi:10.1016/j.jmb.2013.03.018 (2013).
23 Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231-241 (2000).
24 Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8, doi:10.1126/stke.2003.179.re8 (2003).
25 Rucker, A. L. & Creamer, T. P. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci* **11**, 980-985, doi:10.1110/ps.4550102 (2002).
26 Shi, Z., Woody, R. W. & Kallenbach, N. R. Is polyproline II a major backbone conformation in unfolded proteins? *Adv Protein Chem* **62**, 163-240 (2002).
27 Asher, S. A., Mikhonin, A. V. & Bykov, S. UV Raman demonstrates that alpha-helical polyalanine peptides melt to polyproline II conformations. *J Am Chem Soc* **126**, 8433-8440, doi:10.1021/ja049518j (2004).
28 Blanch, E. W. *et al.* Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. *J Mol Biol* **301**, 553-563, doi:10.1006/jmbi.2000.3981 (2000).
29 Pappu, R. V., Srinivasan, R. & Rose, G. D. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc Natl Acad Sci U S A* **97**, 12565-12570, doi:10.1073/pnas.97.23.12565 (2000).
30 Creamer, T. P. Left-handed polyproline II helix formation is (very) locally driven. *Proteins* **33**, 218-226 (1998).
31 Creamer, T. P. & Campbell, M. N. Determinants of the polyproline II helix from modeling studies. *Adv Protein Chem* **62**, 263-282 (2002).
32 Kelly, M. A. *et al.* Host-guest study of left-handed polyproline II helix formation. *Biochemistry* **40**, 14376-14383 (2001).
33 Bartlett, G. J., Choudhary, A., Raines, R. T. & Woolfson, D. N. n-->pi* interactions in proteins. *Nat Chem Biol* **6**, 615-620, doi:10.1038/nchembio.406 (2010).
34 Hinderaker, M. P. & Raines, R. T. An electronic effect on protein structure. *Protein Sci* **12**, 1188-1194, doi:10.1110/ps.0241903 (2003).
35 Berisio, R. *et al.* Effects of microgravity on the crystal quality of a collagen-like polypeptide. *Acta Crystallographica Section D* **56**, 55-61, doi:doi:10.1107/S0907444999014158 (2000).
36 Berisio, R., Vitagliano, L., Mazzarella, L. & Zagari, A. Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3). *Protein Sci* **11**, 262-270, doi:10.1110/ps.32602 (2002).
37 Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. Structural Basis of Collagen Recognition by Integrin α2β1. *Cell* **101**, 47-56, doi:http://dx.doi.org/10.1016/S0092-8674(00)80622-4 (2000).
38 Leikina, E., Mertts, M. V., Kuznetsova, N. & Leikin, S. Type I collagen is thermally unstable at body temperature. *Proc Natl Acad Sci U S A* **99**, 1314-1318, doi:10.1073/pnas.032307099 (2002).
39 Persikov, A. V. & Brodsky, B. Unstable molecules form stable tissues. *Proc Natl Acad Sci U S A* **99**, 1101-1103, doi:10.1073/pnas.042707899 (2002).

40 Olague-Marchan, M. *et al.* A disease-associated glycine substitution in BP180 (type XVII collagen) leads to a local destabilization of the major collagen triple helix. *Matrix Biology* **19**, 223-233, doi:http://dx.doi.org/10.1016/S0945-053X(00)00070-6 (2000).
41 Kuivaniemi, H., Tromp, G. & Prockop, D. J. Mutations in collagen genes: causes of rare and some common diseases in humans. *FASEB J* **5**, 2052-2060 (1991).
42 Beck, K. *et al.* Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc Natl Acad Sci U S A* **97**, 4273-4278, doi:10.1073/pnas.070050097 (2000).
43 Cabral, W. A., Milgrom, S., Letocha, A. D., Moriarty, E. & Marini, J. C. Biochemical screening of type I collagen in osteogenesis imperfecta: detection of glycine substitutions in the amino end of the alpha chains requires supplementation by molecular analysis. *J Med Genet* **43**, 685-690, doi:10.1136/jmg.2005.040493 (2006).
44 Hyde, T. J., Bryan, M. A., Brodsky, B. & Baum, J. Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J Biol Chem* **281**, 36937-36943, doi:10.1074/jbc.M605135200 (2006).
45 Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A. & Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960-14967 (2000).
46 Vitagliano, L., Némethy, G., Zagari, A. & Scheraga, H. A. Stabilization of the triple-helical structure of natural collagen by side-chain interactions. *Biochemistry* **32**, 7354-7359 (1993).
47 Ramshaw, J. A., Shah, N. K. & Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol* **122**, 86-91, doi:10.1006/jsbi.1998.3977 (1998).
48 Burjanadze, T. V. New analysis of the phylogenetic change of collagen thermostability. *Biopolymers* **53**, 523-528, doi:10.1002/(SICI)1097-0282(200005)53:6<523::AID-BIP8>3.0.CO;2-7 (2000).
49 Schumacher, M. A., Mizuno, K. & Bächinger, H. P. The Crystal Structure of a Collagen-like Polypeptide with 3(*S*)-Hydroxyproline Residues in the Xaa Position Forms a Standard 7/2 Collagen Triple Helix. *J Biol Chem* **281**, 27566-27574, doi: 10.1074/jbc.M602797200
50 Vitagliano, L., Berisio, R., Mazzarella, L. & Zagari, A. Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58**, 459-464, doi:10.1002/1097-0282(20010415)58:5<459::AID-BIP1021>3.0.CO;2-V (2001).
51 Holmgren, S. K., Taylor, K. M., Bretscher, L. E. & Raines, R. T. Code for collagen's stability deciphered. *Nature* **392**, 666-667, doi:10.1038/33573 (1998).
52 DeRider, M. L. *et al.* Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J Am Chem Soc* **124**, 2497-2505 (2002).
53 Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L. & Raines, R. T. Conformational stability of collagen relies on a stereoelectronic effect. *J Am Chem Soc* **123**, 777-778 (2001).
54 Bella, J. Collagen structure: new tricks from a very old dog. *Biochem J* **473**, 1001-1025, doi:10.1042/BJ20151169 (2016).
55 Shoulders, M. D. & Raines, R. T. Modulating collagen triple-helix stability with 4-chloro, 4-fluoro, and 4-methylprolines. *Adv Exp Med Biol* **611**, 251-252 (2009).
56 Eberhardt, E. S., Panisik, N. & Raines, R. T. Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *J Am Chem Soc* **118**, 12261-12266, doi:10.1021/ja9623119 (1996).
57 Gorres, K. L. & Raines, R. T. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* **45**, 106-124, doi:10.3109/10409231003627991 (2010).
58 Ishikawa, Y. & Bächinger, H. P. A molecular ensemble in the rER for procollagen maturation. *Biochim Biophys Acta* **1833**, 2479-2491, doi:10.1016/j.bbamcr.2013.04.008 (2013).

