MONASH University

Engineering serpin stability and function

Emilia Maria Marijanovic

B. S. c (Honours)

A thesis submitted for the degree of Doctor of Philosophy at

Monash University in 2018

Department of Biochemistry and Molecular Biology

Monash Biomedicine Discovery Institute

Copyright notice

Table of Contents

Abstract

Serpinopathies are a protein misfolding disease caused by members of the serpin superfamily. These diseases lead to gain-of-toxicity at the site of synthesis and loss-of-function as a result of a decrease in circulating serpin concentration. One of the most studied serpinopathies is α1-antitrypsin (α1-AT) deficiency. α1-AT targets human neutrophil elastase (HNE), controlling the serpin: protease ratio in the lungs. Deficiency in α1-AT (levels below 35% of normal) allows for HNE to degrade the lower lungs, resulting in emphysema. The current treatment to control the activity of HNE is augmentation therapy: intravenous injections of plasma purified α1-AT, aiming to restore the deficient levels in the blood. However, this treatment is cost ineffective and inactive α1-AT has been detected. Therefore, the use of recombinant DNA technology and protein engineering to develop an α1-AT like serpin that is stable, yet functional, could provide a better treatment for α1-AT deficiency.

Previously, consensus-designed protein engineering was used to produce a synthetic serpin, termed conserpin. Conserpin exhibits extreme thermostability and aggregation resistance, while also being functional as an inhibitor towards trypsin. This extreme thermostability and aggregation resistance makes conserpin an attractive starting molecule for engineering of an α1-AT like serpin which also exhibits increased stability.

In this thesis, conserpin was used as a starting molecule to engineer a serpin that functions like α1-AT with an increase in stability. Engineering was approached in two ways: engineer conserpin to target HNE and engineer α1-AT with an increase in stability whilst remaining functional. The folding pathway of conserpin was also explored to gain insight as to why it exhibits this increased stability and aggregation resistance.

Conserpin was engineered to contain the specificity of α1-AT through mutating the sequence of the reactive centre loop (RCL). However, this process did not change the specificity of conserpin, and we found that conserpin acted as a substrate rather than an inhibitor against HNE. This indicates there are regions outside of the RCL that contribute to overall determination of a serpin's specificity. There are two possible explanations. The surface electrostatics between the serpin and protease must be compatible to allow for the formation of a stable serpin: protease complex formation, and the dynamics of the RCL must allow for the RCL to rapidly insert into β-sheet A.

Engineering α1-AT for stability involved grafting the identified stabilizing regions of conserpin onto α1-AT. This process was successful in producing 3 grafts that were more thermostable than WT α1-AT and remained functional as an inhibitor towards HNE. Combining these 3 stable grafts together produced an α1-AT variant that is more stable than the naturally occurring thermophilic serpin thermopin whilst remaining functional against HNE.

The results presented in this thesis shows the use of protein engineering to develop an α1-AT -like serpin that not only retains functionality but also gains stability. This is the start of developing a possible cost-effective replacement for the current augmentation therapy used to treat α1-AT deficiency. The success here may also be applied to other serpins.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: 

Print Name: Emilia Maria Marijanovic

Date: 12/11/2018

Publications during enrolment

1. Porebski, B. T., Keleher, S., Hollins, J. J., Nickson, A. A., **Marijanovic, E. M.**, Borg, N. A., Costa, M. G. S., Pearce, M. A., Dai, W., Zhu, L., Irving, J. A., Hoke, D. E., Kass, I., Whisstock, J. C., Bottomley, S. P., Webb, G. I., McGowan, S., and Buckle, A. M. (2016) Smoothing a rugged protein folding landscape by sequence-based redesign. *Sci. Rep. 6*, 33958.
2. **Marijanovic, E. M.,** Fodor, J., Riley, B. T., Porebski, B. T., Costa, M. G. S., Kass, I., Hoke, D. E., McGowan, S., and Buckle, A. M (2018) Reactive centre loop dynamics and serpin specificity. *Sci Rep. Submitted*

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 1 submitted publication. The core theme of the thesis is using protein engineering to engineer a serpin for function and stability. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Biochemistry and Molecular Biology under the supervision of Associate Professor Ashley M. Buckle

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapter 3 my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status** *(published, in press, accepted or returned for revision, submitted)*	**Nature and % of student contribution**
Chapter 3	Reactive centre loop dynamics and serpin specificity	Submitted	Study design, experimental work, generation of figures, writing of the manuscript, 50%

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%)**	**Co-author(s), Monash student**
James Fodor	Performed molecular dynamics simulations and analysis	10%	No

Blake T. Riley	Analysis of molecular dynamics simulations, generation of figures and wrote the manuscript	20%	Yes
Benjamin T. Porebski	Aid in experimental data and crystallography	1%	No
Mauricio G. S. Costa	Performed frustration analysis	2%	No
Itamar Kass	Assisted in molecular dynamics simulations and analysis	1%	No
David E. Hoke	Aid in experimental data	7%	No
Sheena McGowan	Performed crystallography and wrote the manuscript	2%	No
Ashley M. Buckle	Study design, performed crystallography and wrote the manuscript	7%	No

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature:  **Date:** 12/11/2018

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature:  **Date:** 12/11/2018

Acknowledgements

I would like to dedicate this thesis to two types of people I have come across during my candidature:

1. The few people who told me I could not undertake a PhD without a scholarship

and

2. To whom were unable to complete their candidature and had to withdraw

I would first like to thank my primary supervisor Associate Professor Ashley Buckle for giving me an exciting, yet challenging project, in an aspect of biochemistry that I was most interested in, protein folding. I thank you for your invaluable expertise and enthusiasm, along with your patience and financial support throughout my candidature. I would also like to thank my co-supervisor Dr David Hoke for his teachings, help and suggestions with all the laboratory work and every piece of written work I have given him to read. You and Ashley have helped shape me into the scientist I am today.

Next, I would like to thank Anne Gershenson (University of Massachusetts, USA) and Patrick Wintrode (University of Maryland, USA) for allowing me to spend time in each of your labs. This experience was invaluable, I thank you for giving me this opportunity and for everything you taught me while I was there. I hope we can continue our collaboration in the years to come.

I would like to say thank you to the former and current members of the Buckle lab, and members of other labs for all your help throughout my PhD. I would not have been able to complete this thesis without our discussions. I must also thank all my non-science friends for their support and being someone to talk to when the stress started to become unbearable.

Lastly, I would like to thank my parents and family. You may not understand the science in this thesis, but your support and assistance helped me through my university course and PhD candidature.

Abbreviations

ACT	α1-antichymotrypsin
α1-AT	α1-antitrypsin
arb	Arbitrary units
bis-ANS	4, 4'-Dianilino-1, 1'-Binaphthyl-5, 5'-Disulfonic Acid, Diphosphate salt
β-ME	2-mercaptoethanol
CD	Circular dichroism
°C	Celsius
D	Denatured
DTT	1,4-Dithiothreitol
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
GndHCl	Guanidinium hydrochloride
ΔG	Gibbs free energy
ΔΔG	Difference in Gibbs free energy
HNE	Human neutrophil elastase
IPTG	Isopropyl β-D-thiogalactoside
K	Kelvin
MD	Molecular Dynamics
min	Minute
N	Native
OD_{600}	Optical density at 600nm
PAI-1	Plasminogen activator inhibitor-1
PCA	Principle component analysis
PDB ID:	Protein Data Bank identifier
PEG-3350/8000	Polyethylene glycol 3350/800
RCL	Reactive centre loop
SI	Stoichiometry of inhibition
s	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Serpin	Serine protease inhibitor
smFRET	Single-molecule Förster resonance energy transfer
Tm	Midpoint of thermal denaturation
TEMED	Tetramethylethylenediamine

Chapter 1: Introduction

1.1. Protein folding and stability

Proteins are essential biomolecules for living systems. They are involved in a large range of functions including enzymes, ion channels, structural elements and molecular machines. To function, however, proteins must fold into their functional conformation from an unfolded, polypeptide chain. Under suitable conditions, the unfolded protein will fold into an ordered structure, governed by the proteins amino acid sequence[1,2]. How proteins fold using the information stored in the polypeptide sequence is one of the most sought-after questions in protein chemistry.

1.1.1. Thermodynamics of Protein Folding

The understanding that the amino acid sequence codes the folding pathway and structure of proteins was concluded through the unfolding and refolding studies of ribonuclease by Anfinsen and colleagues[3–5]. These studies also let Anfinsen to postulate the "thermodynamic hypothesis". This hypothesis states that, under physiological conditions, 'the Gibbs free energy of the whole system is lowest', that is, the protein and surrounding solvent is at the lowest energy possible, a global free energy minimum[6,7]. This suggests that the protein will fold into the global minima from any unfolded conformation independent of any initial conditions[7,8]. The Gibbs free energy is defined by 3 terms, entropy S, enthalpy H and temperature T. Therefore, the Gibbs free energy of protein folding is the difference of energy between the unfolded and native states[9]:

$$\Delta G_{(\text{Denatured-Native})} = \Delta H_{(\text{Denatured-Native})} - T\Delta S_{(\text{Denatured-Native})}$$

During protein folding, the entropy decreases as order begins to form in the system, while the enthalpy increases with favourable interactions in the protein's native state. Therefore, the Gibbs free energy, $\Delta G_{(\text{Denatured-Native}),}$ between the denatured and native states of most proteins is approximately -5 to -15 kcal mol^{-1}, largely due to the loss of entropy while folding[10–12].

1.1.2. Kinetics of Protein Folding

Soon after Anfinsen postulated the 'thermodynamic hypothesis' of protein folding, Levinthal recognized that the time it would take for an unfolded protein to search through a vast number of possible conformations to yield a lowest-energy folded structure, is not possible on a

biological timeframe. He postulated that protein folding is under kinetic control, through specific pathways on a biological timeframe ('Leventhal's Paradox')[13,14]. Therefore, protein folding must be under both thermodynamic and kinetic control: folding into a stable conformation in an appropriate timeframe[15]. This can be described in terms of 'energy landscapes' (Figure 1).

The rate at which proteins fold is influenced by its size. Small proteins (100-amino acids or less/ 20kDa) are said to exhibit two-state folding kinetics, that is, an all-or-none transition from unfolded to folded (or reverse). In this mechanism, no folding intermediates have been detected experimentally[9,11,16]. Proteins that are greater than 100 amino acids fold via one or more intermediate state (multi-state kinetics). These intermediates are partially folded entities that may slow down the rate of folding[17,18]. Common to both folding kinetic rates are transition states. The transition state is the rate-limiting state, and the point in the energy landscape where energy is at a maximum. It is not a single structure, but an ensemble of conformations, with constant formation and breakage of many weak bonds[8,11,16].

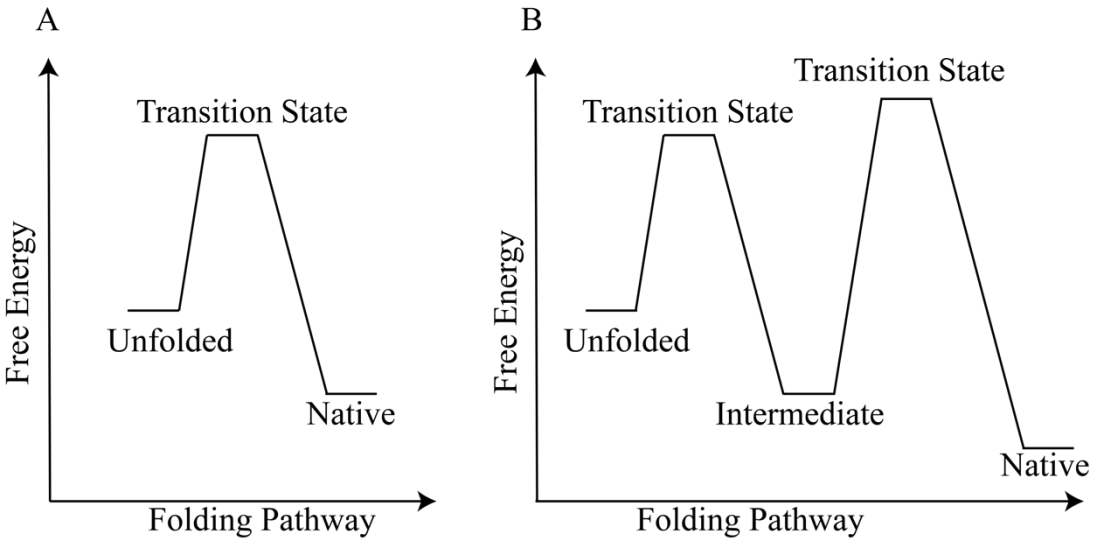


Figure 1. **Free energy diagrams of two-state and multi-state kinetic folders. A**. Two-state kinetic folders fold through one transition state with no populated intermediate. These proteins fold through an all-or-nothing mechanism from unfolded to native conformation. **B**. Multi-state kinetic folders contain at least one transition state and a folding intermediate along the folding pathway.

1.1.3. Folding Funnels: Combining Thermodynamic and Kinetic folding

The energy landscape of a folding protein can be viewed as a folding funnel, first introduced by Leopold and colleagues[19]. The energy landscape theory combines both the thermodynamic and kinetics of folding, imagined as a funnel. At the top of the funnel is the unfolded protein, populating many conformations with high entropy and enthalpy[8]. Here, there is a conflict between the maximum entropy keeping the protein unfolded, and the minimal enthalpy attempting to drag the protein down the folding funnel[11]. As folding proceeds down the funnel, favourable, native interactions accumulate, increasing the stability and reducing the number of conformational states possible until it reaches the bottom of the funnel, the native state at the lowest free energy[2,20–22]. Individual molecule of the same protein may follow different routes to the same native state, with some routes being more populated than others[22]. Folding pathway can contain valleys and hills, corresponding to intermediates or energy-barriers, respectively. While climbing up a hill from a valley (from a folding intermediate towards the native state), the protein does not necessarily unfold, rather it is the breaking of some contacts which will allow the protein to progress to fold to the native conformation[8].

The folding funnel for two-state and multi-state folding kinetics differ. Two-state folders are said to have a smooth funnel, lacking deep valleys (intermediates) and hills (energy barriers)[9,23]. In contrast, multi-state folders have a rugged funnel, with deep valleys hills[24].

1.1.4. Protein stability

For a protein to function, it must fold and remain stable in its native conformation. Just as there are two aspects to protein folding, there are two aspects for protein stability, thermodynamic stability and kinetic stability. Studies on protein stability often refer to the Gibbs free energy between the native and denatured states ($\Delta G_{(Denatured\text{-}Native)}$), and the difference in energy between the wild type and mutant proteins ($\Delta G_{WT\text{-}Mutant)}$). This calculation is the thermodynamic stability, which is the equilibrium between the unfolded, partially folded and the native conformation of the protein[25]. The kinetic stability relates to the high energy barriers that separates the different conformations along the folding pathway (i.e. separation of the unfolded, partially folded and the native conformation), while also providing information on the folding pathway of the protein[25,26].

Various different forces contribute to the stability of a protein, with one of the biggest contribution being the hydrophobic interactions (greater than 50% of the forces)[27–30]. This is due to more than 80% of a proteins non-polar side chains buried into the tightly packed interior of the protein[30]. The large contribution in stability is from the removal of the non-polar side chains from water and the enhanced forces that arise from the tightly packed interior[31]. Each methyl group on a hydrophobic residue contributes to approximately -1.3kcal mol^{-1} [29,32,33]. The importance of the hydrophobic interactions in the core of the protein can be assessed through mutations that decrease the size of the side chain, producing a small cavity[32,33]. Hydrogen bonds also contribute a large contribution to the stability of a protein, by approximately -1.3 kcal mol^{-1} per bond[31,34,35]. Understanding the effect hydrogen bonds have on the stability of a protein can be observed through mutagenesis studies, where an amino acid that has the capability of forming a hydrogen bond (e.g. asparagine) is mutated to an amino acid that does not form hydrogen bonds (e.g. alanine)[34]. Other forces that contribute to the stability of proteins include salt bridges[36–38], van der Waals interactions[31,34,39] and disulfide bonds[40–42]. The addition of every interaction within the native state of a protein would be large, yet, the energy differences between the unfolded and native states of a protein is -5 to -15 kcal/mol, upon folding[11,12,43].

1.1.5. Folding into a metastable conformation

Within the proteome, there are some instances where protein folding is under kinetic control, rather than the combination of kinetic and thermodynamic control. Some proteins do not fold into the lowest-free energy minima conformation, instead adopt a metastable conformation that is necessary for biological function[7,44]. Proteins that fold into a metastable conformation are prevented from reaching the lowest energy minima by either the height of the energy barrier separating the metastable and lowest energy states, or by an external factor (e.g. a pro-region of a protease) which affects the barrier height[7]. Examples of proteins that fold into a metastable conformation include α-lytic protease[45], influenza hemagglutinnin[46], bacterial luciferase[47] and members of the serine protease inhibitor (serpin) superfamily[44]. Here, the serpin superfamily will be discussed.

1.2. Serine Protease Inhibitor (Serpin) superfamily

Serine protease inhibitors (serpins) are the largest and most widely distributed family of protease inhibitors[48] . More than 3000 serpins have been identified within all phyla[49,50]. The identification of this superfamily arose through structural alignment of ovalbumin, antithrombin-III and a1-antitrypsin (α1-AT), determining a shared sequence homology of 30-50%[51]. Despite its name as serine protease inhibitors, not all members inhibit serine proteases. This family also consists of cross-class inhibitors that target cysteine proteases (e.g. SCCA1 and PI9)[52]. Some members are also non-inhibitory, instead perform other roles such as hormone transport (thyroxine-binding globulin)[53], tumour suppression (maspin)[54] and as a molecular chaperone (HSP47)[55].

1.2.1. Folding into a metastable conformation

Serpins are one example where proteins violate the thermodynamic hypothesis of folding into the lowest energy minima conformation. Instead, serpins fold into a functional, metastable native ('stressed', S) conformation that is central to its inhibitory function[56,57]. There is some indication of how serpins fold, based upon biophysical experiments, yet the complete folding pathway has not been determined.

1.2.2. Serpin structure

The conformation serpins fold into provides information on the relationship between structure and function. The serpin's metastable, native conformation consist of 3 β -sheets (A-C) surrounded by 9 α-helices (A-I) and a protruding mobile reactive centre loop (RCL) (Figure 2). The largest element of the secondary structure is a dominant β -sheet (β -sheet A) that is a characteristic of serpins. This sheet is surrounded by a lower α-helical domain at the N-terminus and an upper β-barrel at the C-terminus (β-sheets B and C)[58–60]. Two strands from β-sheet B and Helix-B form a hydrophobic core, perpendicular to β-sheet A[61,62]. The protruding reactive centre loop (RCL) is central for the serpin's inhibitory mechanism, acting as bait for the target protease to bind. The RCL contains a scissile bond, P1-P1', where the target protease binds and cleaves[63]. In the serpins native conformation, this loop extends from strand 5 of β-sheet A (s5A) and strand 1 of β-sheet C (s1C)[64,65]. With its inhibitory mechanism, this loop is cleaved and inserts into β-sheet A, becoming an extra strand within the β-sheet (discussed in Section 1.2.4). Many features contribute to the 'stressed', metastable conformation of a serpin,

including over-packing side-chains in the hydrophobic core[66], unfavourable hydrophobic interactions[67], buried polar groups[68] and surface pockets[69]. This native fold is shared throughout the serpin family, despite overall poor sequence homology between its members[70–72].

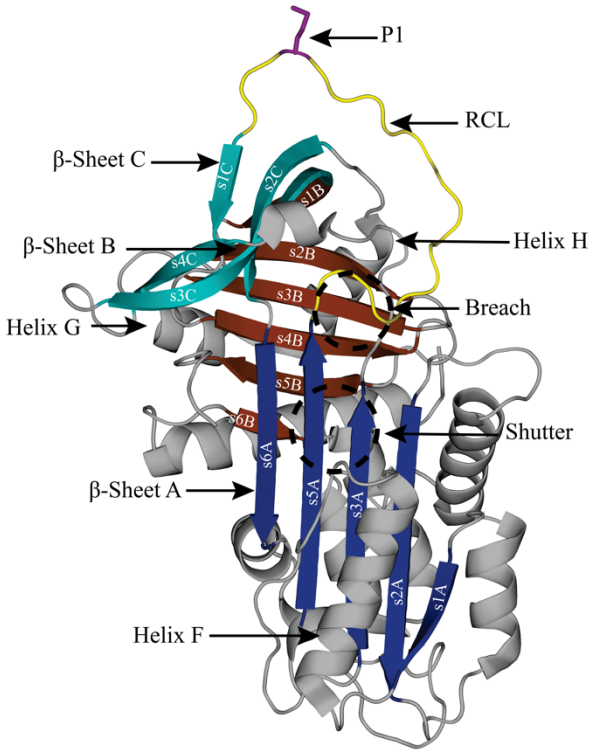


Figure. 2. The X-ray crystallography structure of archetypal serpin α1-antitrypsin (α1-AT) in the native, metastable conformation (PDB: 1QLP[73]). The serpin structure consists of 3 β-sheets (A, *blu*e, B, *brown* and C, *teal, with each strand labelled*) and 9 α-helices (grey). β-sheet A contains highly conserved regions, Breach and Shutter (*black circles*), which are essential in the serpin's inhibitory mechanism. The reactive centre loop (*yellow*) protrudes from the body of the serpin, containing the P1 residue (*purple*).

1.2.3. Folding into a metastable conformation

Many studies, using α1-AT as the archetypal serpin, have attempted to solve how serpins fold into this native conformation. They have revealed that it is a 3-state folding mechanism with at least one spherical, aggregation-prone folding intermediate[74–77]. This folding intermediate

contains approximately 80% of the secondary structural elements present in the native state. These secondary structures include a partially folded β-sheet A, B and C[56,75,78], folded Helix-B[79,80] and disrupted Helix-F[64,79,81]. Within this intermediate, Helix G and H, along with β-sheet B, are hypothesized to serve as a nucleus for folding[76,82,83].

The folding intermediate may serve to prevent any misfolding and folding into a more stable conformation through non-native interactions formed (interactions that are not present in the metastable conformation), restricting the conformational space available to the folding serpin[84]. Helix-F, which lays in front of β-sheet A, may play a role in folding. Biophysical studies show the top of Helix-F inserting between strand 3 and 5 of β-sheet A. This prevents premature insertion of the RCL[81,85].

After the folding intermediate has formed, the C-terminus, including regions s4A-s5A, s1C, s4B-s5B proceed to fold[86]. Which of these segments fold first is not clear. One study states that strand s5A must associate with the already-folded regions before any other strands[86]. However, a hydrogen-deuterium exchange study indicates that there is a 'race' between which region folds first: strands s1C and s4B must fold faster than s5A to prevent the RCL from inserting into β-sheet A[87]. What is known is the RCL is anchored above the body of the serpin last by hydrogen bonds[80,86].

1.2.4. Serpins mechanism of inhibition

The inhibition mechanism of a serpin is unique and viewed as a 'molecular mousetrap', which involves a large conformational change during the process (Figure 3). Inhibition begins with the protease targeting and docking onto the exposed RCL, forming a reversible stoichiometric 1:1 Michaelis- Menton complex[63,88]. Next, the catalytic serine (serine-195) of the protease attacks the P1 residue, progressing from a reversible to an irreversible covalent acyl-enzyme intermediate, disrupting the P1-P1' peptide bond[63,89,90]. Here, the N-terminal region of the RCL is released in a spring mechanism with the protease covalently attached.

The next stage of inhibition is a branched pathway, where the serpin either continues as an inhibitor, or becomes a substrate for the protease. The pathway taken is dependent on the rate the RCL inserts into β-sheet A. If loop insertion is rapid and occurs faster than de-acylation (completion of the proteases catalytic mechanism), the serpin continues as an inhibitor. In this

pathway, the protease, trapped in the acyl-enzyme intermediate, is translocated to the opposite pole of the serpin. In the translocation, the catalytic active site of the protease is distorted through a pulling force exerted on the catalytic triad. The catalytic serine is pulled away from its catalytic partner, histidine (histidine-57), preventing any hydrolysis of the ester bond with P1 residue on the RCL. This leaves a stable covalent serpin: protease complex[59,63,91]. Histidine-57 is also displaced, allowing for the destruction of the oxyanion hole, preventing de-acylation from occurring. Furthermore, the translocation of the covalently attached protease causes a clash between the protease and the body of the serpin, further disrupting the catalytic site of the protease. Since the serpin: protease complex produced after inhibition is covalently linked, it is kinetically stable[73,92].

The other pathway, the substrate pathway, occurs when the rate of RCL insertion into β-sheet A is slower than the de-acylation step of the proteases catalytic mechanism. Here, the RCL is cleaved by the protease and begins to insert into β-sheet A. However, de-acylation occurs, resulting in the release of active protease while the RCL still inserts. The final product is cleaved, inactive serpin and active protease. This pathway occurs with mutations in the RCL that slow the rate of insertion[91,93]. The insertion of the RCL as an extra strand on β-sheet A, via cleavage or insertion producing the latent conformation, releases approximately 32 kcal mol^{-1} of energy[48,94,95], leaving the serpin in an inactive, but the lowest energy thermodynamically stable conformation (‘relaxed’, R).

Stability in the central strands of β-sheet A, s3A and s5A, is important during inhibition as they contain two critical regions for RCL insertion: the breach and the shutter region (Figure 2). These regions facilitate the opening of the sheet and acceptance of the RCL[96]. The breach region is located at the top of the sheet on strands s3A, s5A, s2B and s4B, where the RCL first inserts[71,97], and is involved in in stabilizing acyl-enzyme complex[98]. The shutter region is present in the centre of strands s3A and s5A, as well as portions of strand 5B[99]. The shutter is regarded as the controller of the rest of the molecule, with amino acid packing being extremely important[71]. Therefore, cleavage of the RCL causes the opening of β-sheet A, allowing for a conformational change to occur[99].

Helix-F is hypothesized to be important in the conformational change needed for serpin inhibition. Helix-F acts as a barrier for rapid RCL insertion[81]. The opening of β-sheet A induces a conformational change in Helix-F, where there is partial unfolding and displacement of the

helix. This allows for RCL insertion[99,100]. Mutations in Helix-F affect the conformational change post protease docking. An interaction between isoleucine-157 in Helix-F and β-sheet A must be maintained throughout the inhibitory mechanism, otherwise there is an increase in the substrate pathway[64,81]. Therefore, Helix-F, along with β-sheet A, must have plasticity to allow for the serpin to undergo the large conformational change associated with its inhibitory mechanism.

1.2.5. Adopting an alternative, inactive conformation

The RCL can insert into β-sheet A without cleavage, producing an inactive conformation defined as the latent conformation (Figure 3). For the RCL to insert without cleavage, one strand from β-sheet C (strand 1C) is displaced to allow the RCL to insert. Interestingly, plasminogen activator inhibitor-1 (PAI-1) is regulated by undergoing latency at physiological temperatures[101]. A serpin can be forced to undergo latency under different conditions, including elevated temperatures, low chemical denaturant concentrations, low pH or mutations in specific regions of the serpin (e.g. the B/C barrel)[44,102–104]. Other serpins that undergo latency include α1-AT[102] and a1-antichymotrypsin[105].

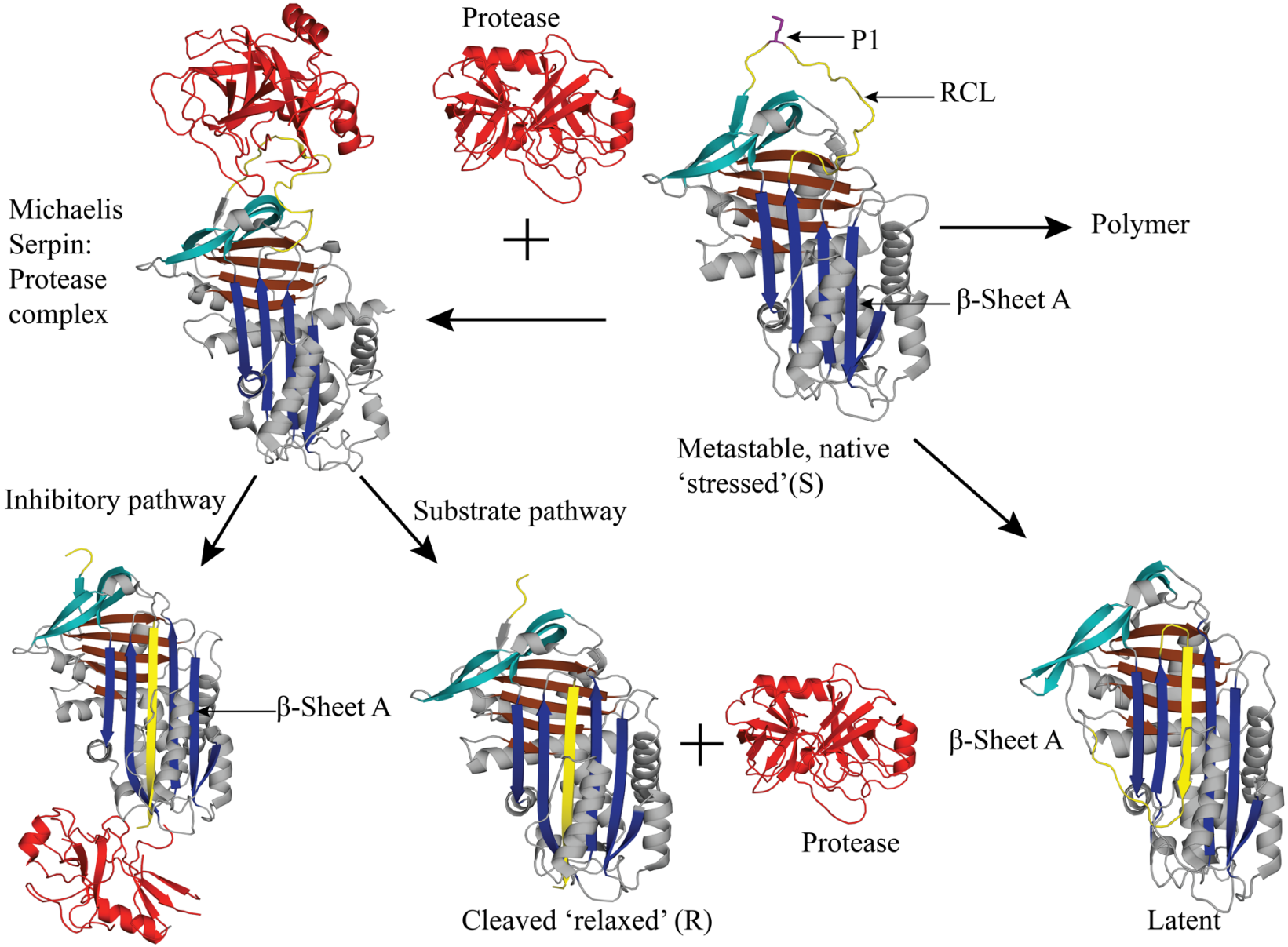

Protease
P1
RCL
Michaelis
Serpin:
Protease
complex
Polymer
β-Sheet A
Metastable, native
'stressed'(S)
Inhibitory pathway
Substrate pathway
β-Sheet A
β-Sheet A
Protease
Cleaved 'relaxed' (R)
Latent

Figure 3. The inhibitory mechanism of serpins. The metastable, native ('stressed', S) conformation contains an exposed RCL (PDB 1QLP[73]). The protease (e.g. trypsin) (PDB: 1S0Q, *unpublished*) binds to the P1 residue on the RCL, producing a Michaelis serpin: protease complex (PDB: 1K9O[106]). Here, the serpin can take the inhibitory pathway or the substrate pathway. The inhibitory pathway involves RCL insertion with the protease covalently attached (PDB: 1EZX[63]). The substrate pathway occurs when the RCL inserts slower than the protease's catalytic mechanism, resulting in RCL insertion and a cleaved serpin ('relaxed', R, PDB: 3NDD[60]), with the release of active protease. From the native conformation, the serpin can insert the RCL without cleavage, producing the latent conformation (PDB: 1IZ2[107]).

1.2.6. Serpin misfolding and polymerization

The ability of serpins to fold into a metastable, native conformation, and use this metastability to insert its RCL into its own β-sheet, renders the molecule susceptible to misfolding caused by mutations[84]. This misfolding often leads to polymerization. Serpin polymerization forms the bases of many diseases, termed serpinopathies. Serpinopathies are characterized by the formation of polymers, leading to gain-of-toxicity, and a decrease in circulating active serpin, causing a loss-of-function disease[84,108,109]. A deficiency in α1-AT, a1-antichymotrypsin, neuroserpin and antithrombin-III results in emphysema[110], chronic obstructive pulmonary disease[111], dementia[112], and thrombosis[113], respectively.

It is difficult to monitor how serpin misfolding leads to polymerization within the human body, therefore *in vitro* studies have used various conditions to promote polymerization. These include altering pH[104,114], low concentration of denaturant [54,81], increased temperature[115,116], cleavage at non-standard positions within the RCL[117,118] and destabilizing mutations[64,119] to promote misfolding and polymerization. Any mutation in the β-sheets B and C can result in the transition from metastable to misfolded conformations. Mutations in this region fail to anchor the RCL into the correct conformation during folding, allowing off-target, misfolding conformations to occur[120].

In vitro misfolding studies have determined that serpin polymerization occurs through a two-step kinetic mechanism[121]. The first step involves the native, active serpin undergoing a rapid conformational change to form a polymerogenic species. The change in conformation is independent of the concentration of native serpin present. Any factor that decreases the stability of the serpin (e.g. pH, heat or mutations) will increase the concentration of the polymerogenic species[104,114,122]. The second step is slower and concentration dependent. Here, the build-up of the

polymerogenic species induces the formation of polymers. The polymerogenic species can also undergo a side pathway, where it folds into the intramolecular latent, inactive conformation.

Serpin polymerization can also occur through a folding intermediate. This folding intermediate, populated under mild denaturant, contains some structural similarities to the polymerogenic intermediate. The folding intermediate and polymerogenic species have disruption in β-sheet A, as detected by intrinsic and extrinsic fluorescence [121,123–125].

1.3. α1-antitrypsin deficiency

One of the most studied serpinopathies is α1-antitrypsin (α1-AT) deficiency. α1-AT is an abundant glycoprotein present in the plasma at a concentration of 20-53uM[126]. Synthesized in the liver, α1-AT travels to the lungs where it primarily functions as an inhibitor of neutrophil elastase, balancing the serpin: protease ratio in the lower respiratory tract. α1-AT deficiency is a hereditary disorder characterized by serum levels below 35% of the normal levels. In the absence of or decreased levels of α1-AT, the equilibrium of inhibitor to protease is shifted, so that neutrophil elastase is free to degrade elastin, increasing the risk of developing respiratory complications such as emphysema[127]. More than 100 allelic variants have been identified, which have been classified into 3 categories: deficient allele variants (decreased level), null alleles (not expressed) and dysfunctional allele variants[126]. Many of the variants identified produce deficient protein, including the most common variant, Z variant (Glu342Lys)[128].

1.3.1. Z variant: the most common cause of α1-AT deficiency

The Z variant is the most common and well- studied α1-AT variant. This variant is found in 4% of the northern European population and results in a decrease to approximately 15% of the normal circulating levels[129–131]. This deficiency does not arise from a reduction in protein synthesis, but rather the increased propensity of the protein to misfold and aggregate[129,132]. Only 15% of the protein expressed is secreted from hepatocytes (liver cells), while the remaining is either degraded in the endoplasmic reticulum or self-associates to form intracellular polymers (which can lead to liver disease)[133]. Of the 15% protein successfully secreted, a small fraction polymerizes in the lungs, further decreasing the amount of circulating functional serpin[134], while the remaining serpin exhibits a decrease in inhibitory function (in comparison to 'wild type' α1-AT)[135].

The Z variant contains an amino acid substitution, from a negatively charged glutamic acid (Glu) to a positively charged lysine (Lys) at a highly conserved residue 342[136]. Located at the top of s5A in the breach region (P17 of the RCL), Glu-342 forms a salt bridge with Lys-270 and hydrogen bonds with Thr-203[97]. Mutating Glu-342 to a lysine disrupts this salt bridge, resulting in a slight conformational change, while the breach region becomes open and highly dynamic compared to 'wild type' (WT) α1-AT, as observed in the x-ray crystallography structure, hydrogen-deuterium exchange (HDX) and molecular dynamics (MD) simulations[78,137,138].

How this mutation leads to polymers is not completely understood. It is suggested that the Z mutation slows the folding of α1-AT, with the aggregation-prone folding intermediate being more than in WT α1-AT. This indicates that the kinetic stability of α1-AT is compromised in the presence of the Z mutation[65,124,139]. During folding, strand 5A associates with the folded, N-terminal portion of the serpin, prior to the folding of the C-terminal portion (as discussed in section 1.2.3.)[86]. Strand 5A must be annealed to the N-terminal portion before the C-terminal portion of the serpin can begin to fold and associate. The Z mutation causes a loss of key interactions in β-sheet A, resulting in the partial association of s5A, delaying the folding process of α1-AT[78]. The delayed folding allows premature association of the C-terminal portion of the serpin. As a result, the formation of C-terminal domain-swap polymers may occur, as observed in an X-ray crystal structure of an α1-AT domain-swapped trimer[116]. While the folding of Z α1-AT has slowed, the unfolding of Z α1-AT from native to folding intermediate occurs faster than WT α1-AT, further emphasizing the kinetic instability of Z α1-AT.

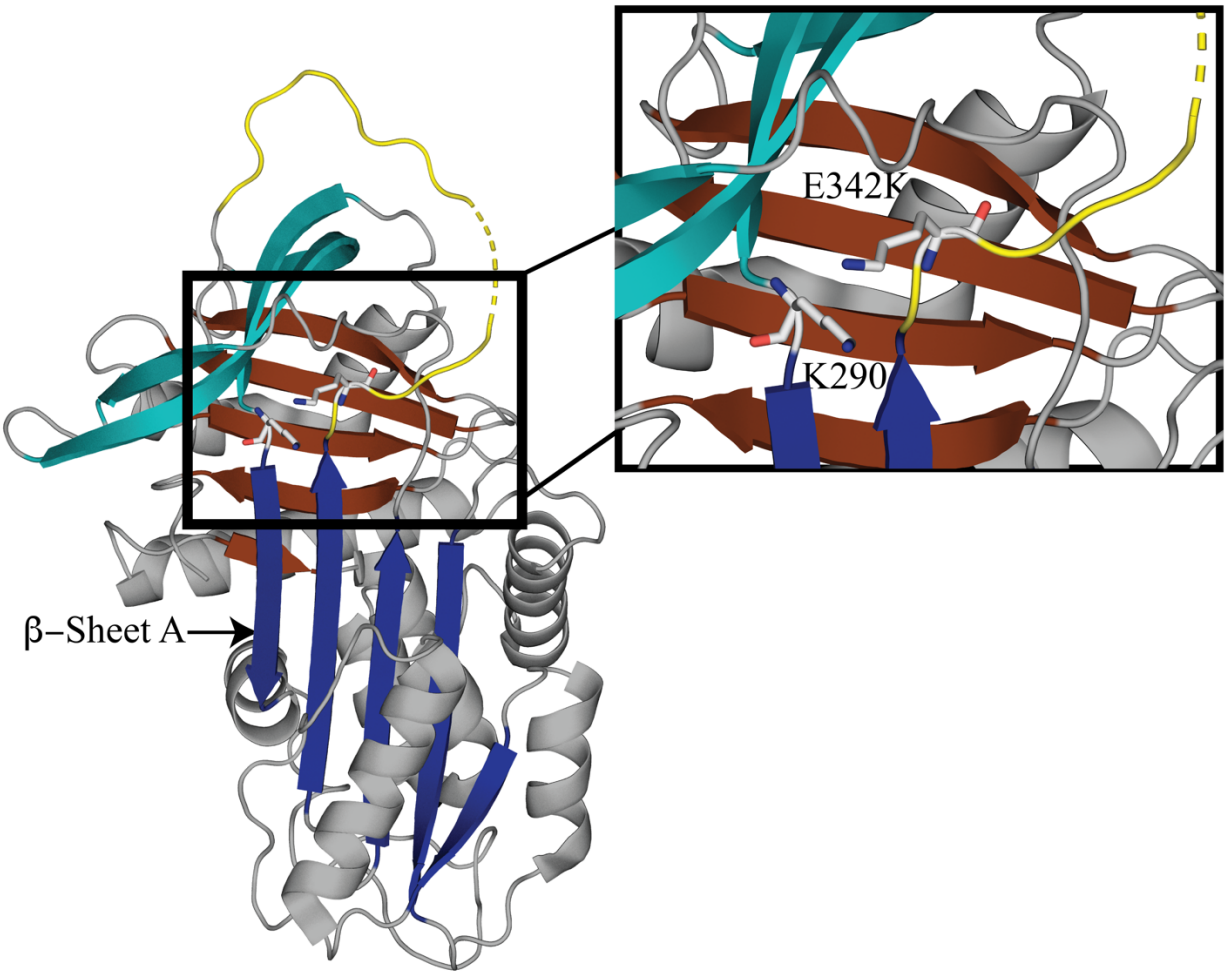


Figure 4. Zoom of the x-ray crystal structure of Z α1-AT (PDB: 5IO1[78]). This variant contains a glutamic acid to lysine mutation at position 342 (white sticks), resulting in a loss of a conserved salt bridge between lysine-290 (K290) and Glutamic acid-342 (E342). Z α1-AT misfolds along the folding pathway leading to the formation of polymers.

1.3.2. Treatment for α1-AT deficiency

The only treatment for emphysema as a consequence for α1-AT deficiency is weekly intravenous injections of plasma-purified α1-AT (augmentation therapy). This aims to increase the circulating α1-AT concentrations to above 11uM, slowing the progression of emphysema with restoring the serpin: protease balance[126,133]. However, significant complications arise, with the most important being the detection of latent α1-AT in the purified sample[140]. Furthermore, augmentation therapy is expensive (USD $60,000-$150,000 per annum) and large quantities are required (60mg/kg of body weight)[126,141,142]. Therefore, alternative treatments are required.

Currently, 4 main approaches have been investigated to reduce the polymerization of Z α1-AT during folding in the liver and increase the levels secreted in the blood: chemical chaperones[143–145], peptide analogues representing the RCL[146,147], small compounds[142] and intrabody treatment[148]. Each of the approaches all prevent polymerization *in vitro*, however, each approach prevented α1-AT from performing its inhibitory mechanism by blocking the insertion of the RCL into β-sheet A. Therefore, another approach to produce a treatment for α1-AT deficiency is necessary. One approach is to use protein engineering to produce a recombinant, α1-AT -like serpin that is stable and aggregation resistant, while remaining functional against HNE. Recombinant protein therapeutics have several benefits, including more cost-effective, production of large quantities of a given therapeutic, reduction in the possible exposure of human or animal diseases, and the ability to modify a protein therapeutic to improve its function[149]. Protein engineering will allow the modification of an α1-AT -like serpin with increased stability and aggregation-resistance.

1.4. Consensus-designed serpin, *Conserpin*

Previously, we employed a protein engineering technique, termed "consensus engineering" to design a synthetic serpin that is active as an inhibitor against serine proteases[150]. Consensus engineering is a sequence-based technique that involves studying the conserved residues in homologous proteins. Through using an alignment of the homologous proteins sequence, the most conserved residue at a given residue number is selected for a consensus sequence. The conserved residues are hypothesized to produce a larger contribution to the proteins structure, function or stability than the non-conserved residues. [151,152]. This consensus approach has been demonstrated to increase the stabilities of a number of proteins[153,154].

This consensus-designed serpin, termed *conserpin,* was produced from an alignment of 212 sequences from the serpin superfamily, sharing a sequence identity of 62% with α1-AT. Despite being a synthetic serpin, conserpin is a functional inhibitor with a consensus-designed RCL, inhibiting trypsin with a stoichiometry of inhibition (SI) of 1.8. Conserpin is extremely thermostable, with a midpoint of thermal unfolding (*Tm*) well above 100°C, in comparison to α1-AT at 61.8°C. Thermal unfolding was only achieved in the presence of 2M guanidinium hydrochloride at 72.5°C. This demonstrated the power of consensus design in increasing protein stability[150].

One of the technical difficulties in studying the folding of many serpins is their tendency to misfold and form aggregate from the unfolded state. Conserpin, however, reversibly folds into the native

conformation from a chemically denatured unfolding with no aggregation observed. This is hypothesized to occur due to conserpin sampling a reduced folding intermediate (as detected with fluorescent dye), populated over a small timeframe. In comparison, α1-AT's folding intermediate is populated over a large timeframe.

Why conserpin is so highly stable may be explained by many favourable interactions throughout the structure, despite having fewer hydrogen bonds and salt bridges in comparison to α1-AT and thermostable serpins thermopin[155] and tengpin[156]. These favourable interactions include favourable polar and non-polar interactions, and improved packing in the hydrophobic core. Many of these interactions are in regions that are suggested to be important in the folding of serpins, including Helix-D and F, β-sheet A and the hydrophobic core, and packing around the B/C-barrel (β-sheets B and C). Therefore, the high thermal stability and reduced propensity of aggregation makes conserpin an excellent model to study the stability and function of serpins.

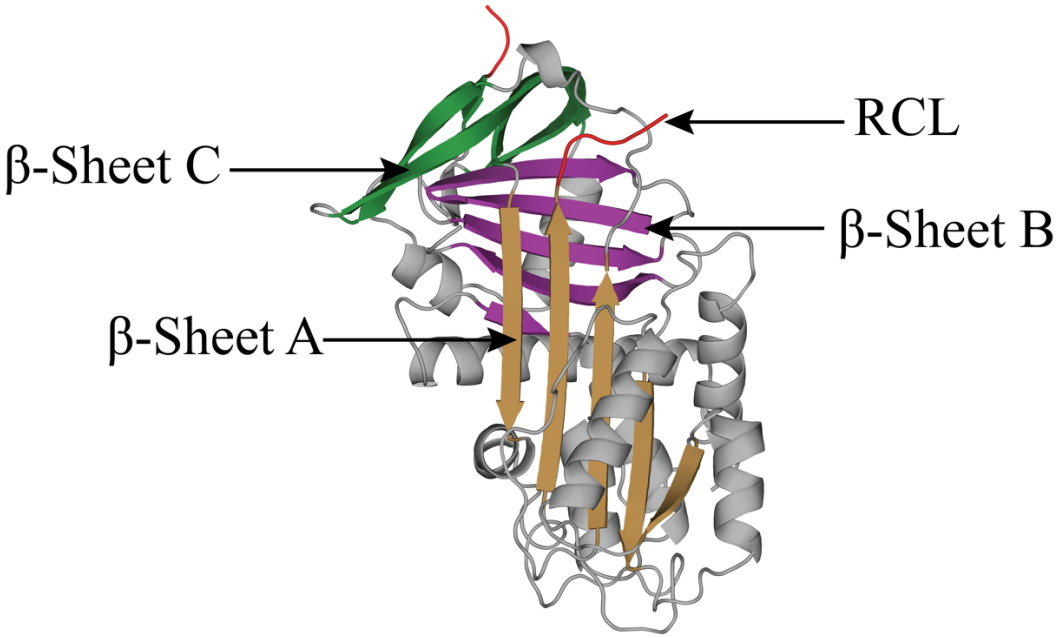


Figure 5. The X-ray crystal structure of Conserpin (PDB: 5CDX[150]). Conserpin folds into a native conformation consisting of 3 β-sheets (A, *sand*, B, *purple* and C, *green*), 9 helices (*grey*) and a protruding RCL (*red*). The RCL was not resolved in the electron density due to its flexibility.

1.5. Aims of this thesis

The use of consensus engineering produced a synthetic serpin, conserpin, containing ideal stability and biophysical properties (e.g. thermostability and aggregation resistance), while also being completely functional with a consensus-designed RCL. These properties make conserpin an ideal base model for further engineering of serpins with enhanced properties and function. The research presented in this thesis focuses on using conserpin to further engineer serpins for function and stability. The aims of this thesis are:

1. Engineering function without compromising stability:

   The consensus-design of conserpin did not provide the serpin with specificity against a given protease. Conserpin is engineered to target human neutrophil elastase (HNE), α1-AT's target, through engineering the reactive centre loop.