59 Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K. & Bächinger, H. P. Biochemical characterization of the prolyl 3-hydroxylase 1.cartilage-associated protein.cyclophilin B complex. *J Biol Chem* **284**, 17641-17647, doi:10.1074/jbc.M109.007070 (2009).
60 Duran, I. *et al.* A Chaperone Complex Formed by HSP47, FKBP65, and BiP Modulates Telopeptide Lysyl Hydroxylation of Type I Procollagen. *J Bone Miner Res* **32**, 1309-1319, doi:10.1002/jbmr.3095 (2017).
61 John, D. C., Grant, M. E. & Bulleid, N. J. Cell-free synthesis and assembly of prolyl 4-hydroxylase: the role of the beta-subunit (PDI) in preventing misfolding and aggregation of the alpha-subunit. *EMBO J* **12**, 1587-1595 (1993).
62 Wilson, R., Lees, J. F. & Bulleid, N. J. Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *J Biol Chem* **273**, 9637-9643 (1998).
63 Yamauchi, M. & Sricholpech, M. Lysine post-translational modifications of collagen. *Essays Biochem* **52**, 113-133, doi:10.1042/bse0520113 (2012).
64 Gjaltema, R. A., van der Stoel, M. M., Boersema, M. & Bank, R. A. Disentangling mechanisms involved in collagen pyridinoline cross-linking: The immunophilin FKBP65 is critical for dimerization of lysyl hydroxylase 2. *Proc Natl Acad Sci U S A* **113**, 7142-7147, doi:10.1073/pnas.1600074113 (2016).
65 Lindert, U. *et al.* Molecular Consequences of the SERPINH1/HSP47 Mutation in the Dachshund Natural Model of Osteogenesis Imperfecta. *J Biol Chem* **290**, 17679-17689, doi:10.1074/jbc.M115.661025 (2015).
66 Jürgensen, H. J. *et al.* A novel functional role of collagen glycosylation: interaction with the endocytic collagen receptor uparap/ENDO180. *J Biol Chem* **286**, 32736-32748, doi:10.1074/jbc.M111.266692 (2011).
67 Prockop, D. J. & Kivirikko, K. I. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* **64**, 403-434, doi:10.1146/annurev.bi.64.070195.002155 (1995).
68 Lamandé, S. R. & Bateman, J. F. The type I collagen pro alpha 1(I) COOH-terminal propeptide N-linked oligosaccharide. Functional analysis by site-directed mutagenesis. *J Biol Chem* **270**, 17858-17865 (1995).
69 Smedsrød, B., Melkko, J., Risteli, L. & Risteli, J. Circulating C-terminal propeptide of type I procollagen is cleared mainly via the mannose receptor in liver endothelial cells. *Biochem J* **271**, 345-350 (1990).
70 Ishikawa, Y. *et al.* Mutation in cyclophilin B that causes hyperelastosis cutis in American Quarter Horse does not affect peptidylprolyl cis-trans isomerase activity but shows altered cyclophilin B-protein interactions and affects collagen folding. *J Biol Chem* **287**, 22253-22265, doi:10.1074/jbc.M111.333336 (2012).
71 Ishikawa, Y. & Bächinger, H. P. A substrate preference for the rough endoplasmic reticulum resident protein FKBP22 during collagen biosynthesis. *J Biol Chem* **289**, 18189-18201, doi:10.1074/jbc.M114.561944 (2014).
72 Ishikawa, Y., Vranka, J., Wirz, J., Nagata, K. & Bächinger, H. P. The rough endoplasmic reticulum-resident FK506-binding protein FKBP65 is a molecular chaperone that interacts with collagens. *J Biol Chem* **283**, 31584-31590, doi:10.1074/jbc.M802535200 (2008).
73 Schwarze, U. *et al.* Mutations in FKBP10, which result in Bruck syndrome and recessive forms of osteogenesis imperfecta, inhibit the hydroxylation of telopeptide lysines in bone collagen. *Hum Mol Genet* **22**, 1-17, doi:10.1093/hmg/dds371 (2013).
74 Chen, Y. *et al.* FKBP65-dependent peptidyl-prolyl isomerase activity potentiates the lysyl hydroxylase 2-driven collagen cross-link switch. *Sci Rep* **7**, 46021, doi:10.1038/srep46021 (2017).
75 Khoshnoodi, J., Pedchenko, V. & Hudson, B. G. Mammalian collagen IV. *Microsc Res Tech* **71**, 357-370, doi:10.1002/jemt.20564 (2008).

76 Beck, K., Boswell, B. A., Ridgway, C. C. & Bächinger, H. P. Triple helix formation of procollagen type I can occur at the rough endoplasmic reticulum membrane. *J Biol Chem* **271**, 21566-21573 (1996).
77 Koide, T. & Nagata, K. in *Collagen: Primer in Structure, Processing and Assembly* (eds Jürgen Brinckmann, Holger Notbohm, & P. K. Müller) 85-114 (Springer Berlin Heidelberg, 2005).
78 Engel, J. & Prockop, D. J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Biophys Chem* **20**, 137-152, doi:10.1146/annurev.bb.20.060191.001033 (1991).
79 Boudko, S. P. *et al.* The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J Biol Chem* **285**, 23721-23731, doi:10.1074/jbc.M110.128405 (2010).
80 Jin, L. *et al.* Ubiquitin-dependent regulation of COPII coat size and function. *Nature* **482**, 495-500, doi:10.1038/nature10822 (2012).
81 Saito, K. *et al.* TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. *Cell* **136**, 891-902, doi:10.1016/j.cell.2008.12.025 (2009).
82 Saito, K. & Katada, T. Mechanisms for exporting large-sized cargoes from the endoplasmic reticulum. *Cell Mol Life Sci* **72**, 3709-3720, doi:10.1007/s00018-015-1952-9 (2015).
83 Ishikawa, Y., Ito, S., Nagata, K., Sakai, L. Y. & Bächinger, H. P. Intracellular mechanisms of molecular recognition and sorting for transport of large extracellular matrix molecules. *Proc Natl Acad Sci U S A* **113**, E6036-E6044, doi:10.1073/pnas.1609571113 (2016).
84 Wilson, D. G. *et al.* Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J Cell Biol* **193**, 935-951, doi:10.1083/jcb.201007162 (2011).
85 Polishchuk, E. V., Di Pentima, A., Luini, A. & Polishchuk, R. S. Mechanism of constitutive export from the golgi: bulk flow via the formation, protrusion, and en bloc cleavage of large trans-golgi network tubular domains. *Mol Biol Cell* **14**, 4470-4485, doi:10.1091/mbc.E03-01-0033 (2003).
86 Canty-Laird, E. G., Lu, Y. & Kadler, K. E. Stepwise proteolytic activation of type I procollagen to collagen within the secretory pathway of tendon fibroblasts in situ. *Biochem J* **441**, 707-717, doi:10.1042/BJ20111379 (2012).
87 Banos, C. C., Thomas, A. H. & Kuo, C. K. Collagen fibrillogenesis in tendon development: current models and regulation of fibril assembly. *Birth Defects Res C Embryo Today* **84**, 228-244, doi:10.1002/bdrc.20130 (2008).
88 Kadler, K. E., Hill, A. & Canty-Laird, E. G. Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. *Curr Opin Cell Biol* **20**, 495-501, doi:10.1016/j.ceb.2008.06.008 (2008).
89 Bruckner, P. Suprastructures of extracellular matrices: paradigms of functions controlled by aggregates rather than molecules. *Cell Tissue Res* **339**, 7-18, doi:10.1007/s00441-009-0864-0 (2010).
90 Herchenhan, A. *et al.* Lysyl Oxidase Activity Is Required for Ordered Collagen Fibrillogenesis by Tendon Cells. *J Biol Chem* **290**, 16440-16450, doi:10.1074/jbc.M115.641670 (2015).
91 Nagai, N. *et al.* Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J Cell Biol* **150**, 1499-1506 (2000).
92 Christiansen, H. E. *et al.* 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, doi:10.1016/j.ajhg.2010.01.034 (2010).
93 Ito, S. & Nagata, K. Mutants of collagen-specific molecular chaperone Hsp47 causing osteogenesis imperfecta are structurally unstable with weak binding affinity to collagen. *Biochem Biophys Res Commun* **469**, 437-442, doi:10.1016/j.bbrc.2015.12.028 (2016).
94 Taguchi, T. & Razzaque, M. S. The collagen-specific molecular chaperone HSP47: is there a role in fibrosis? *Trends Mol Med* **13**, 45-53, doi:10.1016/j.molmed.2006.12.001 (2007).

95 Dafforn, T. R., Della, M. & Miller, A. D. The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J Biol Chem* **276**, 49310-49319, doi:10.1074/jbc.M108896200 (2001).
96 Nagata, K. Hsp47: a collagen-specific molecular chaperone. *Trends Biochem Sci* **21**, 22-26 (1996).
97 Oguro, A. *et al.* The molecular chaperone HSP47 rapidly senses gravitational changes in myoblasts. *Genes Cells* **11**, 1253-1265, doi:10.1111/j.1365-2443.2006.01021.x (2006).
98 Natsume, T., Koide, T., Yokota, S., Hirayoshi, K. & Nagata, K. Interactions between collagen-binding stress protein HSP47 and collagen. Analysis of kinetic parameters by surface plasmon resonance biosensor. *J Biol Chem* **269**, 31224-31228 (1994).
99 Annand, R. R. *et al.* Caspase-1 (interleukin-1beta-converting enzyme) is inhibited by the human serpin analogue proteinase inhibitor 9. *Biochem J* **342 Pt 3**, 655-665 (1999).
100 Hwang, S. R. *et al.* The novel serpin endopin 2 demonstrates cross-class inhibition of papain and elastase: localization of endopin 2 to regulated secretory vesicles of neuroendocrine chromaffin cells. *Biochemistry* **41**, 10397-10405 (2002).
101 Ligoxygakis, P. *et al.* A serpin mutant links Toll activation to melanization in the host defence of Drosophila. *EMBO J* **21**, 6330-6337 (2002).
102 Pike, R. N., Buckle, A. M., le Bonniec, B. F. & Church, F. C. Control of the coagulation system by serpins. Getting by with a little help from glycosaminoglycans. *FEBS J* **272**, 4842-4851, doi:10.1111/j.1742-4658.2005.04880.x (2005).
103 Gettins, P. G. W. Serpin Structure, Mechanism, and Function. *ChemInform* **34**, no-no, doi:10.1002/chin.200307260 (2003).
104 Gettins, P. G. W. The F-helix of serpins plays an essential, active role in the proteinase inhibition mechanism. *FEBS Letters* **523**, 2-6, doi:10.1016/S0014-5793(02)02924-1 (2002).
105 Wei, Z., Yan, Y., Carrell, R. W. & Zhou, A. Crystal structure of protein Z-dependent inhibitor complex shows how protein Z functions as a cofactor in the membrane inhibition of factor X. *Blood* **114**, 3662-3667, doi:10.1182/blood-2009-04-210021 (2009).
106 Rein, C. M., Desai, U. R. & Church, F. C. Serpin–Glycosaminoglycan Interactions. *Methods in Enzymology* **501**, 105-137, doi:http://dx.doi.org/10.1016/B978-0-12-385950-1.00007-9 (2011).
107 Im, H., Seo, E. J. & Yu, M. H. Metastability in the inhibitory mechanism of human alpha1-antitrypsin. *J Biol Chem* **274**, 11072-11077 (1999).
108 Huntington, J. A., Read, R. J. & Carrell, R. W. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**, 923-926, doi:http://www.nature.com/nature/journal/v407/n6806/suppinfo/407923a0_S1.html (2000).
109 Thomson, C. A., Tenni, R. & Ananthanarayanan, V. S. Mapping Hsp47 binding site(s) using CNBr peptides derived from type I and type II collagen. *Protein Sci* **12**, 1792-1800, doi:10.1110/ps.0236903 (2003).
110 Satoh, M., Hirayoshi, K., Yokota, S., Hosokawa, N. & Nagata, K. Intracellular interaction of collagen-specific stress protein HSP47 with newly synthesized procollagen. *J Cell Biol* **133**, 469-483 (1996).
111 Paroutis, P., Touret, N. & Grinstein, S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology (Bethesda)* **19**, 207-215, doi:10.1152/physiol.00005.2004 (2004).
112 Saga, S., Nagata, K., Chen, W. T. & Yamada, K. M. pH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein. *The Journal of Cell Biology* **105**, 517 (1987).
113 Sauk, J. J., Norris, K., Hebert, C., Ordónˆtez, J. & Reynolds, M. Hsp47 Binds to the KDEL Receptor and Cell Surface Expression is Modulated by Cytoplasmic and Endosomal pH. *Connective Tissue Research* **37**, 105-119, doi:10.3109/03008209809028904 (1998).

114 Asada, S., Koide, T., Yasui, H. & Nagata, K. Effect of HSP47 on prolyl 4-hydroxylation of collagen model peptides. *Cell Struct Funct* **24**, 187-196 (1999).
115 Macdonald, J. R. & Bächinger, H. P. 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**, 25399-25403, doi:10.1074/jbc.M102471200 (2001).
116 Thomson, C. A. & Ananthanarayanan, V. S. Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro. *Biochem J* **349 Pt 3**, 877-883 (2000).
117 Ono, T., Miyazaki, T., Ishida, Y., Uehata, M. & Nagata, K. Direct in vitro and in vivo evidence for interaction between Hsp47 protein and collagen triple helix. *J Biol Chem* **287**, 6810-6818, doi:10.1074/jbc.M111.280248 (2012).
118 Koide, T., Takahara, Y., Asada, S. & Nagata, K. Xaa-Arg-Gly triplets in the collagen triple helix are dominant binding sites for the molecular chaperone HSP47. *J Biol Chem* **277**, 6178-6182, doi:10.1074/jbc.M106497200 (2002).
119 Koide, T. *et al.* Specific recognition of the collagen triple helix by chaperone HSP47. II. The HSP47-binding structural motif in collagens and related proteins. *J Biol Chem* **281**, 11177-11185, doi:10.1074/jbc.M601369200 (2006).
120 Koide, T., Asada, S. & Nagata, K. Substrate recognition of collagen-specific molecular chaperone HSP47. Structural requirements and binding regulation. *J Biol Chem* **274**, 34523-34526 (1999).
121 Koide, T. *et al.* Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation. *J Biol Chem* **281**, 3432-3438, doi:10.1074/jbc.M509707200 (2006).
122 Widmer, C. *et al.* Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc Natl Acad Sci U S A* **109**, 13243-13247, doi:10.1073/pnas.1208072109 (2012).
123 Liu, T., Ryan, M., Dahlquist, F. W. & Griffith, O. H. Determination of pKa values of the histidine side chains of phosphatidylinositol-specific phospholipase C from Bacillus cereus by NMR spectroscopy and site-directed mutagenesis. *Protein Sci* **6**, 1937-1944, doi:10.1002/pro.5560060914 (1997).
124 Nanthakumar, C. B. *et al.* Dissecting fibrosis: therapeutic insights from the small-molecule toolbox. *Nat Rev Drug Discov* **14**, 693-720, doi:10.1038/nrd4592 (2015).
125 Cavaluzzi, M. J. & Borer, P. N. Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* **32**, e13, doi:10.1093/nar/gnh015 (2004).
126 Zheng, L., Baumann, U. & Reymond, J. L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115, doi:10.1093/nar/gnh110 (2004).
127 Panja, S., Aich, P., Jana, B. & Basu, T. How does plasmid DNA penetrate cell membranes in artificial transformation process of Escherichia coli? *Mol Membr Biol* **25**, 411-422, doi:10.1080/09687680802187765 (2008).
128 Dagert, M. & Ehrlich, S. D. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. *Gene* **6**, 23-28 (1979).
129 Tabor, S. & Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci U S A* **82**, 1074-1078 (1985).
130 Schmidt, T. G. & Skerra, A. The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Eng* **6**, 109-122 (1993).
131 Schmidt, T. G. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2**, 1528-1535, doi:10.1038/nprot.2007.209 (2007).