2. Engineering stability without compromising function:

   Conserpin's favourable biophysical properties are hypothesized to be a result of regions of favourable interactions throughout the serpin molecule. These regions are grafted onto α1-AT to increase the stability and biophysical properties of the serpin, without compromising α1-AT's function against HNE.

3. Studying the folding of a synthetic serpin using single-molecule Förster Resonance Energy Transfer (smFRET):

   The reversible, rapid folding nature of conserpin differs to mesophilic serpins, but has similarities to thermostable serpins. The folding mechanism of conserpin will be investigated using single-molecule Förster Resonance Energy Transfer (smFRET), providing a detailed outline on its folding pathway. This will provide an insight into how thermostable serpins are capable of folding under extreme temperatures.

Chapter 2:
Materials and Methods

The material and methods described below are the general methods used in the results chapters (Chapters 3, 4 and 5) of this thesis. The chemical and reagents used, along with the supplier, are listed below:

Chemical	**Supplier**
Rubidium chloride (RbCl) Manganese chloride ($MgCl_2$) Calcium chloride ($CaCl_2$) Chloramphenicol (Chlor) Ammonium persulfate (APS) Sodium dihydrogen phosphate (NaH_2PO_4) Potassium dihydrogen phosphate (KH_2PO_4) Dimethyl sulfoxide (DMSO) Trypsin N-Methoxysuccinyl-Ala-Ala-Pro-Val p-nitroanilide Na-Benzyol-L-arginine 4-nitroanilide hydrochloride 4, 4'-Dianilino-1, 1'-Binaphthyl-5, 5'-Disulfonic Acid, Diphosphate salt (Bis-ANS)	Sigma-Aldrich
Potassium acetate MOPS Tris base Sodium dodecyl sulfate (SDS) Imidazole Guanidine hydrochloride (GndHCl) ethylenediaminetetraacetic acid (EDTA**)** Tween 20 2-mercaptoethanol (β-ME)	Amresco
Acetic acid Hydrochloric acid (HCl) Potassium hydroxide (KOH) Sodium hydroxide (NaOH) Yeast extract Peptone (tryptone) Sodium chloride (NaCl) Agar	Merck

Potassium chloride (KCl)	
Glycerol Ampicillin (Amp) 37.5:1 (40%) acrylamide solution 1,4-Dithiothreitol (DTT) Isopropyl β-D-thiogalactoside (IPTG) Glycine Di-sodium hydrogen phosphate (Na_2HPO_4)	Astral Scientific
Tetramethylethylenediamine (TEMED) Commassie brilliant blue G-250 Precision Plus Protein dual colour standards Quick start Bradford 1x Dye reagent Mini-PROTEAN tetra System	Bio-Rad
ISOLATE II plasmid Mini Kit	Bioline
Kanamycin (Kan)	Gibeo by Life Science
Human neutrophil elastase (HNE)	Abcam
ECL start Western Blot detection reagent	Amersham
Anti-His tag HRP-labelled mouse monoclonal IgG antibody	R&D Systems
Methanol	Sharlau
Fuji medical X-ray film	Fujifilm

2.1. Recombinant protein expression

2.1.1. Bacterial cell lines

Cell line	Genotype	Reference
DH5α	F- endA1 hsdR17 (r_K^-, m_K^+) supE44 thi-1 λ^- recA1 gyrA96 relA1 deoR Δ(lacZYA-argF) U169 ϕ80lacZΔM15	[157]
BL21 (DE3) pLysS	F-, ompT, $hsdS_B$ (r_B-, m_B-), dcm, gal, λ(DE3), pLysS, Cm^r	[158,159]
SG13009 (Qiagen)	F^-, Nal^s, Str^s, Rif^s, Thi^-, Lac^-, Ara^+, Gal^+, Mtl^-, $RecA^+$ Uvr^+, Lon^+, Km^r	[160]
Rosetta Blue DE3	endA1, hsdR17 (rK12-mK12+), supE44, thi-1, recA1, gyrA96, relA1, lacF'[proA+B+acIqZΔM15::Tn109tetR)] (DE_3)pLysSRARE (CmR)	Novagen

2.1.2. Preparation of competent *E. coli* cells

Competent cell buffers:

- *Luria-Bertani (LB):* Yeast extract (5g/L), peptone (tryptone, 10/L), NaCl (10g/L)
- *Luria-Bertani (LB) agar:* Yeast extract (5g/L), peptone (tryptone, 10/L), NaCl (10g/L), 1.5% agar
- *TFB1:* 100mM RbCl, 50mM $MgCl_2$, 30mM potassium acetate, 10mM $CaCl_2$, 15% glycerol, pH 5.8 (acetic acid for pH). Cold
- *TFB2:* 10mM MOPS, 10mM RbCl, 75mM $CaCl_2$, 15% glycerol, pH 6.8 (KOH for pH). Cold

Protocol:

Competent cells were prepared from a glycerol stock. A small sample of cells was streaked onto LB agar, containing the appropriate antibiotics for the cell line, and incubated overnight at 37°C. A single colony was picked and inoculated 10ml LB media (with appropriate antibiotics) and grew overnight at 37°C. 1ml of the overnight culture inoculated 100ml of pre-warmth LB media (with antibiotics). Cells grew until the optical density at 600nm (OD_{600}) reached 0.5, followed by cooling of the culture on ice for 5min before collecting cells by centrifugation (4,000xg for 5 minutes at 4°C). The supernatant was discarded, cells re-suspended in cold TFB1 and remained suspended

on ice for 90 minutes. The cells were again collected by centrifugation (4,000xg for 5 minutes at 4°C), supernatant discarded, and cells re-suspended in cold TFB2. The competent cells were aliquoted, snap frozen and stored at -80°C until used.

2.3.2. Transformation of DNA into E. coli cells

Vectors:

The vectors used in this thesis to express recombinant proteins of interest were the pLIC-His and pQE-31 (Qiagen). The vectors containing the DNA for expression of proteins conserpin, conserpin-AAT_{RCL} and WT α1AT (pLIC-His and pQE-31) were obtained from Shani Keleher and Sheena McGowan (Monash University) and the vector containing Δ31-Tengpin (Chapter 4) was obtained from Gordon Lloyd (Monash University). The proteins expressed and studied in Chapter 4 (α1-AT grafts) were synthesized and cloned into the pQE-31 vector by Genscript (USA).

DNA vector		**Transformed into:**	**Reference**
pLIC-His	T7 promotor, ampicillin resistance, N-terminal 6xHis-tag	DH5α (storage) BL21 (DE3) pLysS (expression)	[135,161]
pQE-31	T5 promotor, ampicillin resistance, N-terminal 6xHis-tag	DH5α (storage) SG13009 (expression)	[162,163]
pET-3a	T7 promotor, ampicillin resistance	Rosetta Blue DE3	Novagen

Buffers:

- *LB media*: Yeast extract (5g/L), peptone (tryptone, 10/L), NaCl (10g/L)
- *LB Agar*: Yeast extract (5g/L), peptone (tryptone, 10/L), NaCl (10g/L), 1.5% agar
- *Antibiotics*: Ampicillin (100mg/ml), Chloramphenicol (34mg/ml) and Kanamycin (50mg/ml).

Protocol:

20µl of competent cells was thawed, mixed with 1µl of DNA and incubated on ice for 30min. The competent cells were heat shocked at 42°C for 45 seconds, followed by incubating on ice for 2 minutes. 100µl of LB media was added to the cells, and cells recovered at 37°C for 1 hour. The cells were plated onto LB agar (with appropriate antibiotics) and incubated overnight at 37°C.

2.1.4. DNA sequencing

Vector	**Primer**
pLIC-His	
T7 promotor	5’ TAA TAC GAC TCA CTA TAG GG 3’
T7 terminator	5’ GCT AGT TAT TGA TCA GCG G 3’
pQE-31	
Promotor	5’ GGA GAA ATT AAC TAT GAG AGG 3’
Terminator	5’ GTT CTF AGG TCA ATT ACT GG 3’

The protein-containing vectors were amplified for sequencing by a transformation into DH5α (as by protocol above), and a single colony selected for an overnight culture in LB media. The vector was purified from the bacterial cells using Bioline ISOLATE II plasmid Mini Kit, following manufacture’s protocol. DNA sequencing was performed by Micromon (Monash University), using the primers listed above.

2.1.5. Site-directed mutagenesis

Site-directed mutagenesis[164] was performed to introduce cysteine mutations onto conserpin for single-molecule Forster Resonance Energy Transfer (smFRET) study (Chapter 5).

Site-directed mutagenesis primers:

Mutation	**Direction**	**Primer**	**Annealing temperature (°C)**
D25C	Forward	5’-GTGAGCTTGCAAAGTCTAGTCCTTGCAAAAA CATCTTTTTCTCAC -3’	64
	Reverse	3’-CACTCGAACGTTTCAGATCAGGAACGTTTTTG TAGAAAAAGSAGTG-5’	
D205C/ C209S	Forward	5’-AATACTATCACGATTGTGAGTTAAGTSAGTAA GGTTTTGGAACTTCCCTAC-3’	63
	Reverse	3’-ATATGGTGCTAACACTCAATTCATCATTCCAA	

		AACCTTGAAGGGATG-5’	
S285C	Forward	5’-GGTATTACAGATCTTTTCTGCCCAGGGGCTGACTTG-3’	65
	Reverse	3’-CCATAATGTCTAGAAAAGGCAGGTCCCCGACTAAAC-5’	

Polymerise chain reaction (PCR) reaction mixture:

Chemical	**Amount (µl)**
5x Phusion HF Buffer	10
10mM dNTPs	1
10uM Forward Primer	2.5ul
10mM Reverse Primer	2.5ul
Plasmid DNA	1
ddH20	32.5ul
Phusion DNA polymerase	0.5
Total Volume	50

Polymerase chain reaction (PCR) cycling reaction

Set up		**Temperature (°C)**	**Time (minutes)**	**Cycles**
Initial denaturation		98	2.5	1
Amplification	*Denaturation*	98	0.5	20
	Annealing	X	1	
	Extension	72	6	
Final extension		72	10	1
Hold		6	-	

X: The annealing temperature is dependent on the primers.

Dpn1 digest:

Chemical	**Volume (µl)**
Dpn1	1
Dnp1 digestion buffer	6

After completion of the cycling reaction, the parental plasmid was digested with Dnp1 for 1hour at 37°C.

The PCR product was transformed into DH5α competent cells as stated above, plated onto 2xYT agar plates (with 50µg/µl ampicillin) and incubated overnight at 37°C. A single colony from each agar plate was selected and added into 10ml 2xYT media (with 50µg/µl ampicillin) and incubated overnight at 37°C. The mutated plasmid was purified using the Bioline ISOLATE II plasmid Mini Kit, following manufacture's protocol. The mutation was confirmed by DNA sequencing by Microgen USA.

2.2. Protein expression

Buffers:

- *2xYT media*: Yeast extract (10g/L), peptone (tryptone, 16g/L), NaCl (5g/L)
- *2xYT agar*: Yeast extract (10g/L), peptone (tryptone, 16g/L), NaCl (5g/L), agar (1.5%)
- *Antibiotics*: Ampicillin (100mg/ml), Chloramphenicol (34mg/ml), Kanamycin (50mg/ml) and tetracycline (12.5mg/ml)
- Isopropyl B-D-thiogalactoside (IPTG): 1M

Protocol:

2.2.1. Conserpin, α1-AT and variants

A single colony from a transformation (as protocol above), inoculated an overnight culture of 10ml 2xYT media (with appropriate antibiotics) and incubated at 37°C overnight with constant shaking. This overnight culture inoculated 1 litre of 2xYT, then incubated at 37°C with shaking until the cells reached an OD_{600} of 0.6. Protein expression was induced with IPTG to a final concentration of 1mM, and cells left to express protein for a minimum of 3 hours (3 hours for conserpin and conserpin-AAT_{RCL}, and 4 hours for WT α1-AT and grafts) at 37°C. After the designated protein expression timeframe, the cells were harvested by centrifugation (4,000xg for 15 minutes at 4°C) and cell pallets stored at -20°C.

2.2.2. Δ31-Tengpin

A single colony from a transformation (as protocol above) was used to inoculate an overnight culture of 10ml 2xYT media (with appropriate antibiotics) and incubated at 37°C overnight with constant shaking. 20mls of the overnight culture was used to inoculate 1L 2xYT media (with appropriate antibiotics) and cells incubated at 37°C with shaking. Once the OD_{600} reached 0.6, the cells were cooled to 16 °C and inoculated with IPTG at a final concentration of 1mM. Protein expression occurred overnight. Cells were then harvested by centrifugation (6,000xg for 20 minutes at 4°C), frozen then stored at -20°C.

2.3. Protein Purification

2.3.2. α1-antitrypsin:

WT α1-antitrypsin (α1-AT) was expressed both soluble and insoluble in E. coli, depending on the vector the protein DNA was inserted into. WT α1AT in the pQE-31 vector expressed in the soluble fraction/ bacterial cytoplasm[162], while α1-AT expressed in the pLIC-His vector was in inclusion bodies. All α1-AT grafts (Chapter 4) expressed into inclusion bodies, despite the DNA sequence cloned into pQE-31 vector.

2.3.1.1. Soluble expressed protein:

Cell lysis buffer: 25mM NaH_2PO_4, 500mM NaCl, 1mM β-ME, 10mM Imidazole, pH 8.0

Nickel-NTA (Ni-NTA) Affinity chromatography (native conditions):

- Loose Nickel-NTA resin (Qiagen)
- *Ni-NTA wash buffer:* 25mM NaH_2PO_4, 150mM NaCl, 1mM β-ME, 10mM Imidazole, pH 8.0
- *Ni-NTA elution buffer:* 25mM NaH_2PO_4, 150mM NaCl, 1mM β-ME, 250mM Imidazole, pH 8.0

Ion-exchange chromatography:

- AKTA FPLC (GE Healthcare)
- HiTrap Q FF anion chromatography column (GE Healthcare)
- *Dilution buffer*: 50mM Tris-HCl pH 8.0

- *Anion buffer A:* 50mM Tris-HCl, 50mM NaCl, 1mM EDTA, pH 8.0
- *Anion buffer B:* 50mM Tris-HCl, 1M NaCl, 1mM EDTA, pH 8.0

Protocol:

The cell pallets containing soluble α1-AT were re-suspended in cell lysis buffer and incubated on ice for 20minutes. Cells were disrupted by sonication on ice (6x 25 seconds on/ 35 seconds off) followed by centrifugation at 48,000xg for 20 minutes (4°C) to pallet cellular debris. The soluble fraction was loaded onto pre-equilibrated loose Ni-NTA resin and incubated at 4°C for 1 hour to allow batch binding to occur. Any protein that did not bind was eluted and collected, while loose binding protein was eluted with Ni-NTA wash buffer. Soluble α1-AT was eluted with Ni-NTA elution buffer into 1ml fractions. The fractions that contained protein (as detected by 1:10 protein: Bradford reagent assay), were pooled and diluted with dilution buffer at a 1:1 ratio. The diluted sample was loaded onto a pre-equilibrated HiTrap Q FF anion exchange chromatography column. Ion-exchange chromatography separates the different conformations that serpins can adopt (i.e. separates active native from inactive latent and aggregate conformations). Any protein that did not bind was collected, and native α1-AT was eluted with an increasing sodium chloride (NaCl) concentration. Fractions that contained protein was run on an SDS-PAGE to ensure purity and tested against trypsin for inhibitory activity. Pure, active protein was either snap frozen and stored at -80°C for long term storage or at 4°C for use.

2.3.1.2. Insoluble expressed protein:

Cell lysis buffer: 50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole pH 8.0

Unfolding inclusion body buffer: 8M Urea, 50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole, 10mM β-ME, pH 8.0

Nickel-NTA (Ni-NTA) Affinity chromatography (denaturing conditions):

- Loose Nickel-NTA resin (Qiagen)
- *Ni-NTA wash buffer:* 8M Urea, 50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole, 10mM β-ME, pH 8.0
- *Ni-NTA elution buffer:* 8M Urea, 50mM NaH_2PO_4, 300mM NaCl, 250mM Imidazole, 10mM β-ME, pH 8.0

Refolding buffer: 50mM Tris-HCl, 50mM NaCl, 5mM DTT, pH 8.0

Ion-exchange chromatography:

- AKTA FPLC (GE Healthcare)
- HiTrap Q FF anion chromatography column (GE Healthcare)

- *Anion buffer A:* 50mM Tris-HCl, 50mM NaCl, 1mM EDTA pH 8.0
- *Anion buffer B:* 50mM Tris-HCl, 1M NaCl, 1mM EDTA pH 8.0

Protocol:

The cell pallets that contained insoluble α1-AT (α1-AT expressed into inclusion bodies) were resuspended in cell lysis buffer and incubated on ice for 20 minutes. Cells were disrupted by sonication on ice (30 seconds on/off) and inclusion bodies harvested by centrifugation (48,000xg for 20 minutes at 4°C). As the protein has a 6xHis-tag, an inclusion body preparation was not necessary. Instead, Nickel-NTA (Ni-NTA) affinity chromatography was performed under denaturing conditions to partly purify α1-AT from the inclusion body. The protein within the inclusion body was resuspended in unfolding inclusion body buffer by constant stirring with a magnetic stirrer for 2 hours at room temperature (approximately 21°C). Any protein that was not resuspended was pelleted by centrifugation (35,000xg for 20 minutes at 4°C). The soluble was filtered through 0.8μm filter, loaded onto pre-equilibrated loose Ni-NTA resin and left to batch bind for 1 hour at 4°C with rocking. Any protein that did not bind was collected, while loosely bound protein was eluted with Ni-NTA wash. α1-AT was batch eluted with Ni-NTA elution buffer. To ensure protein was present in the eluted sample, a 1:10 protein: Bradford reagent assay was performed.

The eluted α1-AT was refolded by 1:200 dilution into refold buffer. It is essential that the refold buffer contains DTT to reduce the cysteine residues and prevent disulphide-driven aggregation to occur. The refold was left to occur overnight at 4°C with constant stirring to minimize concentration-dependent aggregation. Any aggregation that occurred was removed through filtering the refold buffer through 0.22μm filter before loading the refold onto a pre-equilibrated HiTrap Q FF anion chromatography column. Any protein that did not bind was collected, and the difference conformations. Protein detected in the fractions were run on an SDS-PAGE to check for purity and tested against trypsin for inhibitory activity. Pure, active α1AT was pooled, and either snap frozen and stored at -80°C for long term storage or at 4°C for use.

2.3.2. Conserpin and conserpin variants

Cell lysis buffer: 50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole pH 8.0

Nickel-NTA (Ni-NTA) Affinity chromatography (native conditions):

- Loose Nickel-NTA resin (Qiagen)
- *Ni-NTA wash buffer:* 50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole pH 8.0

- *Ni-NTA elution buffer:* 50mM NaH_2PO_4, 300mM NaCl, 250mM Imidazole pH 8.0

For purification of cysteine-mutations (Chapter 5), 10mM β-ME was added to all Nickel-NTA affinity chromatography buffers

Size-exclusion chromatography (SEC):

- AKTA FPLC (GE Healthcare)
- Superdex 200 16/60 prep grade column (GE Healthcare)
- *SEC Buffer:* 50mM Tris-HCl, 150mM NaCl pH 8.0

For purification of cysteine-mutations (Chapter 5), 5mM DTT was added to the SEC buffer

Protocol:

Conserpin and variants were express in the soluble fraction/ cell cytoplasm of the bacterial cells. The cells were resuspended in cell lysis buffer and incubated on ice for 20 minutes. Cells were disrupted by sonication on ice (6x 25 seconds on/ 35 seconds off) followed by centrifugation (48,000xg for 20 minutes at 4°C) to pallet cellular debris. The soluble fraction was loaded onto pre-equilibrated loose Ni-NTA resin and incubated at 4°C for 1 hour to allow batch binding to occur. Any protein that did not bind was eluted and collected, while loose binding protein was eluted with Ni-NTA wash buffer. Conserpin and variants were eluted off the loose Ni-NTA resin with Ni-NTA elution buffer into 10ml fractions. A Bradford assay was performed to determine which fractions contained eluted protein. The fractions that contained protein were pooled and concentrated using a 10,000Da (10kDa) concentrator (Millipore) and centrifugation (2,500xg for 15-minute intervals at 4°C). Concentrated protein was further purified by size-exclusion chromatography (SEC) of a Superdex 200 16/60 prep grade SEC column in SEC buffer. Fractions that contained protein was subjected to SDS-PAGE to check for purity. Purified protein was pooled and wither snap frozen and stored at -80°C or 4°C for use.

2.3.3. Δ31-Tengpin

Cell lysis buffer: 50mM Tris-HCl, 150mM NaCl 1mM EDTA, 1% Triton-100 pH 8.0

Dilution buffer: 50mM Tris-HCl, 1mM EDTA pH 8.0

Ion-exchange chromatography:

- HiTrap Q Fast Flow (FF) column (5ml, GE Healthcare)
- *Buffer A:* 50mM Tris-HCl, 10mM NaCl, 1mM EDTA pH 8.0
- *Buffer B:* 50mM Tris-HCl, 1M NaCl, 1mM EDTA pH 8.0

Hydrophobic exchange chromatography:

- *Phenyl stock buffer*: 4M $(NH_4)SO_4$

- HiTrap Phenyl Fast Flow (FF) column (5ml, GE Healthcare)
- *Buffer A:* 50mM Tris-HCl, 1.7M $(NH_4)SO_4$, 1mM EDTA pH 8.0
- *Buffer B:* 50mM Tris-HCl, 0M $(NH_4)SO_4$, 1mM EDTA pH 8.0

Size-exclusion chromatography (SEC):

- AKTA FPLC (GE Healthcare)
- Superdex 200 16/60 prep grade column (GE Healthcare)
- *SEC Buffer:* 50mM Tris-HCl, 50mM NaCl, 1mM EDTA pH 8.0

Protocol:

Tengpin was purified from the soluble fraction of bacterial cells. The cells were resuspended in lysis buffer and cells incubated on ice for 20 minutes. The cells were disrupted by sonication on ice (30 seconds on, 1 minute off), then cellular debris pelleted by centrifugation (35,000xg for 1 hour at 4°C). The soluble fraction was diluted in a 1:4 ratio with dilution buffer, then loaded onto a pre-equilibrated HiTrap Q FF column. Protein was eluted with an increase in salt concentration with a linear gradient. Fractions were assessed to contain tengpin by SDS-PAGE (as per protocol below). Fractions that contained tengpin were pooled and dialysed against dilution buffer (1:50-100 dilution). 3 dialysis buffer changes occurred before the 4th dialysis was left overnight (4°C with constant stirring). The phenyl stock buffer was added to the dialysed protein sample until the final concentration was 1.7M $(NH_4)SO_4$, followed by centrifugation (35,000xg for 15 minutes at 4°C) to remove any precipitated proteins. Tengpin remains soluble in 1.7M $(NH_4)SO_4$, and the tengpin sample was loaded onto a pre-equilibrated HiTrap Phenyl FF column. Protein was eluted with a linear decrease of $(NH_4)SO_4$, and fractions detected to contain protein were observed for tenpin by SDS-PAGE. Fractions that contained tengpin were pooled, concentrated using a 10,000Da (10kDa) concentrator (Millipore) and centrifugation (2,500xg for 15-minute intervals at 4°C), and loaded onto a pre-equilibrated SEC column with the appropriate buffer. Tengpin purity was observed by SDS-PAGE, and fractions positive for tengpin were pooled, snap frozen in liquid nitrogen, and stored at -80°C.

2.4. Protein buffer exchange

- AKTA FPLC (GE Healthcare)
- HiTrap Desalting column (GE Healthcare)

Protocol:

To buffer exchange all proteins into a specific buffer, a HiTrap desalting column was used. The column was pre-equilibrated with the buffer of interest, and protein was loaded onto the HiTrap desalting column through the AKTA FPLC at a 1ml volume. The fractions containing protein were pooled and stored at 4°C for use.

2.5. Determination of protein concentration

The concentration of each protein sample was performed using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific), measuring the absorbance at A_{280}. The extinction coefficient of each protein was included when measuring the A_{280}, calculated by ExPASy ProtParam online tool (https://web.expasy.org/protparam/). The extinction coefficient (at 1%) for each protein is below:

Protein	**Extinction coefficient (1%)**
Serpin	
Conserpin	7.74
Conserpin-AAT$_{RCL}$	7.77
WT α1-AT	4.8
Breach	4.79
Helix-F	5.37
F51	4.8
Citrate	5.13
B/C Barrel	6.03
T59S	4.8
β-sheet C Stapling	4.79
Helix-H	4.8
3stable	5.36
Protease	
HNE	9.85
Trypsin	14.4

2.6. Sodium Dodecyl Sulfate- Polyacrylamide Gel Electrophoresis (SDS-PAGE)

An SDS-PAGE gel was run at the end of each protein purification procedure to check the purity of the protein. The gel was subjected to western blot only during trial expressions to check if the protein of interest was in the soluble (cell cytoplasm) or insoluble (inclusion bodies) fraction of the bacterial cells.

Reagents:

- 37.5% (40%) acrylamide solution
- *4x Resolving buffer:* 0.5M Tris-HCL, 0.4%(w/v) SDS, pH 6.8
- *4x Stacking buffer:* 1.5M Tris-HCL, 0.4% (w/v) SDS, pH 8.8
- ddH_2O
- 20% (w/v) APS
- TEMED
- *1X SDS Running buffer:* 25mM Tris base, 192mM glycine, 0.1% (w/v) SDS
- *6X Laemmli SDS sample loading dye:* 375mM Tris-HCl, 9% SDS, 50% glycerol, 60mM DTT and 0.03% bromophenol blue
- Precision Plus Protein dual colour standards (Bio-Rad)
- *Staining solution:* 50% (v/v) ddH_2O, 40% (v/v) Methanol, 10% (v/v) acetic acid, 0.1% (w/v) Commassie brilliant Blue G-250
- *Destaining solution:* 50% (v/v) ddH_2O, 40%(v/v) Methanol, 10%(v/v) acetic acid
- Bio-Rad Mini-PROTEAN Tetra System

SDS-PAGE gels were constructed with 10% resolving gel and 4% stacking gel. Each gel was constructed as below:

Reagent	*10% Resolving gel*	*4% Stacking gel*
37.5% (40%) acrylamide solution (ml)	2.5	0.325
4x Resolving buffer (ml)	1.87	-
4x Stacking buffer (ml)	-	0.625
ddH_2O (ml)	2.62	1.57

20% APS (µl)	25	12.5
TEMED (µl)	10	5

Protocol:

Each sample subjected to electrophoresis were diluted with 6x reducing sample dye and boiled at 90°C for 5 minutes before loading into the stacking wells of the SDS-PAGE gel. The gel was run at 250V for 30 minutes, or when the dye front reached the bottom of the gel, whichever was first. The gel was either stained with commassie blue protein stain or subjected to western blot. If the gel was to be stained with commassie blue, commassie blue protein stain was added to the gel and left to incubate for 30 minutes while rocking (at room temperature). The stain was then removed and destain added until the background of the gel has destained.

2.7. Western blot

Reagents:

- *TBS-T:* 1x Tris-buffered saline (150mM NaCl, 25mM Tris-HCl, pH 7.4), 0.1% (v/v) Tween 20
- *Blocking buffer:* TBS-T, 5% (w/v) powdered skim milk
- *Transfer buffer:* 25mM Tris base, 190mM glycine, 15% methanol
- Anti-His HRP-labelled mouse monoclonal IgG
- ECL western blot detection reagents (Amersham)
- Fuji medical X-ray film (Fujifilm)

Protocol:

After an SDS-PAGE gel was performed, the proteins were transferred onto a pre-methanol soaked PVDF membrane in transfer buffer using a Bio-Rad transfer system, with a 100V current for 1 hour. The membrane containing the transferred proteins was blocked with blocking buffer for 1.5 hours, followed by the removal of the blocking buffer and addition of anti-His HRP-labelled antibody in TBST (1: 10,000) for another 1 hour. Following antibody binding, the PVDF membrane was washed 4 times for 3 minutes with TBST. The ECL western Blot reagents were mixed and added to the membrane and the membrane exposed to X-ray film for 30 seconds, 1 minute and 30 minutes in a light proof cassette. The film was developed by X-ray developer**.**

2.8. Determination of serpin inhibitory activity

- *Assay buffer:* 50mM Tris-HCL, 150mM NaCl, 0.2%(v/v) PEG 8000, pH 7.4
- *Proteases:* Trypsin (1mM HCl), HNE (50mM sodium acetate, 150mM NaCl, pH 5.5)
- *Chromogenic substrate for trypsin*: Na-Benzyol-L-arginine 4-nitroanilide hydrochloride (in DMSO)
- *Chromogenic substrate for HNE:* N-Methoxysuccinyl-Ala-Ala-Pro-Val p-nitroanilide (in DMSO)
- BMG FLUOstar Optima plate reader (405nm)

2.8.1. Stoichiometry of Inhibition (SI)

Protocol:

The stoichiometry of inhibition (SI) of each serpin against protease was performed similarly as described[165]. Different concentrations of serpin were incubated with a constant concentration of protease (producing a range of 0-2:1 ratios) at 37°C for 30 minutes in assay buffer to allow a serpin: protease complex to form. The residual protease activity was measured for 1 hour with the addition of a target substrate (200μM), using an Optima plate reader set at 405nm. The change in absorbance from addition of substrate to the final reading after 1 hour was plotted as the percentage of active protease against the serpin: protease ratio. The data was normalized, with full protease activity (0:1 serpin: protease) at 100%, and fit with a linear regression, where the interaction with the X-axis determining the SI.

2.8.2. Refolding and determining the stoichiometry of inhibition

- *Unfolding buffer:* 6M GndHCl, 50mM Tris-HCl, 150mM NaCl pH 8.0
- *Refolding buffer:* 50mM Tris-HCl, 150mM NaCl pH 8.0

Protocol:

Serpin was unfolded in unfolding buffer, consisting of 6M GndHCl, for 2 hours before refolding by dilution for another 2 hours. Any aggregate formed was pelleted by centrifugation (desktop centrifuge, 16,000xg for 5 minutes at 4°C) and refolded sample dialysed against the refolding buffer to remove remaining GndHCl. The SI was performed as stated above.

2.8.3. Serpin: protease complex SDS-PAGE gels

- 10% SDS-PAGE gel
- *6X Laemmli SDS loading dye*

Protocol:
Different concentrations of serpin to a constant concentration of protease (producing a 1:1 and 2:1 ratio) was incubated at 37°C for 30 minutes to allow for a serpin: protease complex to form. This reaction was stopped after 30 minutes with the addition of 6X reducing sample buffer and quenching the samples on ice. The samples were subjected to SDS-PAGE, using a 10% acrylamide gel, as stated above. The gel was stained with commassie blue stain and destained to observe any SDS-stable serpin: protease complex formation.

2.9. Biophysical analysis of serpins

2.9.1. Circular Dichroism (CD)

- *Protein buffer:* 1x Phosphate-buffered saline (1x PBS) (136mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, pH 7.4)

All circular dichrosim techniques were performed on a Jasco J-815 circular dichroism spectrometer at a protein concentration of 0.2mg/ml in 1x PBS, pH 7.4, using a path-length of 0.1cm in a quartz cell.

2.9.1.1. Spectral Scan

The observation of protein secondary structure was performed with Far-UV spectra (195-250nm), scanning 100nm/min. The concertation of protein was 0.2mg/ml and the temperature held constant at 20°C.

2.9.1.2. Thermal denaturation

Thermal denaturation was performed with an increase in temperature from 25°C -100°C (unless otherwise stated), at a rate of 1°C/ min, recording the change in signal at 222nm. The protein remained at a concentration of 0.2mg/ml. The sample was the cooled from 100°C to 25°C to observe if any thermal refolding occurred. The midpoint of transition (T_m) was obtained by fitting the data with a Boltzmann sigmoidal curve for both the melting and reverse melting/cooling denaturation experiments.

For proteins that did not undergo an unfolding transition from heating to 100°C, 2M GndHCl was added to aid in the unfolding. The sample with 2M GndHCl was again subjected to an increase in temperature from 25°C -100°C to determine the midpoint of transition.

2.9.1.3. Determining the thermodynamic properties using thermal melts

The thermodynamic properties of the serpins were determined using the unfolding transition slope of the thermal melts. First, the ellipticity signal was normalized, producing a Y-axis of 'Fraction of unfolded' against temperature. The 'fraction of unfolded' when then converted into *Keq* using the following equation:

$$K_{eq} = \frac{f_u}{1 - f_u}$$

Where f_u is the fraction of unfolded protein. K_{eq} was then converted into logK_{eq}. The temperature was converted into kelvin (°C + 273.15), then transformed into 1/ temperature (K). Next, the logK_{eq} was plotted against the corresponding 1/Temperature (K) to produce a van't Hoff analysis[166–168]. The resulting plot was fitted to a linear equation, with the Y-intercept producing the change in entropy, ΔS, using the gas constant, R (8.314 J/K/mol) and the following equation:

$$\frac{\Delta S}{R}$$

The change in enthalpy, ΔH, was calculated using the slope of the line and the gas constant, R, and the equation:

$$\frac{-\Delta H}{R}$$

The Gibbs free energy determined by thermal denaturation was calculated at room temperature (25°C) using the Gibbs free energy equation:

$$\Delta G = \Delta H - T\Delta S$$

The change in Gibbs free energy (ΔΔG) between the WT and grafts was calculated using the equation:

$$\Delta\Delta G = \Delta G_{WT} - \Delta G_{Grafts}$$

2.9.2. Analysis of protein refold by gel filtration

- AKTA FPLC (GE Healthcare)
- Superdex 200 increase 10/300 GL (GE Healthcare)
- *Protein/ refolding buffer:* 1x Tris-buffered saline (1xTBS) (150mM NaCl, 25mM Tris-HCL, pH 7.4)
- *Unfolding buffer:* 8M GndHCl (in 1x TBS pH 7.4)

Protocol:

Analysis of refold was performed by comparing the absorbance peak (at 280nm) of native and refolded protein using a Superdex 200 10/300 column at a final concentration of 2μM. For refolding, the protein was unfolded in 5M GndHCl for 1.5 hours, then refolded by diluting the sample 10 times with refolding buffer until the final concentration of protein was 2μM. The refolded sample was centrifuged to remove any aggregate and a total volume of 500μl loaded onto the column. The yield was analysed by comparing the peak absorbance between native and refolded protein, normalizing the refolded peak against the top absorbance for the native protein (where top absorbance of native protein is 100%).

2.9.3. Bis-ANS unfolding

- Cary Eclipse Fluorescence spectrophotomer (Agilent Technologies)
- *Native buffer:* 1x Phosphate-buffered saline (1x PBS)
- 8M GndHCl (in 1x PBS, pH 7.4)
- 4, 4'-Dianilino-1, 1'-Binaphthyl-5, 5'-Disulfonic Acid, Diphosphate salt (Bis-ANS)

Protocol:

Bis-ANS unfolding experiment was performed in a similar manner to equilibrium unfolding, but in the presence of Bis-ANS. The protein sample (2μM concentration) was incubated for at least 3 hours in different concentrations of GndHCl (ranging from 0-6M) and a 5-molar excess of Bis-ANS (final concentration of 10μM). Reading of Bis-ANS fluorescence was obtained from a 1cm pathlength cuvette with the excitation wavelength was 390nm and emission detected from 400-700nm (5nm slits for both excitation and emission wavelengths). The peak fluorescence was plotted against the corresponding GndHCl concentration to analyse the chemical unfolding intermediate.

2.10. Determination of X-ray crystal structure of native conserpin-AAT$_{RCL}$

2.10.1. Generation of native conserpin-AAT$_{RCL}$ crystals

Native conserpin-AAT$_{RCL}$ crystals were generated using the hanging drop vapour diffusion method. The crystallisation conditions screened were based off the conditions native conserpin was crystallised in[150]. All screening buffers contained 0.1mM Bis-Tris, 0.2M Magnesium chloride ($MgCl_2$), while the concentration of polyethylene glycol-3350 (PEG3350) varied from 15-30%**,** and buffer pH varied from pH 5.5-7.5. Freshly purified protein, concentrated to 10mg/ml (234μM), was mixed with equal amounts of crystallisation buffer (1μl protein mixed with 1μl mother liquor), before being sealed and stored at room temperature. Protein crystals appeared in the crystallisation buffer 0.1mM Bis-Tris, 0.2M $MgCl_2$, 20% PEG-3350 pH 6.5 after 5 days.

2.10.2. Data collection and determination of native conserpin-AAT$_{RCL}$ crystal structure

Diffraction data for native conserpin-AAT$_{RCL}$ crystal was collected at the Australian Synchrotron macro crystallography MX2 beamline by Sheena McGowan (Monash University). A 2.2Å dataset was collected, but resolution cut to 2.48 Å. The diffraction images were processed using iMOSFLM[169]. Scaling of the processed data was performed using SCALA[170] in the CCP4 suite[171] in the C 2 2 2_1 space group. The structural determination was performed by molecular replacement (MR) and the PHASER[172] program, using the native conserpin structure as a search model (PDB ID: 5CDX)[150]. The model was built, refined and structural validation was done using PHENIX[173] and Coot[174] with the $2F_C$-F_O and F_O-F_C electron density maps at 1σ. Any side chains not present in the electron density were cut to carbon-α, while loops without density (e.g. the reactive centre loop) were completely removed in the final model.

2.11. Data analysis and production of figures

All data analysis was performed in Graphpad Prism version 7 for Macintosh, GraphPad Software, La Jolla California, USA, www.graphpad.com. The structural figures were produced using Pymol version 2.0.4[175]. All figures were finalised using Adobe Illustrator 2014.

Chapter 3: Reactive centre loop dynamics and serpin specificity

Reactive centre loop dynamics and serpin specificity

Emilia M. Marijanovic[1], James Fodor[1], Blake T. Riley[1], Benjamin T. Porebski[1,2], Mauricio G. S. Costa[3], Itamar Kass[4], David E. Hoke[1], Sheena McGowan[5], Ashley M. Buckle[1*]

[1]Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Victoria 3800, Australia.

[2]Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom

[3]Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro - RJ, Brazil

[4]Amai Proteins, Prof. A. D. Bergman 2B, Suite 212, Rehovot, 7670504, Israel

[5]Biomedicine Discovery Institute, Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia

*To whom correspondence should be addressed: Ashley M. Buckle: Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Victoria 3800, Australia; ashley.buckle@monash.edu; Tel: +61399029313

Keywords: Serpin, protein engineering, protein stability, molecular dynamics, conformational change, protease inhibitor

Running title: Serpin RCL dynamics and specificity

Abstract

Serine proteinase inhibitors (serpins), typically fold to a metastable native state and undergo a major conformational change in order to inhibit target proteases. However, conformational lability of the native serpin fold renders them susceptible to misfolding and aggregation, and underlies misfolding diseases such as α_1-antitrypsin deficiency. Serpin specificity towards its protease target is dictated by its flexible and solvent exposed reactive centre loop (RCL), which forms the initial interaction with the target protease during inhibition. Previous studies have attempted to alter the specificity by mutating the RCL to that of a target serpin, but the rules governing specificity are not understood well enough yet to enable specificity to be engineered at will. In this paper, we use *conserpin*, a synthetic, thermostable serpin, as a model protein with which to investigate the determinants of serpin specificity by engineering its RCL. Replacing the RCL sequence with that from α1-antitrypsin fails to restore specificity against trypsin or human neutrophil elastase. Structural determination of the RCL-engineered conserpin and molecular dynamics simulations indicate that, although the RCL sequence may partially dictate specificity, local electrostatics and RCL dynamics may dictate the rate of insertion during protease inhibition, and thus whether it behaves as an inhibitor or a substrate. Engineering serpin specificity is therefore substantially more complex than solely manipulating the RCL sequence, and will require a more thorough understanding of how conformational dynamics achieves the delicate balance between stability, folding and function required by the exquisite serpin mechanism of action.

Introduction

Over 1,500 serpins have been identified to date. Inhibitory family members typically fold to a metastable native state that undergoes a major conformational change (termed the stressed [S] to relaxed [R] transition) central for the protease inhibitory mechanism[63]. The S to R transition is accompanied by a major increase in stability. The archetypal serpin fold is exemplified by α1-antitrypsin (α1-AT), a single domain protein consisting of 394 residues, which folds into 3 β-sheets (A → C) and 9 α-helices (A → I) that surround the central β-sheet scaffold[176]. The reactive center loop (RCL) protrudes from the main body of the molecule and contains the scissile bond (P1 and P1' residues), which mediates α1-AT's inhibitory specificity against the target protease, neutrophil elastase (HNE).

The inhibitory mechanism of serpins is structurally well understood[63]. Briefly, a target protease initially interacts with and cleaves the RCL of the serpin. However, following RCL cleavage, but prior to the final hydrolysis of the acyl enzyme intermediate, the RCL inserts into the middle of the serpin's β-sheet A to form an extra strand[63,90]. The opening of β-sheet A is controlled by the shutter and the breach regions[71]. Since the protease is still covalently linked to the P1 residue, the process of RCL insertion results in the translocation of the protease to the opposite end of the molecule. In the final complex, the protease active site is distorted and trapped as the acyl enzyme intermediate[63,76].

In certain circumstances the serpin RCL can spontaneously insert, either partially (delta conformation), or fully (latent conformation) into the body of the serpin molecule without being cleaved[92]. Both latent and delta conformations are considerably more thermodynamically stable than the active, native state although they are inactive as protease inhibitors. Folding to the latent conformation is thought to occur via a late, irreversible folding step that is accessible from the native or a highly native-like state[79,87]. As such, transition to the latent state can be triggered by perturbations to the native state via small changes in solution conditions such as temperature or pH[44,76,123], or by spontaneous formation over long time scales[177,178].

Human α1-AT is an extremely potent inhibitor of its target protease HNE, with a rate of association (k_{ass}) 6×10^7 M^{-1} s^{-1}, forming a serpin−protease complex that is stable for several days[179,180]. The metastable nature of α1-AT is required to facilitate the large conformational change required for its inhibitory function, and the rate of RCL insertion into β-sheet A is the main determinant of whether the acyl linkage between serpin and protease is maintained or disrupted. If RCL insertion is rapid, the inhibitory pathway proceeds. If the RCL insertion is too slow, the serpin becomes a substrate; the de-acylation step of the protease's catalytic mechanism is complete and

cleavage of the P1−P1' bond occurs without protease inhibition. The cleaved, de-acylated RCL still inserts into the body of the serpin, resulting in an inactive inhibitor[93].

Two regions of the RCL appear to govern inhibitory function and specificity. The first, a highly-conserved hinge region (resides P15−P9) consisting of short chain amino acids, facilitates RCL insertion into the A β-sheet. Mutations in the hinge region result in the serpin becoming a substrate rather than an inhibitor[181]. The second region is the P1 residue, thought to determine specificity towards a protease. Serpins with a P1 arginine (e.g. antithrombin III) are known to target proteases of the coagulation cascade, including thrombin and Factor Xa[182,183]. In α1-AT, mutation of P1 methionine to arginine (the Pittsburgh mutation), changes the specificity from HNE to thrombin, resulting in a bleeding disorder[184].

Given the importance of the RCL, it has been the focus of previous attempts aimed at altering serpin specificity, via mutation of RCL residues or swapping RCL sequences between serpins. Chimeric serpins have been made between plasminogen activator inhibitor-1 (PAI-1) and antithrombin-III (ATIII)[185,186], α1-AT and antithrombin-III[187,188], α1-AT and ovalbumin[189], and alpha1-antichymotrypsin (ACT) and α1-AT[180,190,191]. In all cases, specificity could only be transferred partially, as each chimera has a reduced second-order rate constant and a higher SI to a target protease in comparison to the original serpin. The most effective chimera produced, without a cofactor, was ACT with P3−P3' of α1-AT. This chimera achieved a stoichiometry of inhibition (the number of moles of serpin required to inhibit one mole of protease (SI)) of 1.4 and a second-order rate constant (k'/[I]) of 1.1 10^5 M^{-1} s^{-1}, two orders of magnitude slower than that of α1-AT[180]. Therefore, it is highly likely that the determinants of specificity are more complex than the RCL region alone, and other regions may play a role, for example exosite interactions in the serpin−protease complex[186,192–194]

In previous work, we designed and characterized conserpin, a synthetic serpin that folds reversibly, is functional, thermostable and resistant to polymerization[150]. Conserpin was designed using consensus engineering, using a sequence alignment of 212 serpin sequences and determining the most frequently occurring amino acid residue at each position. Since it is thermostable and easier to produce in recombinant form, it is ideally suited as a model in protein engineering studies. Conserpin shares 59% sequence identity to α1-AT, with 154 residue differences scattered throughout the structure. Its RCL sequence is sufficiently different from all other serpins such that it no longer resembles an RCL of any serpin with a known target protease. A recent study that investigated the folding pathway of conserpin engineered the P7-P2' sequence of α1-AT into its RCL[195]. The resulting conserpin/α1-AT chimera inhibits chymotrypsin with an SI of 1.46, however, no SI was calculated against HNE. The chimera forms a weak complex with HNE that

is detectable using SDS-PAGE, however, the majority of the serpin molecules are cleaved without complex formation.

In this study, we have exploited the unique folding characteristics of conserpin and employ it as a model serpin with which to investigate the determinants of specificity. We investigated the effect of replacing the RCL of conserpin with the corresponding sequence from α1-AT on inhibitory specificity towards HNE. Here, the chimera molecule, called conserpin-AAT$_{RCL}$, remains thermostable, yet despite possessing the RCL sequence of α1-AT, specificity against HNE was not restored to the extent of α1-AT. Structural analysis and molecular dynamics simulations indicate that specificity is also governed by other, complex factors involving RCL dynamics, and surface electrostatics of regions external to the RCL.

Results

Biophysical and functional characterisation of a conserpin/α1-AT chimera

With the aim of changing the specificity of conserpin to that of α1-AT, a conserpin/α1-AT chimera was previously produced[195], where 9 residues within the RCL (P7-P2') were swapped with the corresponding residues from α1-AT (Fig. 1A). The resulting chimera, conserpin-AAT_{RCL} (379 aa) has a 61% sequence identity with α1-AT (148 residue differences). Conserpin-AAT_{RCL} was expressed in *E. coli* and purified from the soluble fraction by affinity and size exclusion chromatography as described previously[150].

We first investigated the biophysical properties of conserpin-AAT_{RCL} to ensure that swapping the RCL did not alter them. The majority of serpins irreversibly unfold upon heating with a midpoint temperature transition (T_m) of ~55-65°C[114,121,196]. Using variable temperature far-UV circular dichroism (CD) to measure the thermostability, conserpin-AAT_{RCL} was heated from 35 to 95°C at a rate of 1 °C /min, and upon reaching 95°C, minute changes in signal were observed. Following a subsequent 1 °C /min decrease in temperature from 95 to 35°C, minute changes in signal was observed (Fig. 1B). In addition, far-UV spectral scans before and after thermal unfolding showed minute differences in the signals, suggesting the absence of a large heat-induced conformational change (Fig. 1C). Complete unfolding was only achieved in the presence in 2 M guanidine hydrochloride (GndHCl) with a T_m of 72.2 ± 0.1°C. Upon cooling from 95 to 35°C, no precipitation was observed (Fig. 1D). Thus, high thermostability is consistent with the parent conserpin molecule[150] and indicates that incorporation of the α1-AT RCL does not reduce the thermostability of the conserpin scaffold.

We have previously shown conserpin to be a poor inhibitor of trypsin in comparison to α1-AT (SI = 1.8 vs 1.0 respectively)[150]. Engineering the RCL sequence of α1-AT into conserpin improves the SI against trypsin from 1.8 to 1.64 (conserpin-AAT_{RCL} SI=1.64 ± 0.2 n=3; Fig. 1E, F). Conserpin-$AAT_{RCL,}$ like conserpin, after denaturation and refolding was active against trypsin (SI=2.0). Importantly, conserpin-AAT_{RCL} does not inhibit HNE, the protease target of α1-AT. An SI could not be calculated, as there was residual HNE activity after 30-minute incubation, even with at a 2:1 serpin: protease molar ratio.