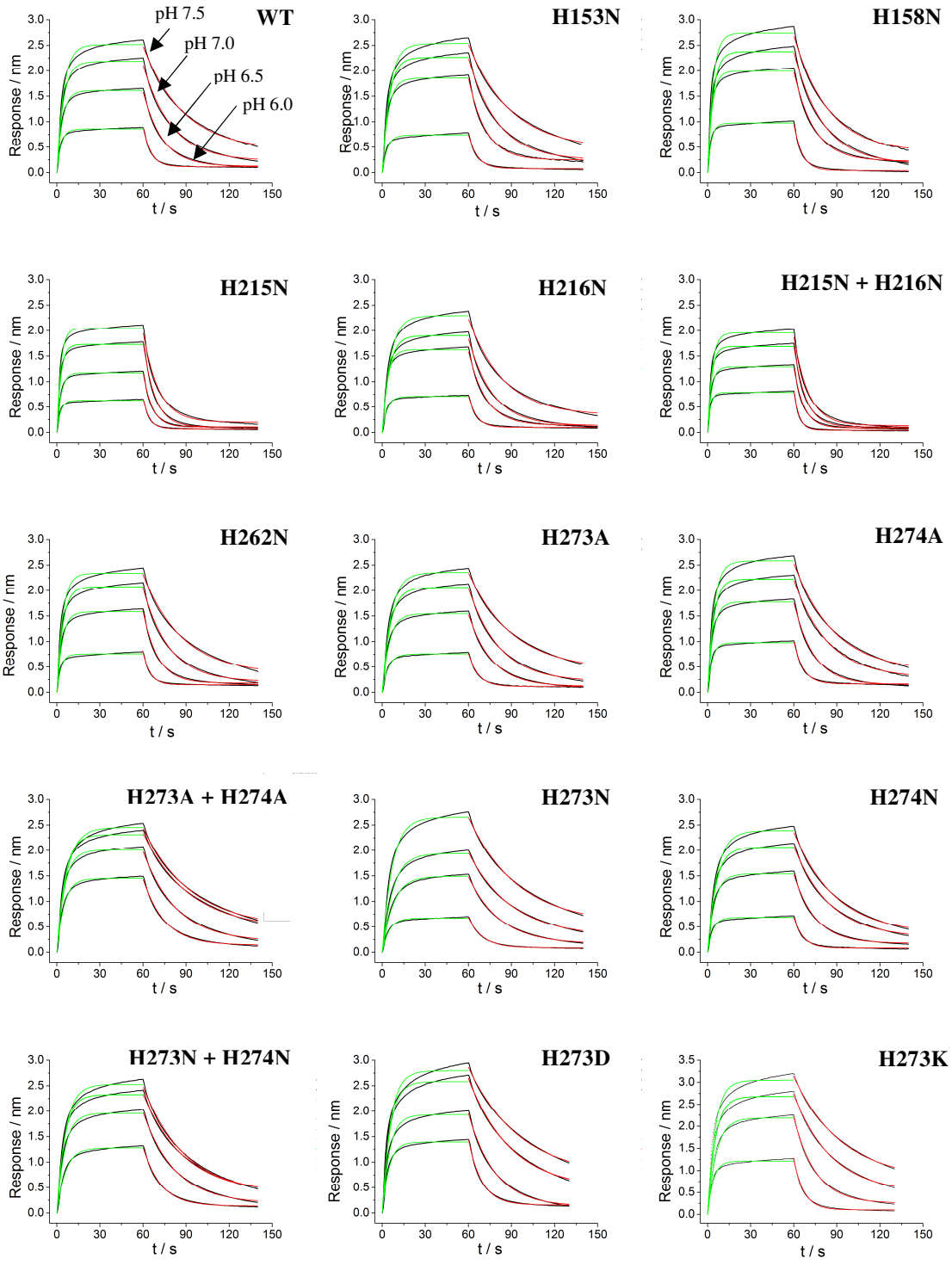
132 Ladner, C. L., Yang, J., Turner, R. J. & Edwards, R. A. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem* **326**, 13-20, doi:10.1016/j.ab.2003.10.047 (2004).
133 Mruk, D. D. & Cheng, C. Y. Enhanced chemiluminescence (ECL) for routine immunoblotting: An inexpensive alternative to commercially available kits. *Spermatogenesis* **1**, 121-122, doi:10.4161/spmg.1.2.16606 (2011).
134 Morpurgo, M., Bayer, E. A. & Wilchek, M. N-hydroxysuccinimide carbonates and carbamates are useful reactive reagents for coupling ligands to lysines on proteins. *Journal of Biochemical and Biophysical Methods* **38**, 17-28, doi:http://dx.doi.org/10.1016/S0165-022X(98)00027-X (1999).
135 Miller, B. T., Collins, T. J., Rogers, M. E. & Kurosky, A. Peptide Biotinylation with Amine-Reactive Esters: Differential Side Chain Reactivity. *Peptides* **18**, 1585-1595, doi:http://dx.doi.org/10.1016/S0196-9781(97)00225-8 (1997).
136 Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411-2423, doi:10.1002/pro.5560041120 (1995).
137 Kuipers, B. J. & Gruppen, H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. *J Agric Food Chem* **55**, 5445-5451, doi:10.1021/jf070337l (2007).
138 Simpson, R. J. SYPRO Orange fluorescent staining of protein gels. *Cold Spring Harb Protoc* **2010**, pdb.prot5414, doi:10.1101/pdb.prot5414 (2010).
139 Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* **11**, 32-51, doi:10.1007/s12575-009-9008-x (2009).
140 Zwier, J. M., Bazin, H., Lamarque, L. & Mathis, G. Luminescent Lanthanide Cryptates: from the Bench to the Bedside. *Inorganic Chemistry* **53**, 1854-1866, doi:10.1021/ic402234k (2014).
141 Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening* **4**, 67-73, doi:10.1177/108705719900400206 (1999).
142 Frank, S. *et al.* Stabilization of short collagen-like triple helices by protein engineering. *J Mol Biol* **308**, 1081-1089, doi:10.1006/jmbi.2001.4644 (2001).
143 Price, W. N. *et al.* Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol* **27**, 51-57, doi:10.1038/nbt.1514 (2009).
144 Beeby, A. *et al.* Non-radiative deactivation of the excited states of europium, terbium and ytterbium complexes by proximate energy-matched OH, NH and CH oscillators: an improved luminescence method for establishing solution hydration states. *Journal of the Chemical Society, Perkin Transactions 2*, 493-504, doi:10.1039/A808692C (1999).
145 Stubenrauch, K. *et al.* Characterization of murine anti-human Fab antibodies for use in an immunoassay for generic quantification of human Fab fragments in non-human serum samples including cynomolgus monkey samples. *J Pharm Biomed Anal* **72**, 208-215, doi:10.1016/j.jpba.2012.08.023 (2013).
146 Green, N. M. Avidin and streptavidin. *Methods Enzymol* **184**, 51-67 (1990).
147 Meyer, C., Notari, L. & Becerra, S. P. Mapping the type I collagen-binding site on pigment epithelium-derived factor. Implications for its antiangiogenic activity. *J Biol Chem* **277**, 45400-45407, doi:10.1074/jbc.M208339200 (2002).
148 Huntington, J. A. *et al.* A 2.6 A structure of a serpin polymer and implications for conformational disease. *J Mol Biol* **293**, 449-455, doi:10.1006/jmbi.1999.3184 (1999).

149 El-Thaher, T. S. H. in *Protein & Peptide Letters* Vol. 3 1-8 (Bentham Science Publishers B. V., 1996).
150 Abdul-Wahab, M. F. *et al.* The pH sensitivity of murine heat shock protein 47 (HSP47) binding to collagen is affected by mutations in the breach histidine cluster. *J Biol Chem* **288**, 4452-4461, doi:10.1074/jbc.M112.409029 (2013).
151 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).
152 Matsuo, K., Sakurada, Y., Yonehara, R., Kataoka, M. & Gekko, K. Secondary-structure analysis of denatured proteins by vacuum-ultraviolet circular dichroism spectroscopy. *Biophys J* **92**, 4088-4096, doi:10.1529/biophysj.106.103515 (2007).
153 Hulme, E. C. & Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *British Journal of Pharmacology* **161**, 1219-1237, doi:10.1111/j.1476-5381.2009.00604.x (2010).
154 Slatter, D. A., Miles, C. A. & Bailey, A. J. Asymmetry in the Triple Helix of Collagen-like Heterotrimers Confirms that External Bonds Stabilize Collagen Structure. *Journal of Molecular Biology* **329**, 175-183, doi:http://dx.doi.org/10.1016/S0022-2836(03)00380-2 (2003).
155 Gauba, V. & Hartgerink, J. D. Self-Assembled Heterotrimeric Collagen Triple Helices Directed through Electrostatic Interactions. *Journal of the American Chemical Society* **129**, 2683-2690, doi:10.1021/ja0683640 (2007).
156 Li, Y., Mo, X., Kim, D. & Yu, S. M. Template-Tethered Collagen Mimetic Peptides for Studying Heterotrimeric Triple-Helical Interactions. *Biopolymers* **95**, 94-104, doi:10.1002/bip.21536 (2011).
157 Ottl, J. & Moroder, L. Disulfide-Bridged Heterotrimeric Collagen Peptides Containing the Collagenase Cleavage Site of Collagen Type I. Synthesis and Conformational Properties. *Journal of the American Chemical Society* **121**, 653-661, doi:10.1021/ja983456d (1999).
158 Sharma, U. *et al.* Structural basis of homo- and heterotrimerization of collagen I. **8**, 14671, doi:10.1038/ncomms14671 https://www.nature.com/articles/ncomms14671#supplementary-information (2017).
159 Bourhis, J.-M. *et al.* Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nat Struct Mol Biol* **19**, 1031-1036, doi:http://www.nature.com/nsmb/journal/v19/n10/abs/nsmb.2389.html#supplementary-information (2012).
160 Sundaramoorthy, M., Meiyappan, M., Todd, P. & Hudson, B. G. Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. *J Biol Chem* **277**, 31142-31153, doi:10.1074/jbc.M201740200 (2002).
161 Kvansakul, M., Bogin, O., Hohenester, E. & Yayon, A. Crystal structure of the collagen alpha1(VIII) NC1 trimer. *Matrix Biol* **22**, 145-152 (2003).
162 Bogin, O. *et al.* Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* **10**, 165-173 (2002).
163 Wirz, J. A., Boudko, S. P., Lerch, T. F., Chapman, M. S. & Bächinger, H. P. Crystal Structure of the Human Collagen XV Trimerization Domain: A Potent Trimerizing Unit Common to Multiplexin Collagens. *Matrix biology : journal of the International Society for Matrix Biology* **30**, 9-15, doi:10.1016/j.matbio.2010.09.005 (2011).
164 Boudko, S. P. *et al.* Crystal Structure of Human Collagen XVIII Trimerization Domain: A Novel Collagen Trimerization Fold. *Journal of Molecular Biology* **392**, 787-802, doi:http://dx.doi.org/10.1016/j.jmb.2009.07.057 (2009).
165 Vanacore, R. *et al.* A Sulfilimine Bond Identified in Collagen IV. *Science* **325**, 1230 (2009).
166 Boudko, S. P., Engel, J. & Bächinger, H. P. Trimerization and triple helix stabilization of the collagen XIX NC2 domain. *J Biol Chem* **283**, 34345-34351, doi:10.1074/jbc.M806352200 (2008).