If the inhibitory pathway of serpin proceeds faster than the substrate pathway, then the SI will be close to 1. If, however, the inhibitory mechanism is too slow and the substrate pathway occurs, the SI is greater than 1[165]. SDS-PAGE using 1:1 and 2:1 serpin: protease molar ratios reveals a faint complex between conserpin-AAT_{RCL} and HNE, but also showed a large amount of cleaved species compared to the complex formation between α1-AT and HNE (Fig. 1G, SI Fig. 1). Since we observe that conserpin-AAT_{RCL} is able to inhibit trypsin, and is still able to transition to the

latent state upon heating, we hypothesized that the RCL mutations do not prevent its insertion into the A-sheet. We therefore sought to investigate the structure and dynamics of conserpin-AAT_{RCL} in order to identify other factors contributing to its inability to inhibit HNE.

The role of electrostatics in the formation of a serpin:protease complex

To understand if there are any structural changes caused by modifying the RCL, we determined the X-ray crystal structure of conserpin-AAT_{RCL} in the native state (Table S1). The overall structure of conserpin-AAT_{RCL} is identical to that of conserpin—a structural alignment reveals a root mean square deviation (RMSD) of 0.2 Å across all Cα atoms. Like conserpin and indeed many other serpins, the RCL of conserpin-AAT_{RCL} is too flexible to be modelled into the electron density. Therefore, all further analyses were performed with the RCL modelled using the structure of wildtype α1-AT (PDB ID: 3NE4[197]).

Effective serpin inhibition of a protease must involve association to form an encounter complex followed by formation of a stereospecific, high-affinity complex that positions the RCL of the serpin to engage with the protease active site. Given the failure to engineer the RCL for α1-AT specificity and inhibition, we reasoned that surface electrostatics may contribute to the formation and stability of a serpin:protease complex and thus protease inhibition. The electrostatic potential surfaces of conserpin, conserpin-AAT_{RCL} and α1-AT differ in several regions. Both conserpin-AAT_{RCL} and α1-AT feature a large electropositive surface centred around the loop connecting strands 2 and 3 of β-sheet B (s2B and s3B) (Fig. 2B, C). In conserpin-AAT_{RCL}, this patch extends to encompass the D-helix, P9−P1 of the RCL, and strand 2 of β-sheet C (s2C)−helix H (Fig. 2E). The corresponding region on α1-AT is much smaller, covering a region under the RCL, some residues of s1B and its connecting loop to helix G, s4B and s5B (Fig. 2F).

A second difference is seen on the top surface of the serpins, directly beneath the RCL. Differences between α1-AT and conserpin-AAT_{RCL}—particularly in s2C, s3C, and the loop between s3A and s3C—lead to a large difference in charge on the surface beneath P9−P1 (Fig. 3A, B). In conserpin-AAT_{RCL} (and conserpin), this region has a large electropositive potential, while the corresponding region in α1-AT is more neutral in charge (Fig. 3A, B).

Functional requirements of an inhibitory serpin's RCL provide selective pressures on its sequence. In inhibitory serpins, the sequence of the RCL must correspond to the specificity of its target proteases[92], maintain a linear, mobile structure in the stressed/native state, and still remain capable of insertion into highly conserved regions in β-sheet A post-cleavage (an example is the requirement of small residues in the hinge region[97,181,198]). Given these known coevolutionary pressures, it follows that there should be either highly conserved residues which are responsible for conferring this polymorphic behaviour, or a coevolutionary signal present in the sequences of

functionally interacting regions within the serpin. As we were interested in the interactions between the residues of the RCL and residues beneath the RCL, we calculated conservation scores using a sequence alignment of 212 serpin sequences, and mapped them onto the structure of α1-AT (Fig. 3C). Residues facing the P1 and P1' residues of the RCL are well conserved, compared to residues on strands s2C and s3C that face the RCL (under the residues N-terminal to P1). We were unable to identify any significant coevolutionary links between residues of the RCL and the region below it on sheet C, though this is most likely a reflection on the limited number of sequences used.

To further investigate the interactions between the RCL and the body of the serpin, we looked at the frustration networks within conserpin-AAT_{RCL} and α1-AT. Frustration analysis labels pairs of residues as 'frustrated' if their interaction is destabilising compared to other combinations of residues in the same location[199]; clusters of frustrated residues are often found near binding sites, suggestive of a stressed conformational state, or otherwise implicated in the function of the protein[200]. In α1-AT, the RCL is minimally frustrated against the body of the serpin, with only the P12-P9 region present in a patch of high frustration. In contrast, there is a more extensive network of frustration in conserpin-AAT_{RCL}, particularly between the RCL and the loop between s3A and s3C (SI Fig. 2). These distinct frustration patterns reflect the differences we observed in the electrostatics on top of the serpin body (Fig. 3), and suggest that the electrostatic compatibility between the body of the serpin and the RCL plays a key role in determining serpin functionality.

Having established clear differences in the surface electrostatics of the serpins, we next investigated possible consequences for engagement with proteases trypsin and HNE. Given contrasting inhibition of these two proteases we compared their electrostatic potential surfaces. The largest difference between the two proteases is found at the active site. Whereas both proteases feature an electronegative potential in the active site cleft, in trypsin it is more extensive, encompassing S2-S4 binding pockets and the surrounding residues (Fig. 4B, 4E). In contrast, the S3-S4 binding pockets and surrounding residues of HNE contains an adjacent large electropositive patch (Fig. 4E). To observe any electrostatic potential clashes during a hypothetical serpin−protease encounter complex, we modelled a conserpin-AAT_{RCL}: trypsin complex, and a conserpin-AAT_{RCL}: HNE complex, each with P1 $M358_{329}$ in the protease active site (Fig. 4A, D), using the x-ray crystallography structure of a Michaelis complex as a starting model (PDB: 1K9O[106]). The electrostatic potential for each protease and serpin were calculated separately, eliminating the influence of one electrostatic potential onto the other.

The calculated electrostatic potential suggests trypsin has greater electrostatic compatibility with conserpin-AAT_{RCL} than HNE. This compatibility can be attributed to the large electropositive surface of the RCL and the body below the RCL. Compatibility will be essential for the formation

and stability of a Michaelis serpin−protease complex, where there is contact between the binding pockets (S4−S1') of the protease and P6−P1' residues of the RCL. The formation and stability of a serpin−protease complex between conserpin-AAT_{RCL} and trypsin can occur with favourable interaction between trypsin's electronegative S3−S4 pockets (Fig. 4B) and the electropositive potential of conserpin-AAT_{RCL} P6−P3 residues (Fig. 4C). Therefore, conserpin-AAT_{RCL} can inhibit trypsin. In comparison, the formation and stability of a complex may be hindered by the charge−charge repulsion between the electropositive S3−S4 binding pockets of HNE (Fig. 4D) and the electropositive surface of P6−P3 of the RCL (Fig. 4E). As a result, conserpin-AAT_{RCL} behaves as a substrate to HNE rather than as an inhibitor.

RCL dynamics are important for protease inhibition

Given the large conformational changes involved in serpin function, and specifically the central role played by the RCL in protease engagement and subsequent insertion into the A-sheet, an investigation of the dynamics of the RCL of conserpin-AAT_{RCL} may provide some insight into its inhibitory properties. We therefore performed molecular dynamics (MD) simulations of conserpin-AAT_{RCL} and compared the results to those of α1-AT and conserpin simulations we performed previously[150]. Although we are unable to perform simulations for long enough to observe the RCL insertion into the A-sheet, MD is able to reveal the intrinsic dynamics of the RCL and specifically the lifetime of its interactions with the body of the serpin. After reaching equilibrium at around 150 ns, the root mean square deviation (RMSD) indicated that the simulations remained stable with no large conformational changes observed. Given the importance of RCL conformation in facilitating the S→R transition following protease engagement, we analyzed the dynamics of the A-sheet and the RCL, and also the interactions between the RCL and the body of the serpin during the time course of the MD simulations. The central A-sheet contains two conserved regions, the shutter and breach, which are critical for the insertion of the RCL; mutations in these regions often render the serpin susceptible to misfolding and aggregation[71].

Substitution of residues P7−P2' of the RCL for the corresponding region of α1-AT did *not* serve to reduce the flexibility of the RCL region to the lower level observed in α1-AT simulations[150]. Instead, while the region of conserpin-AAT_{RCL} around residues 353_{324}-362_{333} showed a reduced root mean square fluctuation (RMSF) (corresponding to less conformational variability), the region around residues 342_{314}-352_{323} of the RCL showed an increased RMSF (Fig. 5D). This increase in flexibility is evident in comparing MD snapshots of the three systems, where the substantially increased flexibility of the lower RCL region of conserpin-AAT_{RCL} is clearly evident (Fig. 5A-C). This highly dynamic region encompasses the hinge region of the RCL, the first residues that insert into the A-sheet.

As the fate of the serpin as either a substrate or inhibitor is determined by the competition between rates of RCL insertion and the de-acylation of the protease[93], it is likely that an increase in RCL loop dynamics would slow the rate of insertion. This would allow a protease with a fast catalytic mechanism (such as HNE), to escape inhibition, thereby pushing the serpin down the substrate pathway.

To understand the RCL conformations adopted by conserpin-AAT$_{RCL}$ throughout the simulations, we performed principal component analysis on the conformations of the RCL backbone (between P17-P1') over all simulations (α1-AT, conserpin and conserpin-AAT$_{RCL}$), followed by a clustering. This produced a total of 9 clusters, with RCL conformations within each cluster being structurally close but clearly distinguishable from others (SI Fig. 3). These analyses show that α1-AT's RCL maintains a reasonably close set of conformations throughout the three independent simulations, while conserpin's RCL explores a broad variety of conformations that are exclusive of those explored by α1-AT's RCL (Fig. 6A). Conserpin-AAT$_{RCL}$'s RCL not only adopts conformations that overlap with those of the other serpins, but also explores conformations that were not seen in α1-AT and conserpin simulations.

Conserpin's RCL explored 5 different conformations (5 clusters), most of which have an extended conformation in which the hinge region (P12-P9) of the RCL is moved away from the breach region of β-sheet A (Fig. 6B). This preference for an extended RCL hinge in conserpin is surprising, as conserpin has an extended salt bridge network in the breach region in comparison to α1-AT[150], which was hypothesised to stabilise conserpin's native state.

For α1-AT, the RCL explores 2 similar conformations, with both conformations containing the hinge region primed for insertion between s3A and s5A. This expands on the RMSF analysis (Fig. 5D), where α1-AT's RCL was seen to be relatively rigid over the course of the simulations (in comparison to conserpin and conserpin-AAT$_{RCL}$). The rigidity of α1-AT's RCL suggests that there are interactions between the body and the RCL that reduce the dynamics of the loop, and possibly prime the hinge region between s3A and s5A strands.

The RCL of conserpin-AAT$_{RCL}$ explores 4 conformations: one overlapping with a cluster seen in conserpin, one overlapping with a cluster seen in α1-AT, and 2 conformations unique to conserpin-AAT$_{RCL}$. All of these conformations (except the α1-AT-like one) include an extended hinge region away from β-sheet A. This could possibly be a consequence of the interactions between the residues on β-sheet C and the RCL, as stated previously[190]. Interestingly, one of the conformations include a slight helical turn from P10-P7 (similarly to an α1-AT / α1-antichymotrypsin chimera[201]), possibly responsible for pulling the hinge region away from β-sheet A. Importantly, one of conserpin-AAT$_{RCL}$'s RCL conformation is similar to α1-AT's, where the hinge region is primed to insert into β-sheet A. The ability of conserpin-AAT$_{RCL}$ to access this

conformation may explain the increase in inhibitory activity against trypsin over conserpin, as the conserpin-AAT_{RCL} RCL could insert faster than conserpin from this pose. However, despite this primed hinge region conformation, conserpin-AAT_{RCL} primarily remains a substrate against HNE. It is possible that HNE could negatively impact on the conformation of the RCL upon encounter, or even prevent formation of a stable serpin: protease complex, a scenario in which HNE's catalytic mechanism occurs more rapidly than trypsin, allowing for rapid cleavage of the RCL followed by substrate rather than inhibitor behaviour. A structural difference was also observed by calculating the phi-psi angles of the RCL for each serpin over the course of the simulations. Replacement of P7-P2' of α1-AT onto conserpin has failed to reproduce the conformational pattern seen in α1-AT. Specifically, while the conformations adopted by conserpin-AAT_{RCL} in the region around residues 353_{324}-362_{333} (SI Fig. 4, red) are more similar to those of α1-AT than conserpin (SI Fig. 4, blue and black, respectively), those in the region around residues 342_{314}-352_{323} (SI Fig. 4) show a distinctly different set of conformations. Together these observations indicate that the conformational landscape sampled by the structure in and around the RCL region, including areas near the breach region, is important in the process of RCL insertion, and thus ultimately serpin inhibitory specificity.

Discussion

Conserpin shares high sequence identity to α1-AT (59%), is extremely stable, polymerisation-resistant and yields large quantities when expressed through recombinant techniques. It is therefore an attractive model system for investigating the folding, stability and function of serpins. In this study, to investigate the determinants of serpin specificity, we used a conserpin/α1-AT chimera, which we call conserpin-AAT_{RCL}, in which the residues in the RCL are replaced with those of α1-AT. The resulting hybrid retained the thermostability and polymerisation resistance of conserpin. However, despite containing the RCL sequence of α1-AT, which is thought to be a key determinant of inhibitory specificity, conserpin-AAT_{RCL} showed only minor improvement as an inhibitor of trypsin, in comparison to conserpin, and like conserpin behaved mostly as a substrate against HNE.

We attempted to rationalise the substrate behaviour of conserpin-AAT_{RCL} using a structural and molecular modelling/simulation approach. Although the X-ray crystal structure of conserpin-AAT_{RCL} revealed no significant differences with the parent molecule, we were able to provide insights into the failure to transfer specificity by analysing electrostatic differences and changes in the flexibility of RCL, hinge, breach and shutter regions with molecular dynamics simulations.

For a serpin to perform its inhibitory function, the serpin and protease must come into contact with each other. Reasoning that, like other protein−protein complexes[202,203], cognate serpins and proteases must exhibit complementary electrostatic surfaces to ensure rapid, and high affinity association, we identified several differences between the electrostatic surface characteristics of α1-AT and conserpin-AAT_{RCL} that may contribute to their contrasting inhibitory properties. HNE contains a shallow active site that interacts with P6−P3' residues of the RCL, therefore the electrostatic surface of this region must be complementary to ensure efficient binding to the P1 methionine residue[191,204]. In comparison to α1-AT, conserpin-AAT_{RCL} harbours several regions where poor charge complementarity may explain the diminished capacity to form a complex with HNE, and subsequently why it acts as a substrate rather than an inhibitor. One of these regions includes the electrostatic potential beneath the RCL. The role of electrostatics has been investigated for several serpins. For example, single-pair Förster resonance energy transfer (spFRET) studies of the inhibition of anionic rat and cationic bovine trypsin by α1-AT showed only partial translocation of anionic rat trypsin compared to full translocation of cationic bovine trypsin[205,206]. This indicates that the electrostatic potential between the protease and serpin are important for formation of a serpin:protease complex and protease inhibition. Similarly, for the serpin PAI-1, the Michaelis complex between tissue-type plasminogen activator (tPA) and PAI-1 was observed to have more complementary electrostatic interactions than the complex between urokinase-type plasminogen activator (uPA) and PAI-1. This was used to explain the difference in

second-order inhibitory rate constants between the two proteases: tPA is inhibited at a faster rate (2.6×10^7 M^{-1} s^{-1}) compared to uPA (4.8×10^6 M^{-1} s^{-1}) [207,208]. Furthermore, an arginine to glutamic acid substitution produced a 'serpin-resistant' tPA variant, where the glutamic acid produced a repulsion to PAI-1, leading to a failure to inhibit this variant. This tPA variant was only inhibited through creating a complementary PAI-1 with the opposite glutamic acid to arginine mutation[209,210], further emphasising the importance of surface potential in the formation of a stable serpin:protease complex. Any possible repulsive interactions may destabilize a serpin:protease complex and therefore prevent inhibition.

Dynamics in the RCL is important for its insertion into β-sheet A during protease inhibition. We therefore investigated the difference in RCL dynamics between conserpin-AAT_{RCL} and α1-AT using molecular dynamics simulations. Despite P7-P2' of conserpin-AAT_{RCL} being identical to the corresponding region in α1-AT, the overall flexibility of the RCL region as a whole was not reduced to the level of α1-AT, while the hinge region of the RCL, which inserts into β-sheet A first during insertion, exhibited higher flexibility in conserpin-AAT_{RCL} compared to α1-AT. A plausible explanation for this is the additional residue at P2'. Conserpin was designed without an isoleucine at P2', producing an RCL length that fits onto the serpin body. The addition of P2' isoleucine in conserpin-AAT_{RCL} may force the RCL to adopt a non-ideal conformation, possibly increasing the dynamics of the hinge region.

The conformation of the RCL is likely highly tailored to the particular inhibitory specificity of each serpin. α1-AT, a potent inhibitor of HNE, has an RCL that is in a primed position for insertion into β-sheet A. That is, the hinge region is poised between strands 3A and 5A, allowing for rapid insertion during HNE inhibition. Conserpin and conserpin-AAT_{RCL} contain RCL conformations that are extended, with the hinge region away from β-sheet A, likely reducing the rate at which the RCL can insert. One conformation that conserpin-AAT_{RCL} explores contains a primed hinge region, which possibly explains the increase in its inhibitory activity against trypsin (compared to conserpin), but is not enough to produce inhibition against HNE. Furthermore, the breach region of α1-AT 'loosens' and opens over the course of the simulations[150], while the breach region in conserpin and conserpin-AAT_{RCL} remains rigid due to the extended salt bridge network. Therefore, it is possible that α1-AT inhibits HNE at a rapid rate due to the primed position of the hinge region and opening of the breach region, allowing inhibition before HNE's de-acylation step of cleavage. Our observation that the RCL of conserpin-AAT_{RCL} sampled this primed hinge conformation out of 4 possible conformations, and the relatively rigid nature of its breach region, suggests that the initial steps of RCL insertion into A-sheet are slower than the de-acylation step of HNE's cleavage.

Previous studies that have attempted to convert the specificity of ACT to that of α1-AT by swapping RCL residues have been generally unsuccessful, as the chimeras had a greater SI and

slower inhibitory rate compared to α1-AT. This suggests that other factors may play important roles, including interactions between the RCL and the body of the serpin, and the structure of the chimeric RCL[180,190,191,201]. Engineering of ACT/α1-AT chimeras show that HNE's proteolytic mechanism occurs on a shorter timescale in comparison to ACT's catalytic mechanism[180]. It is also known that an increase in the dynamics of the RCL can affect the serpin's ability to inhibit a protease. Notably, loss of a salt bridge in the breach region in α1-AT Z variant increases RCL dynamics and subsequently leads to an SI increase (from 1.0 to 1.8) and decrease in rate of inhibition (from 6.9 to 2.3 × 10^6 M^{-1} s^{-1})[78,135,137,138]. This implies that the rate of RCL insertion occurs slower than the de-acylation step of HNE's catalytic mechanism, producing a substrate rather than an inhibitor of HNE. It is likely that RCL-protease interactions will vary for each protease, influencing the dynamics of the RCL[193] and the conformational change needed for RCL insertion[211]. With the use of fluorescent labels, it was observed that the two protease targets of plasminogen activator inhibitor-1 (PAI-1), tissue-type plasminogen activator (tPA) and urokinase-type plasminogen activator (uPA), rests differently on the P1−P1' bond and change the dynamics of the RCL when bound. tPA affects the C-terminus of the RCL through exosite interactions, while retaining dynamics observed with free PAI-1. In contrast, uPA affects the N-terminus with different exosite interactions, restricting the dynamics and immobilising the RCL. This difference in RCL dynamics also contributes to the difference in the rates the proteases are inhibited by PAI-1[193]. Taken together, the dynamics of the RCL is critical for the rate of insertion during protease inhibition. Fast insertion favours protease inhibition while slow insertion forces the serpin to undergo the substrate pathway.

Along with the possibility of electrostatic repulsion and increased RCL dynamics, the failure to transfer specificity onto conserpin-AAT_{RCL} could be a consequence of the delicate balance between stability and function. Serpins use the metastable conformation to undergo the large conformational change necessary for its inhibitory function. Increasing the stability of this metastable state may decrease the dynamics and plasticity required to undergo the S→R transition during inhibition of a target protease. For example, increasing the stability of α1-AT more than 13 kcal mol^{-1} than the wild type α1-AT compromises its inhibitory activity[212]. For conserpin, the very high stability, although still functional, can be attributed to certain key regions important for the serpin's inhibitory mechanism and S→R transition[150]. Structural plasticity is required in the breach region, as this region is important in controlling the insertion of the RCL and conformational change to allow for protease inhibition. The extensive salt-bridge network in the breach region in conserpin and conserpin-AAT_{RCL} increases rigidity and slows the opening of β-sheet A between strands s3A and s5A. The rigidity of the breach, along with displacement of conserpin and conserpin-AAT_{RCL} hinge region away from β-sheet A, may explain the reduced inhibitory activity

of conserpin towards trypsin, and the failure of conserpin-AAT_{RCL} to inhibit HNE. Furthermore, helix F, which packs tightly against the A-sheet, may act as a barrier to RCL insertion via A-sheet opening, and must partially unfold to allow rapid RCL insertion[64,81,99]. Mutations on the helix F/A-sheet interface of α1-AT can relieve this tight packing, increasing the stability but also decreasing activity. Therefore, the tight packing between helix F and A-sheet contributes to the metastability and that is relieved in the S→R transition. In conserpin, this interface is tightly packed, but not to the extent of α1-AT. As a result, the tight packing at this interface may not have the strain observed in α1-AT, slowing the partial unfolding of helix-F to allow for rapid RCL insertion.

In conclusion, we utilized a serpin chimera to investigate the rules that govern serpin specificity, by studying the effect of replacing the RCL of conserpin, a model synthetic serpin, with the corresponding sequence from α1-AT. Despite possessing the RCL sequence of α1-AT, specificity against trypsin or HNE was not restored to that of α1-AT. Crystal structural analysis and molecular dynamics simulations indicate that, although the RCL sequence may partially dictate specificity, electrostatic surface potential coupled with dynamics in and around the RCL likely play an important role. Although beyond the scope of the current study, systematic mutational studies on conserpin-AAT_{RCL} that alter its electrostatic complementarity with HNE will ultimately allow our hypotheses to be tested. The dynamics of the RCL appears to govern the rate of insertion during protease inhibition, dictating whether it behaves as an inhibitor or a substrate. The unusual mechanism of serpin action also requires a delicate balance between stability, dynamics and function[213,214]. Engineering serpin specificity is therefore substantially more complex than solely manipulating the RCL sequence, and although may be guided by the general principles discussed in this work, each serpin will most likely present unique challenges. Notwithstanding this, further characterisation of the role of dynamics will be required to advance our understanding of how serpins perform their exquisite inhibitory functions.

Experimental procedures

Design of conserpin-AAT$_{RCL}$

The design of conserpin-AAT$_{RCL}$ was based of the RCL sequence of α1-AT. Residues P_7-P_2' of the α1-AT RCL were mutated onto the original conserpin molecule to provide specificity against trypsin and neutrophil elastase. The residue numbering adheres to that adopted as previous[150]: $Q105_{\alpha1\text{-}AT}$ and corresponding conserpin-AAT$_{RCL}$ residue $R79_{conserpin\text{-}AATRCL}$ is written as $Q105R_{79}$.

Expression constructs

The plasmid encoding conserpin-AAT$_{RCL}$ was generated using ligation-independent cloning with the pLIC-HIS vector[161] using standard protocols, adding an N-terminal 6His-tag to conserpin-AAT$_{RCL}$; this construct was transformed into BL21(DE3) pLysS *E. coli*.

Protein expression and purification

Protein was expressed 2xYT media and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 1. Expression was continued for 3 hours before cells were harvested and lysed in 10 mM imidazole, 50mM NaH_2PO_4, 300 mM NaCl, pH 8.0. Following centrifugation, batch bound to nickel-NTA loose resin (Qiagen) and washed with 50 mL of 20 mM imidazole, 50mM NaH_2PO_4, 300mM NaCl, pH 8.0. Any conserpin-AAT$_{RCL}$ bound to the nickel-NTA resin was eluted with 250mM imidazole, 50mM NaH_2PO_4, 300mM NaCl, pH 8.0, into 5mL fractions. Fractions containing conserpin-AAT$_{RCL}$ were loaded into a Superdex 200 16/60 column for further purification and eluted with 50mM tris-HCl, 150mM NaCl pH 8.0. The N-terminal His-tag remained attached to conserpin-AAT$_{RCL}$.

Characterisation of inhibitory properties

The stoichiometry of inhibition against bovine trypsin (Sigma-Aldrich) was performe similarly as described[150,165]. Briefly, various concentrations of conserpin-AAT$_{RCL}$ (0−200 nM in 25 nM increments) was incubated with a constant trypsin (105 nM) concentration at 37°C for 30 min in 50mM tris-HCl, 150mM NaCl, 0.2% v/v PEG 8000 pH 8.0. The residual trypsin activity was measured at 405nm using the substrate Na-benzoyl-L-arginine 4-nitroanilide hydrochloride (Sigma-Aldrich).

To test for activity after refolding, conserpin-AAT$_{RCL}$ was unfolded in 6M guanidine hydrochloride (GndHCl) 50mM tris-HCl, 150mM NaCl pH 8.0 for 2 hours before refolding via dilution for another 2 hours, so the final concentration of guanidine hydrochloride was 0.2M. Any aggregate was pelleted by centrifugation and the sample dialysed against the same buffer to

remove any remaining GndHCl. The SI assay against trypsin was performed as stated above (constant trypsin concentration of 210nM and varying conserpin-AAT_{RCL} concentrations from 0−450nM in 50nM increments).

To observe an SDS-stable serpin: protease complex, different ratios of serpin were incubated with protease for 30 minutes at 37°C. Reducing SDS sample buffer was added to each sample and quenched on ice to stop any further reaction. Samples were loaded onto a 10% SDS-PAGE.

Circular dichroism scans and thermal denaturation

Circular dichroism (CD) measurements were performed on a Jasco J-815 CD spectrometer at a protein concentration of 0.2mg/mL with PBS using a quartz cell with a path-length of 0.1cm. Far-UV scans were performed at 190−250nm. For thermal denaturation, a heating rate of 1°C/min from 35°C to 95°C was used, with the change in signal measured at 222nm. For samples containing 2 M GndHCl, refolding was measured directly after the thermal melt by holding the temperature at 95°C for 1 min before the temperature was decreased to 35°C at the same rate. The midpoint of transition (T_m) was obtained by fitting the data with a Boltzmann sigmoidal curve in accordance with the method described[150] for both forward and reverse thermal denaturation experiments.

Crystallization, X-ray data collection, structure determination and refinement

Crystals of conserpin-AAT_{RCL} were obtained using hanging drop vapour diffusion, with 1:1 (v/v) ratio of protein to mother liquor (1µL of conserpin-AAT_{RCL} mixed with 1µL of mother liquor. The protein was concentrated to 10mg/mL and crystals appeared in 0.2M magnesium chloride, 0.1mM Bis-Tris and 20% PEG 3350, pH 6.5 after 5 days.

Diffraction data was collected on the MX2 beamline at the Australian Synchrotron. The diffraction data was processed with iMOSFLM[169] to 2.48Å, followed by scaling with SCALA[170] in the CCP4 suite[171]. The structure was determined by molecular replacement (MR) with Phaser[172] using the conserpin structure (native state) as a search probe (PDB 5CDX)[150]. The model was built and refined using PHENIX[173] and Coot[174].

Computational resources

Atomistic MD simulations were performed on Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE), and in-house hardware (NVIDIA TITAN X Pascal GPU).

Atomic coordinates, modelling and graphics

The RCL was modelled onto the x-ray crystal structure using MODELLER[215]. In MD simulations, atomic coordinates were obtained from the following PDB entry: 3NE4[197]. α1-AT and conserpin molecular dynamics simulations used for the analysis were run previously in the original Conserpin paper[150]. The residue numbering remained as determined by crystal structure, that is, the glutamine from the TEV cleavage tag remained as residue -1. Structural representations were produced using PyMOL version 2.0.4[175] and VMD 1.9.4[216]. Trajectory manipulation and analysis was performed using MDTraj[217] and VMD 1.9.4[216]. Electrostatic calculations were performed with the APBS plugin[218,219] on PyMOL. Serpin:protease complexes were modelled based on the X-ray crystal structure of a serpin:protease Michaelis complex (*Manduca sexta* serpin 1B with rat trypsin (S195A), PDB: 1K9O)[106].

Molecular dynamics (MD) systems setup and simulation

Each protein, with protonation states appropriate for pH 7.0 as determined by PROPKA[220,221], was placed in a rectangular box with a border of at least 10 Å, explicitly solvated with TIP3P water[222], sodium counter-ions added, and parameterized using the AMBER ff99SB all-atom force field[223–225]. After an energy minimization stage consisting of at least 10,000 steps, an equilibration protocol was followed in which harmonic positional restraints of 10 kcal Å mol^{-1} were applied to the protein backbone atoms. The temperature was incrementally increased while keeping volume constant from 0K to 300K over the course of 0.5ns, with Langevin temperature coupling relaxation times of 0.5ps. After the target temperature was reached, pressure was equilibrated to 1atm over a further 0.5ps using the Berendsen algorithm[226]. Following equilibration, production runs were performed in the NPT ensemble using periodic boundary conditions and a time step of 2 fs. Temperature was maintained at 300K using the Langevin thermostat with a collision frequency of 2 ps, and electrostatic interactions computed using an 8Å cutoff radius and the Particle Mesh Ewald method[227]. Three independent replicates of each system were simulated for 500 ns each using Amber 14[228]. The three independent replicates for each system were concatenated, and RMSD, RMSF, and phi and psi angles computed over 500ps timesteps using VMD 1.9.4[216].

Sequence methods

In calculating construct sequence identities, construct sequences were aligned using MUSCLE[229,230] v3.8.1551. The 6xHIS-TEV-SacII N-terminal peptide was removed from the alignment so as not to inflate alignment statistics. Percentage identities were calculated as %id = 100% × number of identity columns / length of aligned region (including gaps).

Mapping of sequence conservation on structure α1-AT[231] was performed using the Consurf 2016 server[232] using the previously designed alignment of 212 serpins[97]. Sequence coevolution

analysis was performed using the OMES χ^2 residue independence test[233], as well as the SCA[234] and ELSC[235] perturbation-based residue covariance methods.

RCL principal component analysis & clustering

From the nine trajectories described above, trajectories of the 72 atoms describing the backbone (N, CA, C, O) from P17 (E342)−P1' (S359) were extracted using MDTraj[236]. These trajectories were concatenated together into a 8993-frame trajectory, and Scikit-learn[236] was used to calculate eigenvectors describing 216 principal component vectors. The top three PCA vectors describe 35.64%, 16.67%, and 11.72% respectively of the variance across all conformations in the concatenated trajectory. The nine trajectories were then transformed into this PCA space, and plotted using matplotlib[237].

The concatenated trajectory, as expressed in PCA coordinates, was clustered using the HDBSCAN algorithm[238,239], using default parameters, except a minimum cluster size of 1% of the total trajectory (90 frames).

Frustration calculation

Local frustration analysis of the modelled serpin: protease complexes was conducted with the Frustratometer2 web server[240]. Essentially, the energetic frustration is obtained by the comparison of the native state interactions to a set of generated "decoy" states where the identities of each residue are mutated. The constant *k* used to model the electrostatic strength of the system was set to its default value (4.15). A contact is defined as "minimally frustrated" or "highly frustrated" upon comparison of its frustration energy with values obtained from the decoy states.

Acknowledgements

BTP is a Medical Research Council Career Development Fellow. SM acknowledges fellowship support from the Australian Research Council (FT100100960). We thank the Australian Synchrotron for beam-time and technical assistance. This work was supported by the Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au). We acknowledge the Monash Protein Production Unit and Monash Macromolecular Crystallization Facility.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

EMM and AMB designed the study. EMM performed the protein expression, purification, CD thermal melt and assay experiments with assistance from BTP and DEH. EMM, SM and AMB performed the crystallography, with assistance from BTP. JF performed molecular dynamics simulations. JF and BTR analyzed simulations, with assistance from IK. MGSC performed frustration analysis. EMM and BTR generated figures. EMM, BTR, SM and AMB wrote the manuscript.

Figures

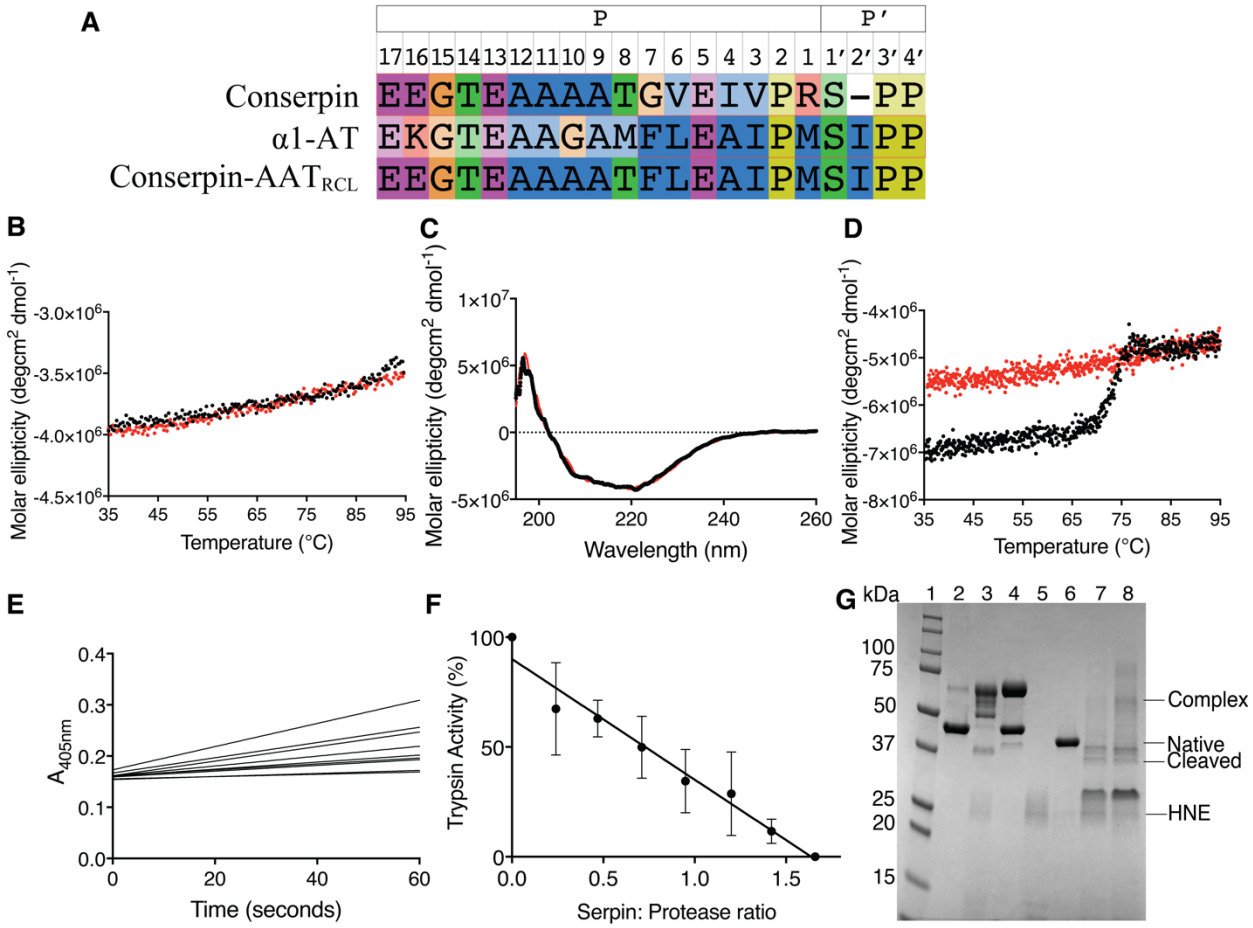


Figure 1. Stability and inhibitory activity of conserpin-AAT$_{RCL}$. (A) RCL sequence alignment indicating which residues of conserpin were replaced with the corresponding residues in α1-AT; (B) Variable temperature thermal melt of conserpin-AAT$_{RCL}$, heating to 95°C (black line) and cooling to 35°C (red line), measured by CD at 222 nm; (C) Spectral scan before (black line) and after (red line) variable temperature thermal melt; (D) Variable temperature thermal melt in the presence in 2 M GndHCl (heating to 95°C; black line, cooling: red line); (E) Inhibitory activity assay and (F) SI against trypsin (n=3); (G) A cropped SDS-PAGE showing a serpin:protease complex formed between HNE and AAT, but less complex formed between HNE and conserpin-AAT$_{RCL}$. From left to right: 1. Molecular weight markers (kDa); 2. α1-AT alone; 3. 1:1 ratio of α1-AT: HNE; 4. 2:1 ratio of α1-AT:HNE; 5. HNE alone; 6. conserpin-AAT$_{RCL}$ alone; 7. 1:1 ratio of conserpin-AAT$_{RCL}$:HNE; 8. 2:1 ratio of conserpin-AAT$_{RCL}$:HNE. The full length SDS-PAGE gel is presented in Supplementary Figure 1.

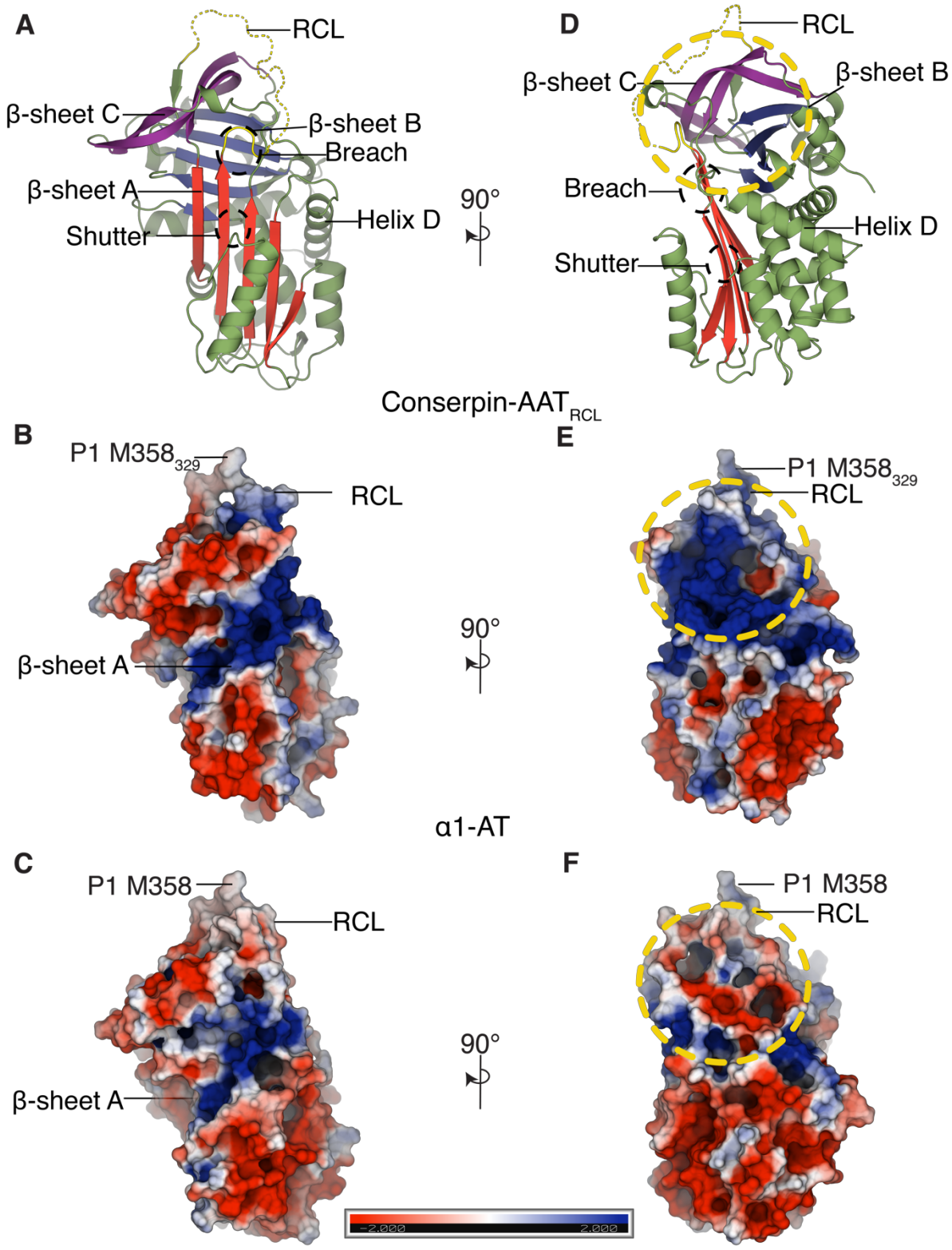


Figure 2. Structure and electrostatics of conserpin-AAT$_{RCL}$. (A, D) X-ray crystal structure of native state conserpin-AAT$_{RCL}$ represented as a cartoon. The breach and shutter regions are marked with black broken circles. (B–F) A comparison of electrostatic potential surfaces (blue=+ve, red=-ve) of (B, E) conserpin-AAT$_{RCL}$ and (C, F) α1-AT. Both conserpin-AAT$_{RCL}$ and α1-AT feature a large electropositive surface centred around the loop connecting strands 2 and 3 of β-sheet B (s2B and s3B) (B, C). A large surface patch between helix D and the RCL, highlighted with yellow

broken circles, has a generally positive potential in conserpin-AAT$_{RCL}$ (E), and negative potential in α1-AT (F).

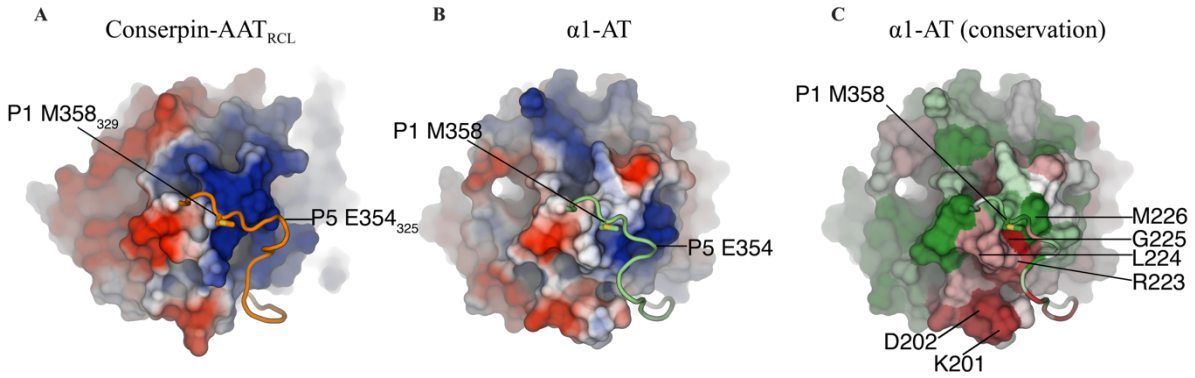


Figure 3. Electrostatic potential surfaces of the RCL differs between conserpin-AAT$_{RCL}$ and α1-AT. While we have grafted the α1-AT RCL (cartoon) from P7−P2' onto conserpin (surface), the electrostatic surface potential between conserpin-AAT$_{RCL}$ and α1-AT differs beneath the RCL. (A) In conserpin-AAT$_{RCL}$, the region below the RCL contains a large electropositive potential, while in α1-AT (B), the corresponding region is more neutral in charge. (C) ConSURF conservation scores for the serpin superfamily, mapped onto the surface of α1-AT as colours from forest green (highly conserved) to brick red (highly variable). This depicts poor conservation (red) of residues 201−202 and 223−225 of α1-AT, suggesting that these residues may be responsible for contributing to protease specificity within the serpin family.

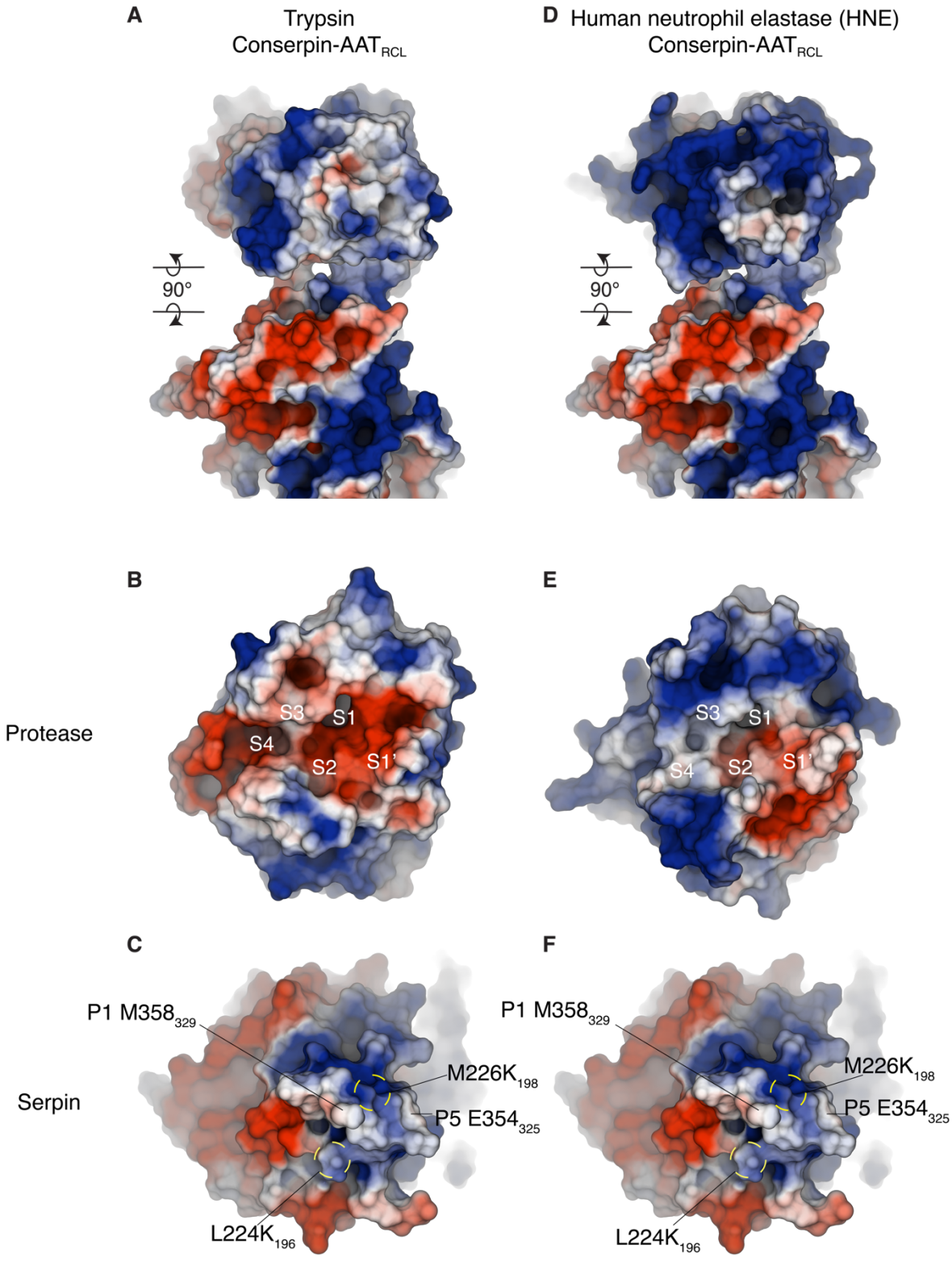


Figure 4. Electrostatic compatibility between serpin and protease (A) Electrostatic surfaces of a modeled complex between trypsin and conserpin-AAT$_{RCL}$, and (D) between HNE and conserpin-AAT$_{RCL}$. Associated complexes are separated into individual proteins by rotating each molecule by 90° around the horizontal axis in the plane of the paper (clockwise for the top molecules, anti-clockwise for the bottom molecules). (B) Electrostatic surface for the active site of trypsin and (E) HNE shows that trypsin has a more electronegative binding cleft than HNE. Comparing this to the

electrostatic surface of (C, F idem.) conserpin-AAT$_{RCL}$ suggests a greater electrostatic compatibility between trypsin, particularly the electropositive surface below the RCL. However, the electropositive surface of S3-S4 binding pocket in HNE suggests there may be a charge repulsion with the electropositive surface potential of conserpin-AAT$_{RCL}$ at P6-P3.