167 Boudko, S. P. & Bächinger, H. P. The NC2 domain of type IX collagen determines the chain register of the triple helix. *J Biol Chem* **287**, 44536-44545, doi:10.1074/jbc.M112.417543 (2012).
168 Rockey, D. C., Bell, P. D. & Hill, J. A. Fibrosis--a common pathway to organ injury and failure. *N Engl J Med* **372**, 1138-1149, doi:10.1056/NEJMra1300575 (2015).
169 Bringardner, B. D., Baran, C. P., Eubank, T. D. & Marsh, C. B. The role of inflammation in the pathogenesis of idiopathic pulmonary fibrosis. *Antioxid Redox Signal* **10**, 287-301, doi:10.1089/ars.2007.1897 (2008).
170 Cheresh, P., Kim, S. J., Tulasiram, S. & Kamp, D. W. Oxidative stress and pulmonary fibrosis. *Biochim Biophys Acta* **1832**, 1028-1040, doi:10.1016/j.bbadis.2012.11.021 (2013).
171 Morry, J. *et al.* Dermal delivery of HSP47 siRNA with NOX4-modulating mesoporous silica-based nanoparticles for treating fibrosis. *Biomaterials* **66**, 41-52, doi:10.1016/j.biomaterials.2015.07.005 (2015).
172 Naitoh, M. *et al.* Upregulation of HSP47 and collagen type III in the dermal fibrotic disease, keloid. *Biochem Biophys Res Commun* **280**, 1316-1322, doi:10.1006/bbrc.2001.4257 (2001).
173 Yang, S. F., Tsai, C. H. & Chang, Y. C. The upregulation of heat shock protein 47 expression in human buccal fibroblasts stimulated with arecoline. *J Oral Pathol Med* **37**, 206-210, doi:10.1111/j.1600-0714.2007.00633.x (2008).
174 Razzaque, M. S., Nazneen, A. & Taguchi, T. Immunolocalization of collagen and collagen-binding heat shock protein 47 in fibrotic lung diseases. *Mod Pathol* **11**, 1183-1188 (1998).
175 Kuroda, K., Tsukifuji, R. & Shinkai, H. Increased expression of heat-shock protein 47 is associated with overproduction of type I procollagen in systemic sclerosis skin fibroblasts. *J Invest Dermatol* **111**, 1023-1028, doi:10.1046/j.1523-1747.1998.00437.x (1998).
176 Sato, Y. *et al.* Resolution of liver cirrhosis using vitamin A-coupled liposomes to deliver siRNA against a collagen-specific chaperone. *Nat Biotechnol* **26**, 431-442, doi:10.1038/nbt1396 (2008).
177 Ishiwatari, H. *et al.* Treatment of pancreatic fibrosis with siRNA against a collagen-specific chaperone in vitamin A-coupled liposomes. *Gut* **62**, 1328-1339, doi:10.1136/gutjnl-2011-301746 (2013).
178 Thomson, C. A., Atkinson, H. M. & Ananthanarayanan, V. S. Identification of Small Molecule Chemical Inhibitors of the Collagen-Specific Chaperone Hsp47. *Journal of Medicinal Chemistry* **48**, 1680-1684, doi:10.1021/jm049148+ (2005).
179 Andrews, J. P., Marttala, J., Macarak, E., Rosenbloom, J. & Uitto, J. Keloids: The paradigm of skin fibrosis — Pathomechanisms and treatment. *Matrix Biology* **51**, 37-46, doi:http://dx.doi.org/10.1016/j.matbio.2016.01.013 (2016).
180 Yamaguchi, Y. *et al.* A Peptide Derived from Endostatin Ameliorates Organ Fibrosis. *Science Translational Medicine* **4**, 136ra171 (2012).
181 Nishimoto, T., Mlakar, L., Takihara, T. & Feghali-Bostwick, C. An endostatin-derived peptide orally exerts anti-fibrotic activity in a murine pulmonary fibrosis model. *Int Immunopharmacol* **28**, 1102-1105, doi:10.1016/j.intimp.2015.07.039 (2015).
182 Nakayama, S. *et al.* Pirfenidone inhibits the expression of HSP47 in TGF-β1-stimulated human lung fibroblasts. *Life Sciences* **82**, 210-217, doi:http://dx.doi.org/10.1016/j.lfs.2007.11.003 (2008).
183 Katarkar, A., Haldar, P. K. & Chaudhuri, K. De novo design based pharmacophore query generation and virtual screening for the discovery of Hsp-47 inhibitors. *Biochem Biophys Res Commun* **456**, 707-713, doi:10.1016/j.bbrc.2014.12.051 (2015).
184 Oecal S, Socher E, Uthoff M, et al. The pH-dependent Client Release from the Collagen-specific Chaperone HSP47 Is Triggered by a Tandem Histidine Pair. *J Biol Chem* **291(24)**, 12612-12626 (2016)

Appendix

A.1 Figures



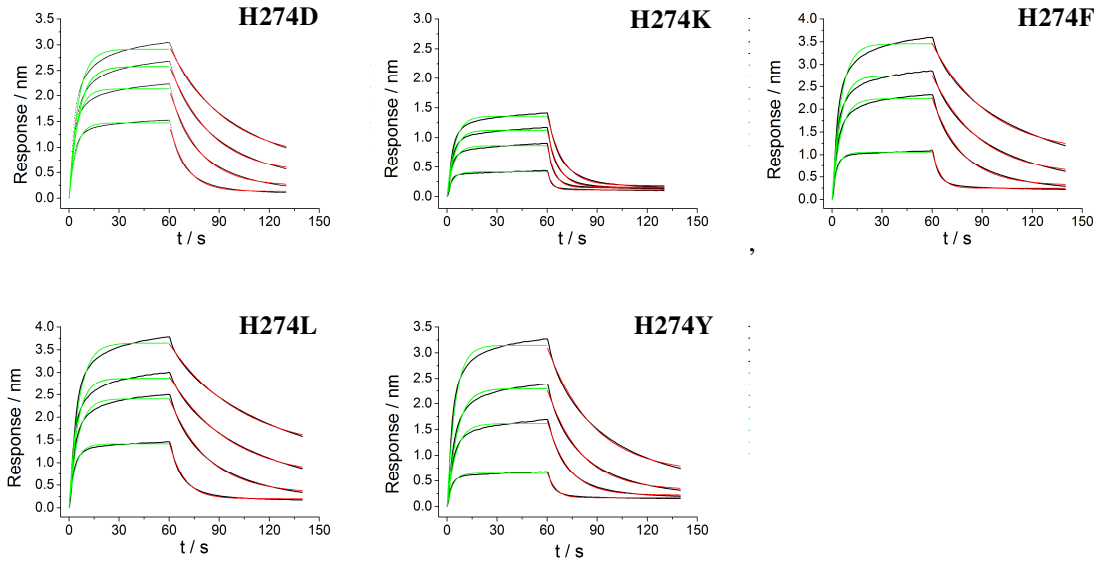


Figure A.1 | Sensorgrams of the interaction of wild-type and mutant HSP47 with CMP-R18 at various pH. The binding of 2.5µM HSP47 to immobilized CMP-R18 was monitored via BLI at various pH. The pH corresponding to the individual binding curves is indicated for the wild-type. The association and dissociation phases were fitted separately using a monophasic exponential function *(green and red curves)* to yield the k_{on} and k_{off} values, respectively.