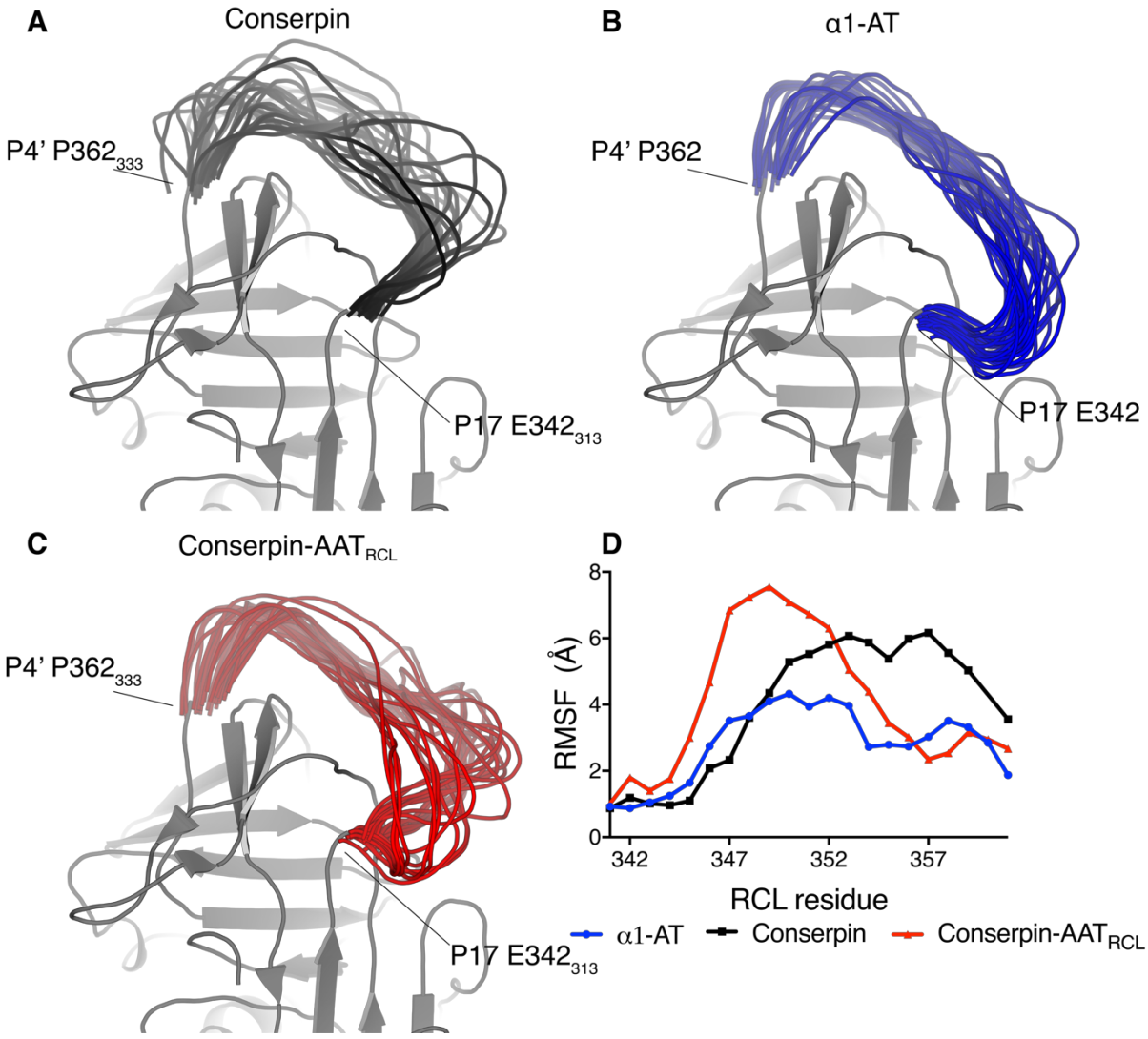


Figure 5. The dynamics of the RCL is important for inhibition. Snapshots of conformations of the RCL from the MD runs at 50 ns intervals overlaid on static structure for the rest of the molecule, showing that (A) conserpin prefers an extended-hinge RCL conformation, (B) α1-AT prefers a bent-hinge RCL conformation, and (C) conserpin-AAT$_{RCL}$ occupies both of these conformations. the increased flexibility of the lower RCL region (residues 342$_{314}$-352$_{323}$) relative to both conserpin and α1-AT. (D) Root mean square fluctuation (RMSF) calculated for the RCL region from the molecular dynamics simulations shows that the conserpin-AAT$_{RCL}$ (red) has lower flexibility than conserpin (black) in the 353$_{324}$-362$_{333}$ region but a higher flexibility in the 342$_{314}$-352$_{323}$ region than conserpin and α1-AT (blue) (α1-AT numbering), reflecting the structural differences between the two conformational clusters occupied by the RCL.

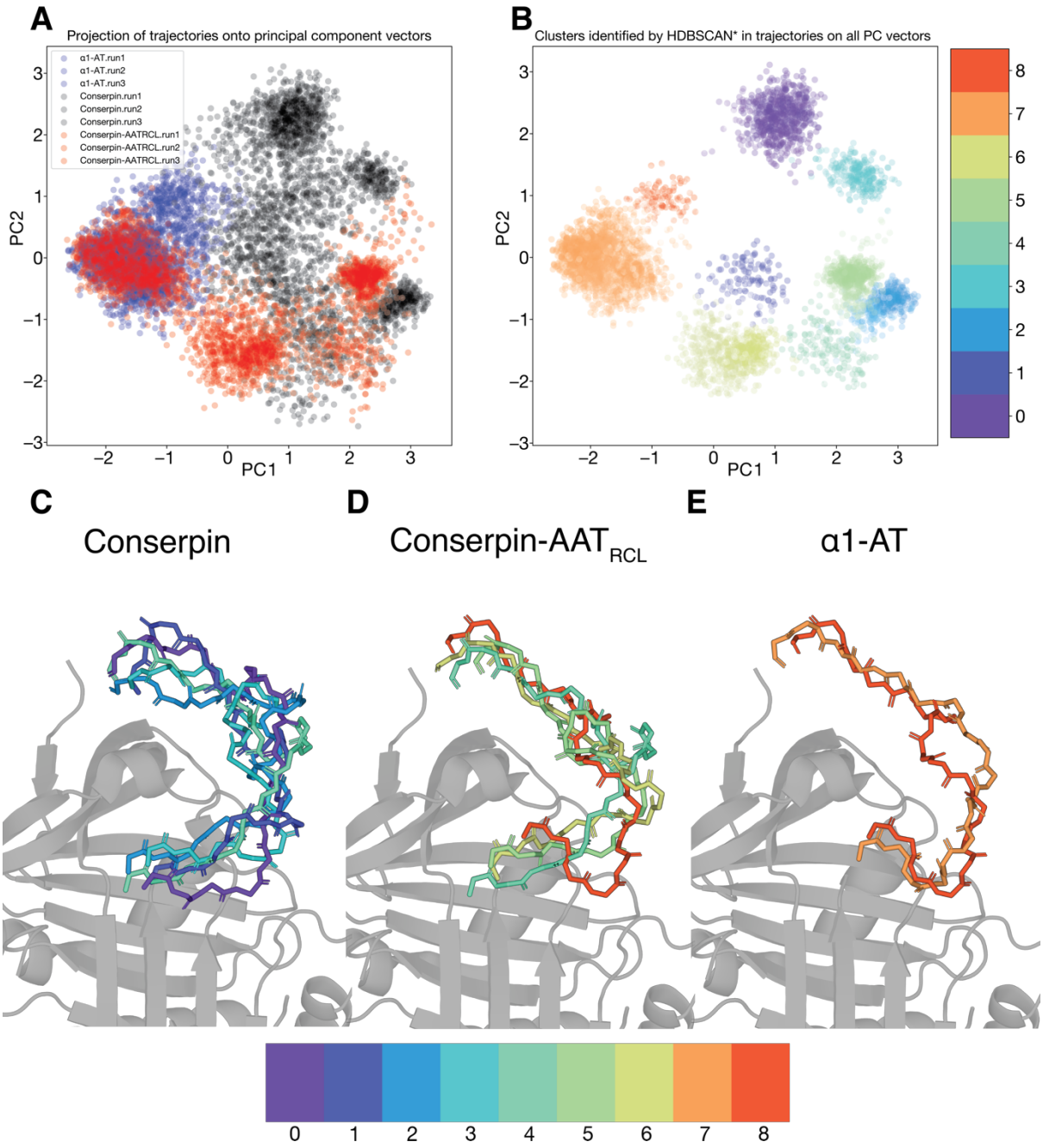


Figure 6: RCL conformational cluster determination by principal component analysis. To describe the motion of the RCL across all simulations, principal component vectors were determined for all RCL backbone conformations. (A) The trajectories of each RCL (α1-AT: blue, conserpin: black, conserpin-AAT_{RCL}: red) are projected on the first 2 PC axes, and (B) these conformations were grouped into 9 clusters. For (C) conserpin, (D) conserpin-AAT_{RCL} and (E) α1-AT, representative RCL backbone conformations for the clusters explored by each serpin over the course of the molecular dynamics simulations, are shown atop a serpin body (grey cartoon α1-AT (PDB: 3NE4)).

Supporting Information

Reactive Centre Loop Dynamics and Serpin Specificity

Emilia M. Marijanovic, James Fodor, Blake T. Riley, Benjamin T. Porebski, Mauricio G. S. Costa, Itamar Kass, David E. Hoke, Sheena McGowan, Ashley M. Buckle

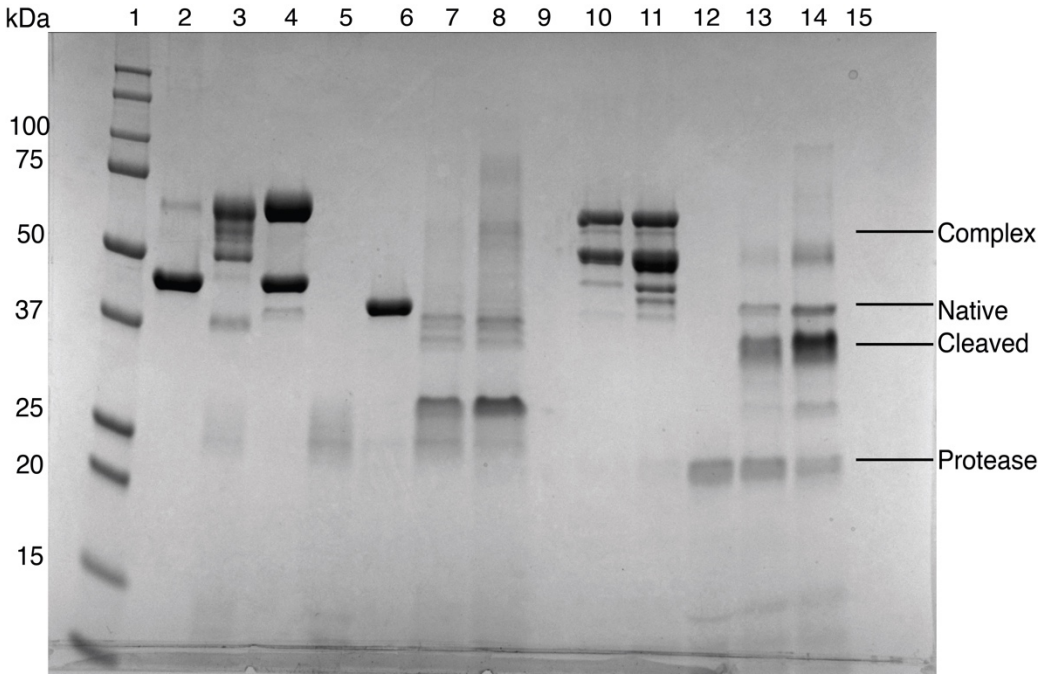
SI Tables

Table S1*: Data collection and refinement statistics.*

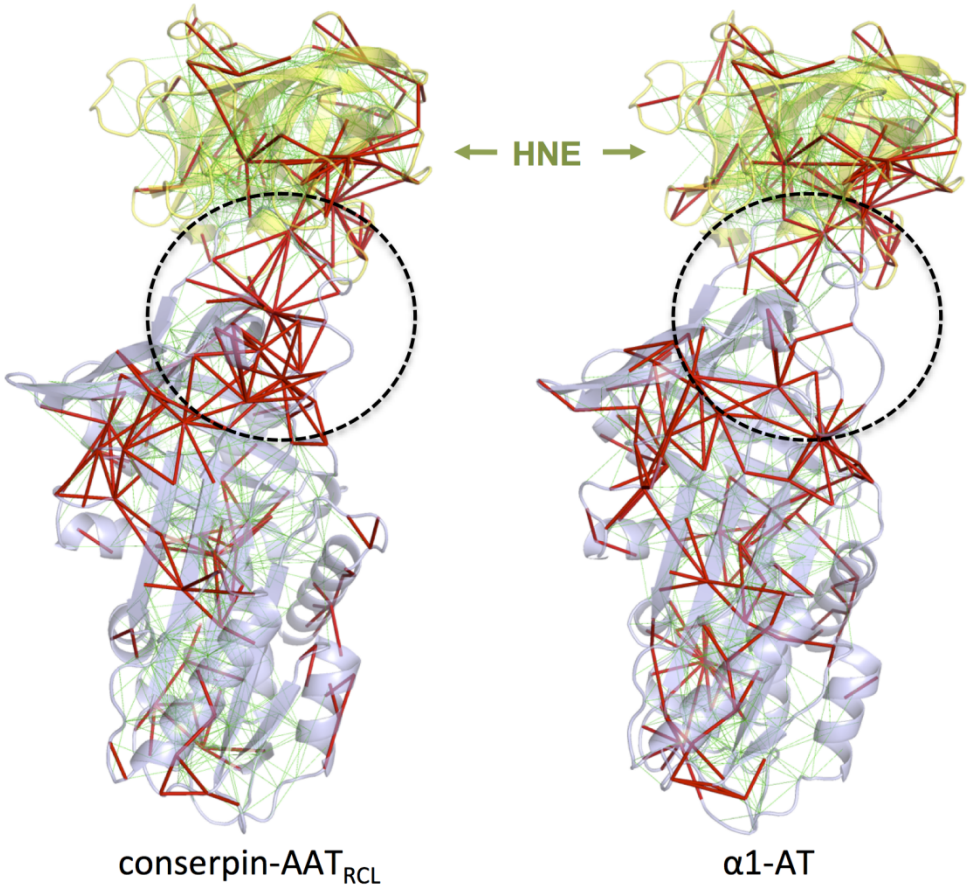
Data collection	Native conserpin-AAT_{RCL}
Wavelength (Å)	0.9537
Space group	C 2 2 2_1
Unit cell dimensions (Å)	68.06, 75.28, 150.45, 90, 90, 90
Resolution (Å)	2.48
Number of measured reflections	13376
Number of unique reflections	13455
Completeness (%)	95.47
Redundancy	29.2
R_{pim}	
$\langle I/\sigma I \rangle$	99.73
Structure refinement	
Number of reflections	13376
Number of protein atoms	2561
Number of water molecules	9
R_{work} (%)	0.1986

R_{free} (5% of data) (%)	0.2593
CC1/2	0.998
CC*	1
RMSD bond lengths (Å)	0.008
RMSD bond angles (°)	1.17
Average B-factor ($Å^2$)	82.60
Protein	82.70
Solvent	59.30
Ramachandran	
Favoured (%)	91
Outliers (%)	0.57
MolProbity score	2.32, 85th percentile* (N=6912, 2.48Å ± 0.25Å)
PDB ID	6EE5

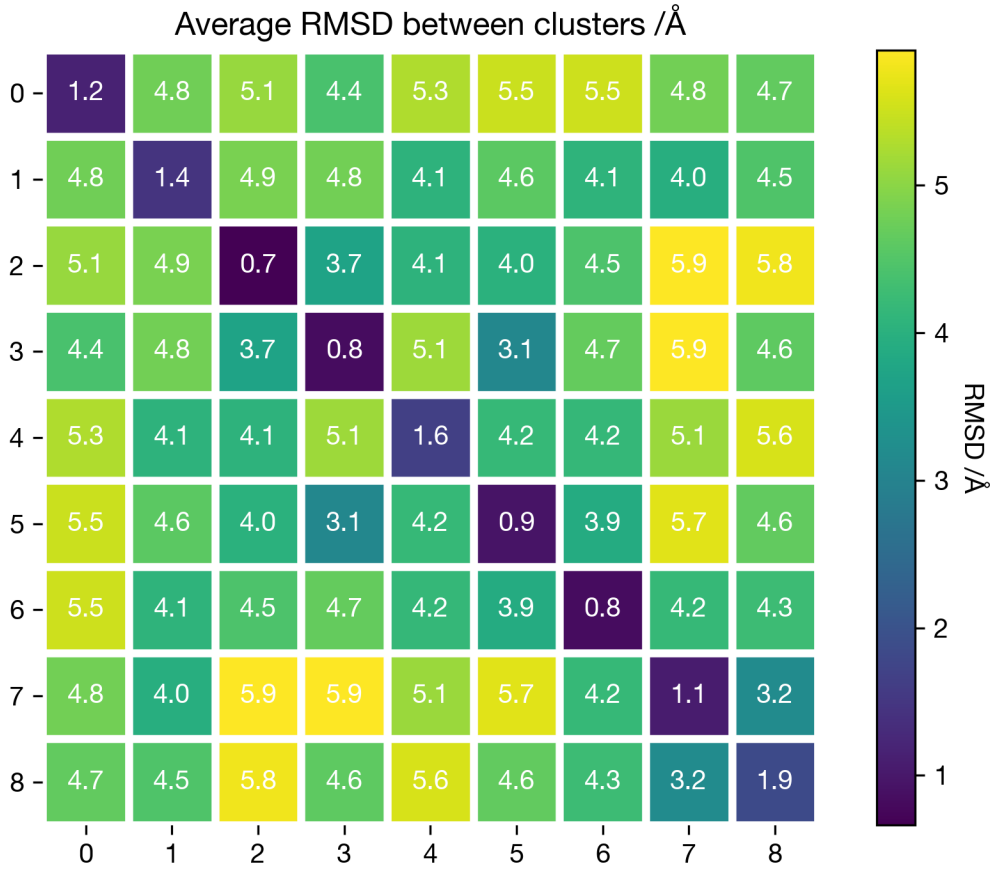
SI Figures



SI Figure 1: The full length SDS-PAGE gel, cropped in Figure 1. The gel was cropped from lane 1-8. Lanes 9-14 show the formation of a serpin: protease complex between α1-AT and conserpin-AAT_{RCL} with trypsin. From left to right: 1. Molecular weight markers (kDa); 2. α1-AT alone; 3. 1:1 ratio of α1-AT: HNE; 4. 2:1 ratio of α1-AT:HNE; 5. HNE alone; 6. conserpin-AAT_{RCL} alone; 7. 1:1 ratio of conserpin-AAT_{RCL}:HNE; 8. 2:1 ratio of conserpin-AAT_{RCL}:HNE; 9. blank; 10. 1:1 ratio of α1-AT: trypsin; 11. 2:1 ratio of α1-AT: trypsin; 12. Trypsin alone; 13. 1:1 ratio of conserpin-AAT_{RCL}:trypsin; 14. 2:1 ratio of conserpin-AAT_{RCL}:trypsin; 15. blank.

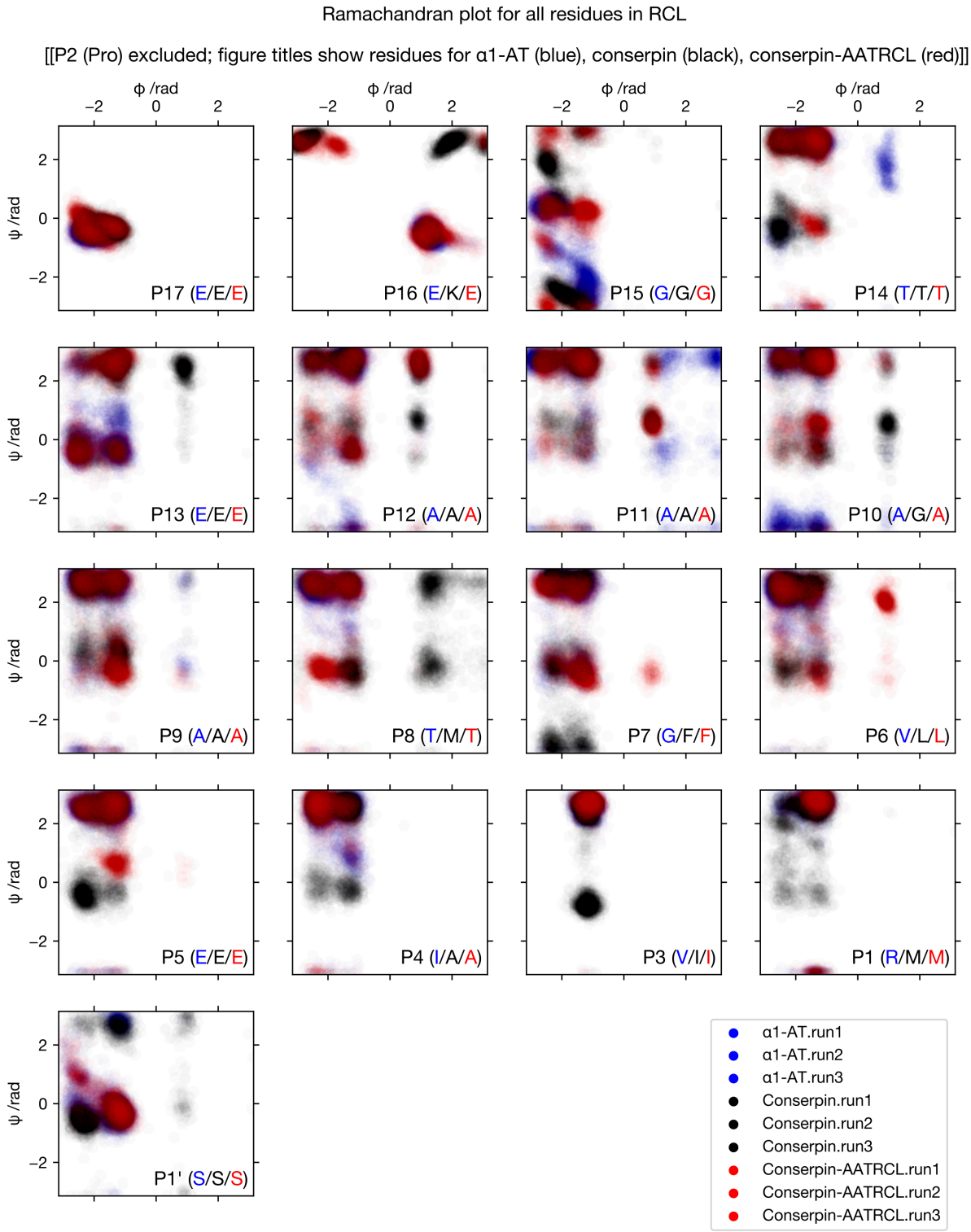


SI Figure 2: Visualization of the frustration networks mapped onto the modeled complexes between conserpin-AAT$_{RCL}$ or α1-AT with HNE (left and right, respectively). Minimally and highly frustrated contacts are represented in green and red, respectively.



SI Figure 3: Average cross-cluster RMSD heatmap.

After clustering, a set of exemplar structures was extracted from each cluster. Pairwise RMSDs were calculated for all frames within a cluster, and also all frames between clusters. These pairwise RMSDs were then averaged, and displayed here as a heatmap. The low RMSD along the main diagonal shows that the exemplar structures from each cluster are self-similar, and the large RMSD outside the main diagonal shows that each cluster is distinct to all other clusters.