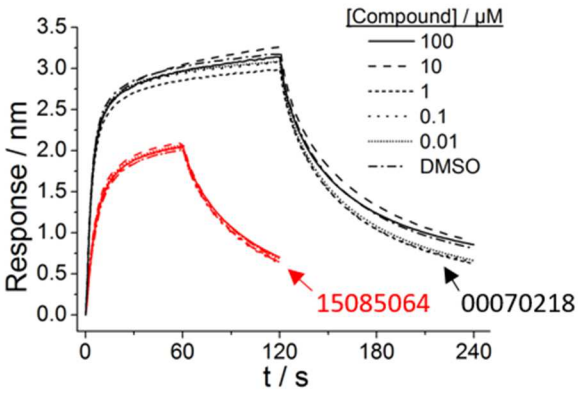


Figure A.2 | BLI testing of inhibitor efficacy. The binding HSP47 to immobilized CMP-R18 was monitored in presence of a serial dilution of compounds reported to inhibit HSP47 - collagen complex formation by Katarkar *et al.*[183] *(red)* or Thomson *et al.*[178]*(black).* The concentration of HSP47 was 1.5µM and 2.5µM, respectively. The 8-digit numbers represent the ZINC database ID of the compounds. An effect of the compounds on the HSP47 - CMP interaction was not observed.

A.2 Tables

Table A.1 | Client-dissociation rates of HSP47 mutants at various pH values

	k_{off} / s^{-1}			
	pH 7.5	**pH 7.0**	**pH 6.5**	**pH 6.0**
WT	0.028 ± 0.000	0.042 ± 0.001	0.075 ± 0.001	0.192 ± 0.002
H153N	0.029 ± 0.001	0.045 ± 0.001	0.074 ± 0.006	0.179 ± 0.003
H158N	0.031 ± 0.003	0.045 ± 0.001	0.070 ± 0.002	0.187 ± 0.002
H215N	0.097 ± 0.005	0.126 ± 0.005	0.159 ± 0.015	0.313 ± 0.023
H216N	0.039 ± 0.001	0.055 ± 0.003	0.088 ± 0.004	0.184 ± 0.013
H215N + H216N	0.094 ± 0.002	0.124 ± 0.003	0.168 ± 0.019	0.273 ± 0.005
H238N	n.d.	n.d.	n.d.	n.d.
H262N	0.030 ± 0.001	0.047 ± 0.004	0.096 ± 0.002	0.236 ± 0.018
H273N	0.028 ± 0.001	0.038 ± 0.003	0.063 ± 0.003	0.148 ± 0.009
H274N	0.029 ± 0.001	0.038 ± 0.001	0.075 ± 0.004	0.170 ± 0.010
H273N + H274N	0.024 ± 0.001	0.026 ± 0.001	0.038 ± 0.002	0.071 ± 0.002
H273A	0.023 ± 0.002	0.030 ± 0.002	0.048 ± 0.004	0.111 ± 0.002
H274A	0.029 ± 0.001	0.036 ± 0.002	0.071 ± 0.008	0.169 ± 0.005
H273A + H274A	0.028 ± 0.002	0.031 ± 0.001	0.043 ± 0.002	0.080 ± 0.002
H273D	0.020 ± 0.001	0.028 ± 0.001	0.051 ± 0.002	0.110 ± 0.003
H273K	0.023 ± 0.001	0.034 ± 0.001	0.057 ± 0.005	0.134 ± 0.006
H274D	0.024 ± 0.001	0.032 ± 0.002	0.048 ± 0.004	0.100 ± 0.002
H274K	0.103 ± 0.010	0.124 ± 0.012	0.167 ± 0.006	0.285 ± 0.011
H273F	0.022 ± 0.000	0.029 ± 0.000	0.050 ± 0.002	0.139 ± 0.003
H273L	0.017 ± 0.000	0.024 ± 0.001	0.039 ± 0.001	0.112 ± 0.002
H273Y	0.029 ± 0.001	0.043 ± 0.003	0.076 ± 0.002	0.215 ± 0.001

Table A.2 | Client-association rates of HSP47 mutants at various pH values

	k_{on} / $mM^{-1} \cdot s^{-1}$			
	pH 7.5	**pH 7.0**	**pH 6.5**	**pH 6.0**
WT	35.04±2.31	34.24±0.95	30.82±0.12	28.96±2.48
H153N	38.03±0.93	34.28±0.94	38.88±3.89	29.82±0.58
H158N	56.19±1.50	45.44±3.68	51.33±3.37	49.51±6.59
H215N	61.06±5.44	57.60±5.50	44.56±6.71	35.08±3.68
H216N	32.90±2.81	35.67±1.38	32.48±3.34	23.98±6.29
H215N + H216N	57.86±3.78	57.72±3.75	40.09±10.19	59.57±6.05
H238N	n.d.	n.d.	n.d.	n.d.
H262N	45.31±3.38	42.49±7.74	37.77±1.54	24.67±8.63
H273N	28.67±0.17	30.15 ± 0.67	27.17±2.18	21.75±4.81
H274N	44.19±1.52	41.75±2.26	35.93±2.82	29.33±2.71
H273N + H274N	27.86±2.57	30.80±0.19	30.27±0.77	27.19±0.10
H273A	28.23±2.60	20.38±1.78	18.41±0.40	18.63±3.68
H274A	23.89±2.11	31.99±1.35	30.10±1.56	26.39±6.29
H273A + H274A	26.29±0.63	31.42±1.67	25.74±2.55	27.11±6.05
H273D	31.47±1.48	31.34±0.83	29.37±0.41	27.87±4.40
H273K	33.41±0.36	30.66±0.54	28.76±5.49	26.65±2.46
H274D	32.54±0.33	29.38±3.39	34.48±2.67	25.18±2.68
H274K	21.65±2.91	17.46±0.91	11.415±2.83	5.02±3.63
H273F	35.58±2.08	30.19±5.46	23.38±4.14	25.77±0.34
H273L	37.57±4.94	32.91±2.01	30.89±0.67	29.45±2.33
H273Y	35.07±0.69	28.24±3.22	24.05±0.71	20.84±3.79

Table A.3 | Client-dissociation constants of HSP47 mutants at various pH values

	K_D / μM			
	pH 7.5	**pH 7.0**	**pH 6.5**	**pH 6.0**
WT	0.74 ± 0.02	1.14 ± 0.04	2.25 ± 0.24	6.23 ± 0.88
H153N	0.76 ± 0.02	1.20 ± 0.04	1.93 ± 0.36	6.00 ± 0.21
H158N	0.55 ± 0.05	1.00 ± 0.08	1.38 ± 0.13	3.85 ± 0.52
H215N	1.60 ± 0.19	2.21 ± 0.27	3.70 ± 0.96	8.93 ± 1.20
H216N	1.18 ± 0.08	1.55 ± 0.14	2.76 ± 0.04	7.65 ± 1.93
H215N + H216N	1.63 ± 0.12	2.15 ± 0.12	4.70 ± 1.96	4.58 ± 0.92
H238N	n.d.	n.d.	n.d.	n.d.
H262N	0.67 ± 0.06	1.13 ± 0.20	2.55 ± 0.17	11.39 ± 5.18
H273N	0.97 ± 0.04	1.27 ± 0.13	2.34 ± 0.29	7.27 ± 2.01
H274N	0.65 ± 0.03	0.92 ± 0.05	2.12 ± 0.26	5.86 ± 0.75
H273N + H274N	0.85 ± 0.09	0.86 ± 0.02	1.26 ± 0.09	2.60 ± 0.08
H273A	0.82 ± 0.09	1.48 ± 0.21	2.58 ± 0.16	6.21 ± 1.20
H274A	1.23 ± 0.12	1.12 ± 0.05	2.37 ± 0.19	6.85 ± 1.93
H273A + H274A	1.07 ± 0.07	1.00 ± 0.06	1.69 ± 0.22	3.17 ± 0.92
H273D	0.64 ± 0.04	0.89 ± 0.02	1.74 ± 0.07	4.04 ± 0.58
H273K	0.69 ± 0.02	1.09 ± 0.06	2.11 ± 0.56	5.08 ± 0.67
H274D	0.73 ± 0.04	1.11 ± 0.18	1.40 ± 0.07	4.02 ± 0.44
H274K	2.98 ± 0.28	3.40 ± 0.78	5.17 ± 0.81	16.29 ± 7.56
H273F	0.61 ± 0.04	1.01 ± 0.20	2.20 ± 0.32	5.39 ± 0.20
H273L	0.47 ± 0.06	0.72 ± 0.06	1.28 ± 0.06	3.84 ± 0.26
H273Y	0.83 ± 0.02	1.53 ± 0.24	3.16 ± 0.20	10.37 ± 1.90

Table A.4 | Primers for Site-directed Mutagenesis

Primer	Sequence
H153N-Fw	GCAGAACTACAATTGCGAG
H153N-Rv	CAATTGTAGTTCTGCTTGCTGCT
H158N-Fw	CGAGAACAGCAAAATCAACTTTC
H158N-Rv	GATTTTGCTGTTCTCGCAATTGT
H215N-Fw	AATTTCAATCACAAAATGGTCGAT
H215N-Rv	TTTGTGATTGAATTTCTCATCCCA
H216N-Fw	TTCCACAATAAAATGGTCGATAAT
H216N-Rv	CATTTTATTGTGGAATTTCTCATC
H215N+H216N-Fw	TTCGCCGCCAAAATGGTCGATAATC
H215N+H216N-Rv	CATTTTGGCGGCGAATTTCTCATC
H262N-Fw	GGCCAACAAGCTGTCCTCTC
H262N-Rv	GCTTGTTGGCCAGCGGC
H273A-Fw	ATGCCGGCTCACGTGGAGCCGCTG
H273A-Rv	CACGTGAGCCGGCATCAAGATGATC
H273D-Fw	ATGCCGGACCACGTGGAGCC
H273D-Rv	ACGTGGTCCGGCATCAAGATGATC
H273F-Fw	ATGCCGTTTCACGTGGAGCC
H273F-Rv	CACGTGAAACGGCATCAAGATG
H273K-Fw	ATGCCGAAGCACGTGGAGCC
H273K-Rv	ACGTGCTTCGGCATCAAGATGATC
H273L-Fw	ATGCCGCTTCACGTGGAGCC
H273L-Rv	CACGTGAAGCGGCATCAAGATG
H273Y-Fw	ATGCCGTATCACGTGGAGCC
H273Y-Rv	CACGTGATACGGCATCAAGATG
H274A-Fw	CCGCATGCCGTGGAGCCGCTGGAGC
H274A-Rv	CTCCACGGCATGCGGCATCAAGATC
H274D-Fw	CCGCATGACGTGGAGCCGCTGGAGC
H274D-Rv	CTCCACGTCATGCGGCATCAAG
H274K-Fw	CGCATAAGGTGGAGCCGCTG
H274K-Rv	CTCCACCTTATGCGGCATCAAGATC
H274A+H274A-Fw	CCGGCTGCCGTGGAGCCGCTG
H274A+H274A-Rv	CTCCACGGCAGCCGGCATCAAGATC
H315N-Fw	GTAACGAACGATCTGCAGAAAC
H315N-Rv	GATCGTTCGTAACTTCGACCAC
Y383F-Fw	CTGTTTTTCGCGGATCATCCG
Y383F-Rv	GATGATCCGCGAAAAAACAGC
D385N-Fw	TTACGCGAATCATCCGTTCATTTTCCTG
D385N-Rv	GGATGATTCGCGTAAAACAGCTTCG
HSP47 -Cstrep +XhoI	ATACTCGAGCAGCTCATCGCGCATTTTG
R222A-Fw	GATAATGCAGGTTTCATGGTTAC

(continued) **Table A.4 | Primers for Site-directed Mutagenesis**

Primer	Sequence
R222A-Rw	GAAACCTGCATTATCGACCATTT
R222E-Fw	GATAATAAGGGTTTCATGGTTAC
R222E-Rv	GAAACCCTTATTATCGACCATTT
R222K-Fw	GATAATGAGGGTTTCATGGTTAC
R222K-Rv	GAAACCCTCATTATCGACCATTT
D247A-Fw	TACTACGCAGACGAAAAAGAGAAAC
D247A-Rv	TTCGTCTGCGTAGTAGTTATACAGAC

Table A.5 | Crystallization Screen 1 & 2 (Low-Tacsimate, High and Low pH)

1% T7 16% P 10% H	1%TM 16% P 10% H	1% T8 16% P 10% H	2% T7 16% P 10% H	2%TM 16% P 10% H	2% T8 16% P 10% H	3% T7 16% P 10% H	3%TM 16% P 10% H	3% T8 16% P 10% H	4% T7 16% P 10% H	4%TM 16% P 10% H	4% T8 16% P 10% H
5% T7 16% P 10% H	5%TM 16% P 10% H	5% T8 16% P 10% H	6% T7 16% P 10% H	6%TM 16% P 10% H	6% T8 16% P 10% H	7% T7 16% P 10% H	7%TM 16% P 10% H	7% T8 16% P 10% H	8% T7 16% P 10% H	8%TM 16% P 10% H	8% T8 16% P 10% H
1% T7 18% P 10% H	1%TM 18% P 10% H	1% T8 18% P 10% H	2% T7 18% P 10% H	2%TM 18% P 10% H	2% T8 18% P 10% H	3% T7 18% P 10% H	3%TM 18% P 10% H	3% T8 18% P 10% H	4% T7 18% P 10% H	4%TM 18% P 10% H	4% T8 18% P 10% H
5% T7 18% P 10% H	5%TM 18% P 10% H	5% T8 18% P 10% H	6% T7 18% P 10% H	6%TM 18% P 10% H	6% T8 18% P 10% H	7% T7 18% P 10% H	7%TM 18% P 10% H	7% T8 18% P 10% H	8% T7 18% P 10% H	8%TM 18% P 10% H	8% T8 18% P 10% H
1% T7 20% P 10% H	1%TM 20% P 10% H	1% T8 20% P 10% H	2% T7 20% P 10% H	2%TM 20% P 10% H	2% T8 20% P 10% H	3% T7 20% P 10% H	3%TM 20% P 10% H	3% T8 20% P 10% H	4% T7 20% P 10% H	4%TM 20% P 10% H	4% T8 20% P 10% H
5% T7 20% P 10% H	5%TM 20% P 10% H	5% T8 20% P 10% H	6% T7 20% P 10% H	6%TM 20% P 10% H	6% T8 20% P 10% H	7% T7 20% P 10% H	7%TM 20% P 10% H	7% T8 20% P 10% H	8% T7 20% P 10% H	8%TM 20% P 10% H	8% T8 20% P 10% H
1% T7 22% P 10% H	1%TM 22% P 10% H	1% T8 22% P 10% H	2% T7 22% P 10% H	2%TM 22% P 10% H	2% T8 22% P 10% H	3% T7 22% P 10% H	3%TM 22% P 10% H	3% T8 22% P 10% H	4% T7 22% P 10% H	4%TM 22% P 10% H	4% T8 22% P 10% H
5% T7 22% P 10% H	5%TM 22% P 10% H	5% T8 22% P 10% H	6% T7 22% P 10% H	6%TM 22% P 10% H	6% T8 22% P 10% H	7% T7 22% P 10% H	7%TM 22% P 10% H	7% T8 22% P 10% H	8% T7 22% P 10% H	8%TM 22% P 10% H	8% T8 22% P 10% H

% = percent volume
T7 = Tacsimate (pH 7), TM = 1:1 mixture Tacsimate (pH 7) and Tacsimate (pH 8), T8 = Tacsimate (pH 8)
P = PEG-3350
H = 1M HEPES (pH 7.5 or pH 6)

Table A.6 | Crystallization Screens 3 & 4 (High-Tacsimate, High and Low pH)

9% T7 16% P 10% H	9%TM 16% P 10% H	9% T8 16% P 10% H	10% T7 16% P 10% H	10%TM 16% P 10% H	10% T8 16% P 10% H	11% T7 16% P 10% H	11%TM 16% P 10% H	11% T8 16% P 10% H	12% T7 16% P 10% H	12%TM 16% P 10% H	12% T8 16% P 10% H
13% T7 16% P 10% H	13%TM 16% P 10% H	13% T8 16% P 10% H	14% T7 16% P 10% H	14%TM 16% P 10% H	14% T8 16% P 10% H	15% T7 16% P 10% H	15%TM 16% P 10% H	15% T8 16% P 10% H	16% T7 16% P 10% H	16%TM 16% P 10% H	16% T8 16% P 10% H
9% T7 18% P 10% H	9% TM 18% P 10% H	9% T8 18% P 10% H	10% T7 18% P 10% H	10%TM 18% P 10% H	10% T8 18% P 10% H	11% T7 18% P 10% H	11%TM 18% P 10% H	11% T8 18% P 10% H	12% T7 18% P 10% H	12%TM 18% P 10% H	12% T8 18% P 10% H
13% T7 18% P 10% H	13%TM 18% P 10% H	13% T8 18% P 10% H	14% T7 18% P 10% H	14%TM 18% P 10% H	14% T8 18% P 10% H	15% T7 18% P 10% H	15%TM 18% P 10% H	15% T8 18% P 10% H	16% T7 18% P 10% H	16%TM 18% P 10% H	16% T8 18% P 10% H
9% T7 20% P 10% H	9% TM 20% P 10% H	9% T8 20% P 10% H	10% T7 20% P 10% H	10%TM 20% P 10% H	10% T8 20% P 10% H	11% T7 20% P 10% H	11%TM 20% P 10% H	11% T8 20% P 10% H	12% T7 20% P 10% H	12%TM 20% P 10% H	12% T8 20% P 10% H
13% T7 20% P 10% H	13%TM 20% P 10% H	13% T8 20% P 10% H	14% T7 20% P 10% H	14%TM 20% P 10% H	14% T8 20% P 10% H	15% T7 20% P 10% H	15%TM 20% P 10% H	15% T8 20% P 10% H	16% T7 20% P 10% H	16%TM 20% P 10% H	16% T8 20% P 10% H
9% T7 22% P 10% H	9%TM 22% P 10% H	9% T8 22% P 10% H	10% T7 22% P 10% H	10%TM 22% P 10% H	10% T8 22% P 10% H	11% T7 22% P 10% H	11%TM 22% P 10% H	11% T8 22% P 10% H	12% T7 22% P 10% H	12%TM 22% P 10% H	12% T8 22% P 10% H
13% T7 22% P 10% H	13%TM 22% P 10% H	13% T8 22% P 10% H	14% T7 22% P 10% H	14%TM 22% P 10% H	14% T8 22% P 10% H	15% T7 22% P 10% H	15%TM 22% P 10% H	15% T8 22% P 10% H	16% T7 22% P 10% H	16%TM 22% P 10% H	16% T8 22% P 10% H

% = percent volume
T7 = Tacsimate (pH 7), TM = 1:1 mixture Tacsimate (pH 7) and Tacsimate (pH 8), T8 = Tacsimate (pH 8)
P = PEG-3350
H = 1M HEPES (pH 7.5 or pH 6)

List of Abbreviations

Å	Ångstrom
ADAMTS	A Disintegrin and Metalloproteinase with Thrombospondin Motifs
APS	Ammonium persulfate
AT	Anti-trypsin
AU	Absorbance unit
BLI	Biolayer interferometry
BMP	Bone morphogenic protein
BSA	Bovine serum albumin
°C	Degrees Celsius
C-terminal	Carboxyterminal
CBB	Coomassie brilliant blue
CBP	Collagen binding protein
CD	Circular dichroism
CMP	Collagen model peptide
COL	Collagenous (Domain)
COMP	Cartilage Oligomeric Matrix Protein
COP	Coat protein complex
CpHMD	Constant pH molecular dynamics
CRM	Cytokine response modifier
CRTAP	Cartilage-associated protein
CSM	Collagen-related structural motif
cTAGE	Cutaneous T-cell lymphoma-associated antigen
CUL	Cullin ubiquitin-ligase
CypB	Cyclophilin B
dNTP	Deoxynucleotide triphosphate
ddNTP	Dideoxynucleotide triphosphate
DMSO	Dimethylsulfoxide
DSF	Differential scanning fluorimetry
DTT	1,4-Dithiothreitol
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERGIC	Endoplasmic reticulum - golgi intermediary compartment
EVH	Enabled/vasodilator-stimulated phosphoprotein homology
FACIT	Fibril associated collagens with interrupted triple-helices
FKBP	FK506-binding protein
FPLC	Fast protein liquid chromatography

FRET	Förster resonance energy transfer
GPC	Golgi-to-plasma membrane carriers
GRP	Glucose-regulated protein
GST	Glutathione-S-transferase
HBS	HEPES-buffered saline
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	Hight pressure liquid chromatography
HRP	Horseradish Peroxidase
HSP	Heat-shock protein
HTRF	Homogenous time-resolved Förster resonance energy transfer
IC_{50}	Inhibiting concentration (at 50% of maximal activity)
IMAC	Immobilied metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Dissociation constant
k_{off}	Dissociation rate
k_{on}	Association rate
LB	Lysogeny broth
LH	Lysine hydroxylase
LOX	Lysyl oxidase
mAb	Monoclonal antibody
MACIT	membrane associated collagens with interrupted triple-helices
MARCO	Macrophage scavenger receptor
MD	Molecular dynamics
MMP	Matrix metalloproteases
NC	Non-collagenous (Domain)
NHS	N-Hydroxysuccinimide
NTA	Nitrilotriacetic acid
N-Terminal	Aminoterminal
OD	Optical density
OI	Osteogenesis imperfecta
PAGE	Polyacrylamide gel electrophoresis
PCI	Protein C inhibitor
PDB	Protein database
PDI	Protein disulphide isomerase
PEG	Polyethylene glycol
PPII	Polyproline helix type II
PVDF	Polyvinylidene fluoride
P3H	Proline-3-hydroxylase
P4H	Proline-4-hydroxylase
RCL	Reactive center loop

RCSB	Research Collaboratory for Structural Bioinformatics
rER	Rough endoplasmatic reticulum
RFU	Relative fluorescence units
SA	Streptavidin
SDS	Sodium dodecylsulfate
SEC	Size-exclusion chromatography
SERPIN	Serine-protease inhibitor
SH	Src homology
SLRP	Small leucine-rich proteoglycans
SPR	Surface plasmon resonance
TAE	Tris-acetate-EDTA
TANGO	Transport and golgi organization
TBS	Tris-bufferd saline
TCE	2,2,2-Trichloroethanol
TEMED	Tetramethylethylenediamine
TLL	Tolloid-like
TREN	tris(2-aminoethyl)amine
Tris	tris(hydroxymethyl)aminomethane
TSPN	Thrombospondin-1 N-terminus-like
uPARAP	Urokinase plasminogen activator receptor-associated protein
WT	Wild-type
Xaa	Any amino acid
Yaa	Any amino acid
vWF	von Willebrand Factor

List of Figures

All figures depicted in this work were (if not explicitly stated otherwise) prepared by the author.

List of Tables

Danksagung

Ohne die Hilfe und Unterstützung vieler Personen wäre die Bewerkstelligung dieser Arbeit nicht denkbar gewesen. In vorderster Linie möchte ich mich bei Prof. Dr. Ulrich Baumann bedanken, nicht nur für die mir gegebene Möglichkeit unter seiner Aufsicht forschen und promovieren zu dürfen, sondern auch für seine geduldige, hilfsbereite und freundliche Art sowie die leidenschaftliche Herangehensweise an die Wissenschaft, welche für mich immer eine Inspiration war. Weiterhin bin ich Prof. Dr. Karsten Niefind zum Dank verpflichtet, sowohl für die Übernahme des Zweitgutachtens als auch für die Interessanten Einsichten in die Kinasenforschung zu der ich aufgrund der besonderen Nähe beider Arbeitsgruppen immer Zugang hatte. Zusätzlich möchte ich mich für Prof. Dr. Eric von Elert dafür bedanken, dass er zugestimmt hat den Prüfungsvorsitz zu übernehmen.

Besonderer Dank gebührt Dr. Jan Gebauer, der mir immer sowohl mit Rat und Tat zur Seite stand und über die ganze Zeit eine große Hilfe war.

Bei den mitleidenden Doktoranden der Arbeitsgruppen Baumann und Niefind, Alex, Anna M, Anna K, Christian, Christine, Matthias und Swantje, möchte Ich mich für die unvergessliche Zeit und der gegenseitigen Unterstützung angesichts des oft mit Widrigkeiten auftrumpfenden Doktorandenalltags bedanken. Bei Doro, Elena, Monika und Sandra möchte Ich mich für die unentbehrliche Hilfe im und um das Labor bedanken, sowie für die vielen Unterhaltungen.

Bei Andreas Oder vom FMP Berlin möchte Ich mich für die Hilfe bei der Entwicklung des HTRF Assays und für die enstpannte aber produktive Zusammenarbeit bedanken.

Zuallerletzt möchte Ich mich bei meiner Familie bedanken, für all das was Ihr mir ermöglicht habt, und für all den Rückhalt den Ihr mir gegeben habt.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie -abgesehen von unten angegebenen Teilpublikationen- noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Ulrich Baumann betreut worden.

……. …….

Ort, Datum Sinan Öcal