SI Figure 4: Ramachandran plots show changes in dynamics of the RCL (P17-P1'): to localize conformational differences between each system on a residue-by-residue basis, phi and psi angles were calculated for all residues (excluding P2 Proline), and plotted here. The conformations explored in α1-AT simulations are coloured blue; in conserpin simulations, black; in conserpin-AAT_{RCL} simulations red. In particular, it shows that α1-AT and conserpin-AAT_{RCL} occupy a Lα torsion at P16 (bent hinge), whereas conserpin does not. contributing to the insertion of the hinge

region. Additionally, conserpin is prone to forming α-helices between P5−P3, in contrast to the other two groups.

Chapter 4: Enhancing the biophysical properties of α1-antitrypsin

4.1. Introduction

Proteins have evolved over millions of years to contain certain properties, including their function[241]. For a protein to function, it must fold into a native conformation that is stable[242]. Protein stability is often referred to as Gibbs free energy difference between the unfolded state and the folded state of a protein, between 5-15 kcal/mol[12,243]. However, there are two aspects to protein stability. Referring to a protein's Gibbs free energy is the thermodynamic stability of the protein, which is the equilibrium between the unfolded, partially folded and the native conformation of the protein[244]. The other aspect is the kinetic stability. This type of stability relates to the high free-energy levels between the different conformations along the folding pathway between the unfolded and native states, including folding intermediates, and provides information on how a protein folds[26,244]. The stability of the protein could be enhanced thermodynamically by increasing the stability of the native state, or decrease the stability of the unfolded state (pushing the protein to fold into the native conformation)[245], or kinetically, by decreasing the rate of unfolding through increasing the energy barrier that separates the native state to the folding intermediates[8,33,246].

4.1.1. Engineering protein stability

Engineering stability can occur with or without a structure of the target protein. Two non-structural methods include directed evolution and mutating conserved residues from homologous proteins. The directed evolution method mimics natural evolution in a laboratory setting. Here, random mutations are introduced onto a target proteins DNA sequence through polymerase chain reaction (PCR). The resulting mutant proteins are then screened for increased stability, with the successful mutations (containing an increase in stability) selected and sequenced to determine the new residue and position on the protein[247–251]. The successful mutations can undergo further mutagenesis. The second method involves determining the conserved residues between homologous proteins. The conserved residues are hypothesized to contribute to structure, function or stability[152]. The conserved residues identified can be introduced onto a target protein[252,253].

The structure of a target protein allows for the use of rational design to engineer stability. Through mutagenesis studies on small, reversible folding proteins, several strategies for increasing protein stability were observed, including reducing the rigidity of a proteins backbone by replacing glycine residues with other amino acids, or introducing prolines[168,254]; introducing disulphide bonds and salt bridges[38,40,41,47,255–258] or improve the packing in the hydrophobic core[29,33,259,260].

Lastly, understanding how homologous thermophilic proteins function at increased temperatures can aid in increasing the stability of the mesophilic counterparts[261]. Through studying the structural and sequence differences between homologous thermophilic and mesophilic proteins, any residues and interactions that confer stability can be identified, then introduced onto the target mesophilic protein. This method has previously been successful in increasing the stability of the target protein[262–264].

4.1.2. Protein stability/ function trade-off

Proteins must remain stable (folded) to perform their function. However, it is possible that increasing the stability could be detrimental to its function. It is hypothesized that proteins may contain a stability-function trade off, where increasing the stability of a protein results in a compromise of function, particularly when mutating the active site or binding regions of proteins[265]. This is true in some cases[262,265,266], while other proteins remain just as functional compared to wild type, when the stability increased[250,263,267]. This indicates the stability-function trade off occurs at an individual residue level for different proteins.

4.1.3. Using stability for function: Serpins

Not all proteins fold into the most thermodynamically stable conformation, instead they pause folding once reaching a particular intermediate. Serine protease inhibitors (serpins) fold into a metastable native conformation, and use this metastability for their function (also known as the native state)[56,57]. Due to this metastability, serpins are susceptible to misfolding and aggregation under various conditions, including heat[115,116] and mutations[64,119]. Therefore, increasing the stability of the metastable native state (both thermodynamically and kinetically) could prevent misfolding and aggregation, all while not affecting its inhibitory function. Here, the archetypal serpin α1-antitrypsin (α1-AT) is the protein of focus.

This metastable state provides information of the relationship between structure, stability and function. Folding into this metastable conformation is a result of over-packing side chains, unfavourable interactions, unfavourable buried polar groups and large surface pockets[66,67,69,268]. The serpins metastable native conformation consists of 3 β-sheets (A-C), 9 α-helices (A-I) and a protruding mobile loop, noted as the reactive centre loop (RCL) (Figure 1). The largest element of the serpin structure is a central β-sheet, β-sheet A, which lies behind Helix-F. The hydrophobic core is comprised of strands 2-6 of β-sheet B (s2B, s3B, s4B, s5B and s6B) and helix-B, while β-

sheet B also forms the B/C barrel with β-sheet C[58]. The protruding RCL is central to the serpins inhibitory mechanism, acting as a bait for the target protease to bind[58–60]. The loop is connected from strand 5 of β-sheet A and stand 1 of β-sheet C.

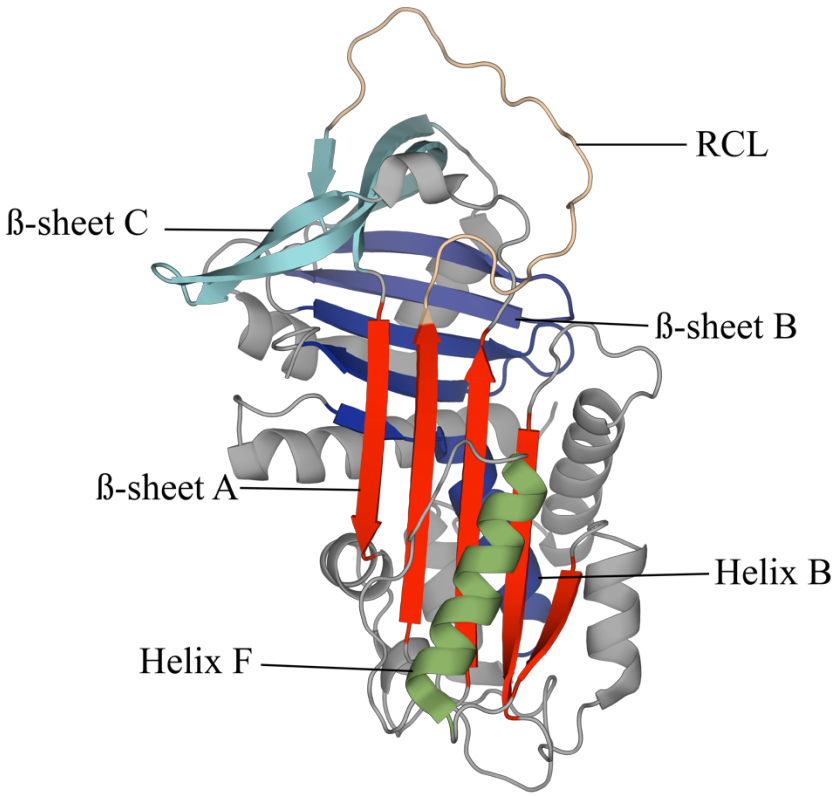


Figure 1. The x-ray crystal structure of native α1-AT (PDB ID: 3NE4[269]). α1-AT is comprised of 3 β-sheets (sheet A, *red*; B, *blue*; and C, *cyan*) and 9 α-helices (Helix-B, *blue*; *Helix-F*, green). The reactive centre loop (RCL) protrudes from the body of the serpin. Helix-B and β-sheet B form the hydrophobic core (blue), while β-sheet B and β-sheet C form the B/C barrel.

The inhibitory mechanism of a serpin is unique, which involves using the serpin's metastability to perform a large conformational change to inhibit the target protease. The inhibitory mechanism begins with the target protease binding to the P1-P2' scissile bond, forming a reversible, stoichiometric 1:1 Michaelis-Menton complex[63,88]. From here, the protease's catalytic serine cleaves the P1-P1' bond, releasing the N-terminus of the RCL in a spring-like mechanism, with the protease covalently attached. Next, the RCL inserts into β-sheet A, between strands 3 and 5 (s3A and s5A), becoming an extra strand (strand 4, s4A), while covalently attached protease translocated 71Å to the opposite pole of the serpin. This translocation process disrupts the protease's active site, rendering the protease inactive[63]. The insertion of the RCL as an extra strand on β-sheet A releases approximately 32 kcal mol^{-1} of energy[48,94,95], leaving the serpin in an inactive, but a thermodynamically lower, stable energy conformation.

Alternatively, the insertion of the RCL does not always occur with protease cleavage. Insertion without cleavage, known as the latent conformation, can occur spontaneously at physiological temperatures[101], at elevated temperatures, with specific mutations (e.g. in the B/C barrel[107]), at low pH or in the presence of low denaturant concentrations[102–104].

4.1.4. Stabilizing the native state of α1-AT without compromising function

There have been a few studies to date that have attempted to stabilize the native state of α1-AT without compromising its function. Each of these studies have used single-point mutations to observe the effect on thermodynamic stability (measuring stability through chemical equilibrium unfolding and determining the change in Gibbs free energy between the wild type and mutant, ΔΔG) and inhibitory function. Many of these mutations were not subjected to analysis to observe any change in folding pathway or folding rates. The stabilizing single-point mutations that did not compromise function were in regions that were not involved in the inhibitory mechanism, such as the hydrophobic core[66,69,211,268,270], while mutations that were in regions important for function, such as β-sheet A, decreased the inhibitory mechanism of the serpin. Furthermore, only random single point mutations were introduced, rather than potentially stabilizing cluster of mutations. Only one study combined multiple single mutations together to observe the change in stability. This constructed α1-AT, known as the Multi7, contains mutations mostly in the hydrophobic core (Phe-57Ala, Thr-68Ala, Ala-70Gly, Met-374Ile, Ser-381Ala, Lys-387Arg)[211] (Figure 2). This Multi7 α1-AT has a thermodynamic stability much higher than WT α1-AT, almost as stable as non-inhibitory serpin ovalbumin, while remaining just as active as WT. There is no statement of the midpoint of thermal denaturation on the Multi7 serpin. Furthermore, the Multi7 variant was only successful in increasing its thermodynamic stability as many of the mutations are present in the hydrophobic core. It was previously stated that mutations in functional areas increased stability while compromising its function[213], yet these mutations were introduced randomly. To overcome this, rational design should be employed, using a thermostable, active serpin as a reference.

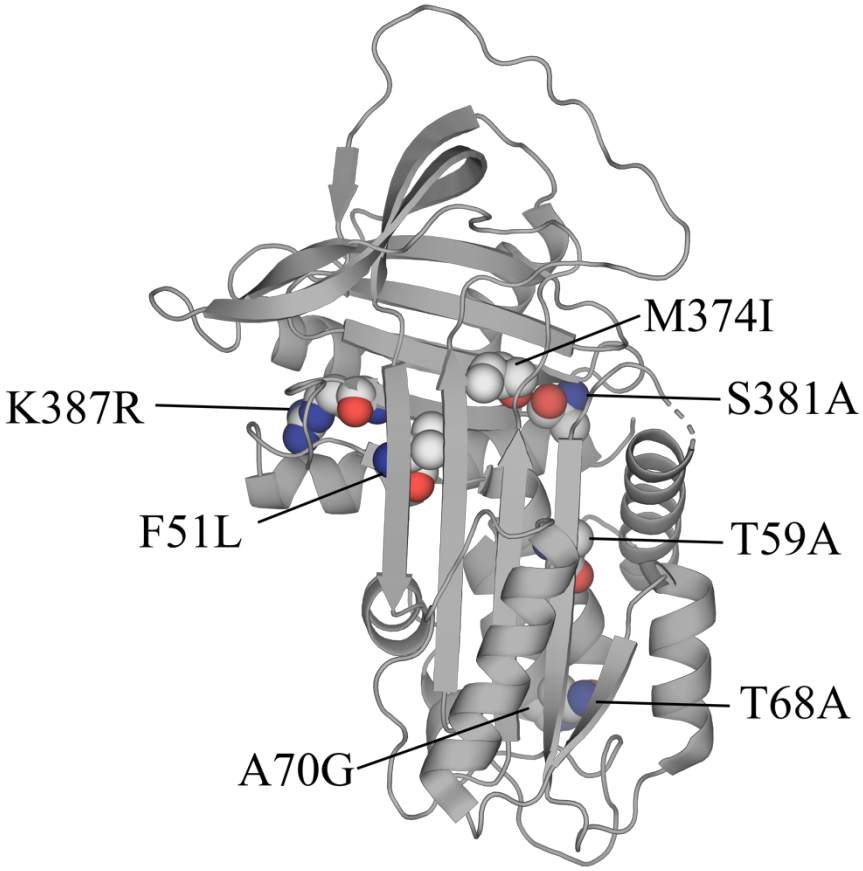


Figure 2. The x-ray crystal structure of highly stable Multi7 α1-AT (PDB ID: 5NBV, *unpublished).* Many of the stabilizing mutations (spheres) are present in the hydrophobic core of the serpin.

4.1.5. Engineering stability of α1-AT using consensus-designed serpin, conserpin

It is possible to increase the stability of mesophilic proteins by studying their homologous thermophilic counterparts[262–264]. Thermostable proteins have evolved to withstand high temperatures, yet perform the same function as its mesophilic counterpart[152,261]. Therefore, observing the structural features in a thermostable protein can aid in increasing the stability of a mesophilic protein. Previously, consensus engineering was employed to produce a consensus-designed serpin, termed conserpin[150]. Biophysical analysis of conserpin revealed it was extremely thermostable (with a midpoint of thermal unfolding, *Tm* above 100°C), folds through a poorly populated intermediate (populated over a small timeframe) with a Gibbs free energy ($\Delta G_{Denatured-Native}$) of -23.2 kcal mol^{-1}, while being an active inhibitor against the serine protease trypsin. The determination of the x-ray crystal structure of native state conserpin highlights why it is thermostable. Despite having less salt bridges and hydrogen bonds compared to other thermophilic

serpins, conserpin contains several regions throughout its structure that have improved packing of the residue side chains throughout the structure, additional salt bridges in key regions of the serpin structure and favourable hydrophobic packing. Each of these features is hypothesized to contribute to the overall stability of conserpin[150].

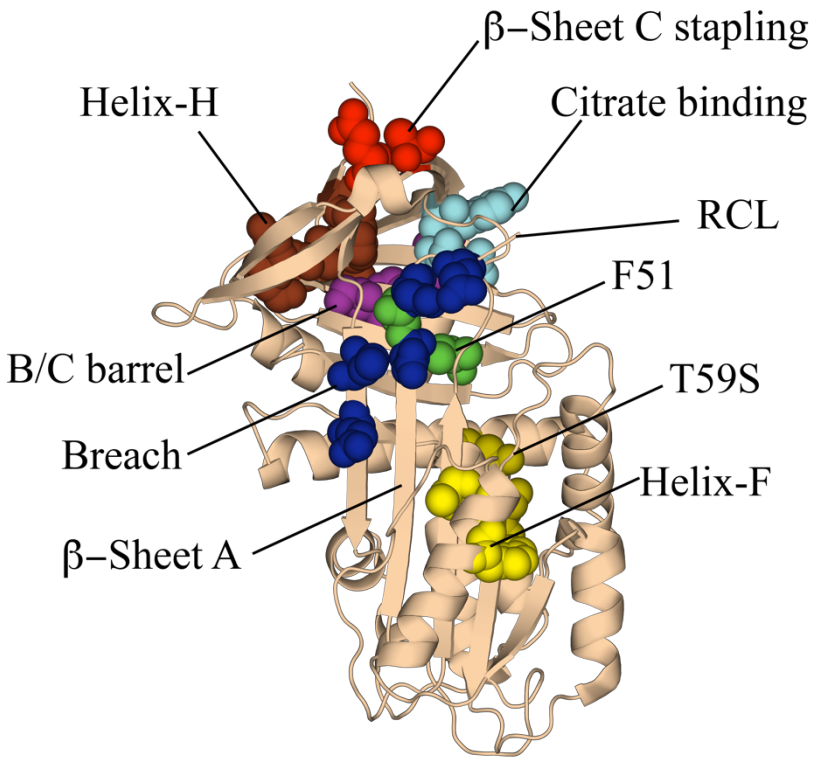


Figure 3. The x-ray crystal structure of native conserpin (PDB ID: 5CDX[150]). Each different colour of sphere represents a region that is hypothesized to contribute to conserpin's ideal biophysical properties

4.1.6. Aim of study

The structural determination of conserpin in the native state highlights several regions of the synthetic serpin that are hypothesized to contribute to its thermostability. In this study, the hypothesized stabilising regions were grafted onto α1-AT in an attempt to increase the thermostability of α1-AT's native state without compromising its function. Some of these stabilizing regions are present in regions of the serpin that are important for function. The other regions are spread throughout the serpin molecule, such as the B/C barrel and a couple of helices. Therefore, we hypothesize that rational design will be superior in engineering the thermostability of α1-AT without compromise of function, rather than random mutations.

4.2. Results

4.2.1. Designing α1-AT grafts

Throughout the conserpin molecule, there are various clusters of amino acids that are hypothesized to contribute to thermostability[150]. These clusters were identified through determining the x-ray crystallography structure and contribute to the thermostability through improved packing in the hydrophobic core, the addition of favourable interactions throughout various regions or increased number of salt bridges. To improve the biophysical properties of α1-AT, each amino acid cluster was grafted onto α1-AT. All clusters were modelled to ensure all incoming amino acids fit and did not produce any unfavourable interactions. A total of eight grafts were produced, with a varying number of mutated amino acids; from two to six residues (Figure 4 a and b, Table 1). Below is information of each graft, clustered based on the position on the serpin:

Hydrophobic core:

- T59S: introduction of three polar, uncharged side chains to create favourable polar and non-polar interactions. Present on Helix-A and Helix-B
- F51: contrary to the name, phenylalanine 51 was not mutated. Present on β-sheet A facing towards the hydrophobic core, three residues surrounding residue 51 were mutated. Two involved a decrease in side chains, while 1 has an increase in size chain size. This improves the packing against β-sheet A.

B/C Barrel:

- B/C barrel: Removal of two charged residues on Helix-H, facing into the B/C barrel, and a decrease in a side chain on β-sheet B facing towards Helix-H. This is to improve the hydrophobic packing within this region.
- Helix-H: An addition of six residues between Helix- H and β-sheet B, four of which include the addition of a charged residue. This creates a coordinated salt bridge network to stabilise the native structure. The remaining two residues involve either an increase (valine to phenylalanine) or decrease (phenylalanine to alanine) in side chain size. A combination of all mutations improves the residue packing, providing further stabilization (along with the salt bridge network).
- Citrate-binding: Two mutations were introduced in the B/C barrel, one on β-sheet B and another on β-sheet C. These mutations are in the region where citrate binds and stabilizes the native state of α1-AT through binding into a surface pocket. Introducing the two

mutations, both of which contain a larger side chain, fill this surface pocket, stabilizing the native state.

- β-sheet C stapling: The addition of two charged residues on strands 2 and 3 of β-sheet C (s2C and s3C) creates a salt bridge, rigidifying and stabilizing β-sheet C.

Functional regions:

- Breach: based at the top of β-sheet A and base of the RCL, five mutations were introduced to increase the salt bridge network already present. The additional salt bridges will prevent β-sheet A from opening up as easily as in WT α1-AT, stabilizing the native conformation.
- Helix-F: Laying in front of β-sheet A, three mutations introduced to improve the packing between Helix-F and β-sheet A, through the increase of 1 side chain and a decrease of the other two. This will rigidify the helix, possibly preventing easy shifting when β-sheet A opens.

Graft	**Mutations**
Breach	S292K, T294E, T339E, D341N, K343E
B/C barrel	F253I, F275W, E279L
β-sheet *C stapling*	L224K, S285E
Citrate-binding	N228Y, L241E
F51	L291F, 1340V, M374I
Helix-F	G115A, Y160W, Y187A
Helix-H	L232D, K234E, V364F, F366A, N367D
T59S	L30N, A58S, T59S

Table 1. A list of all the mutations introduced for each graft

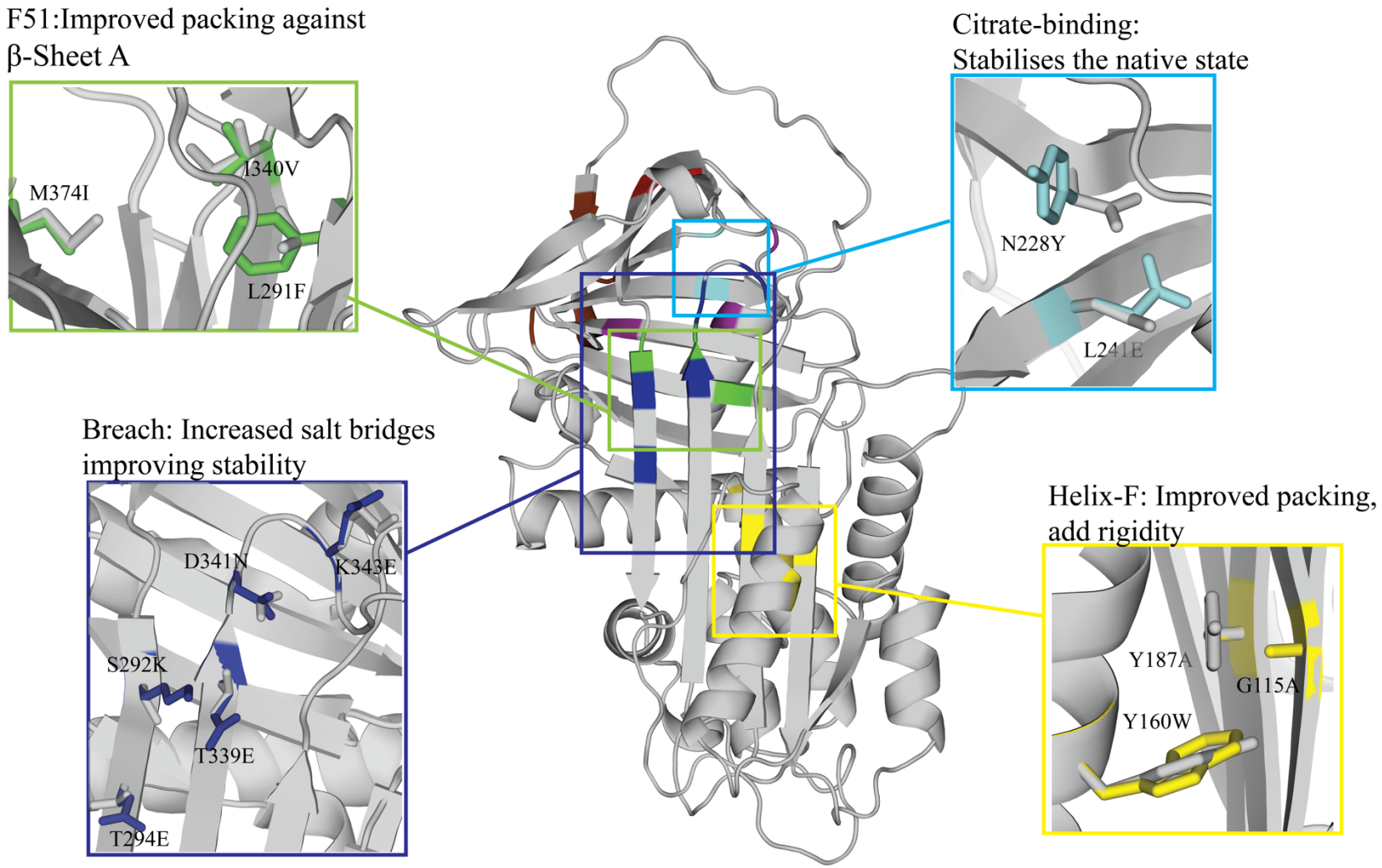


Figure 4a. Front view of which regions of conserpin were grafted onto α1-AT (PDB ID: 3NE4[269]). Each region grafted from conserpin onto α1-AT is hypothesized to contribute to conserpin's ideal biophysical properties through various favourable interactions (*grey:* WT α1-AT, *colour*: introduced mutations).

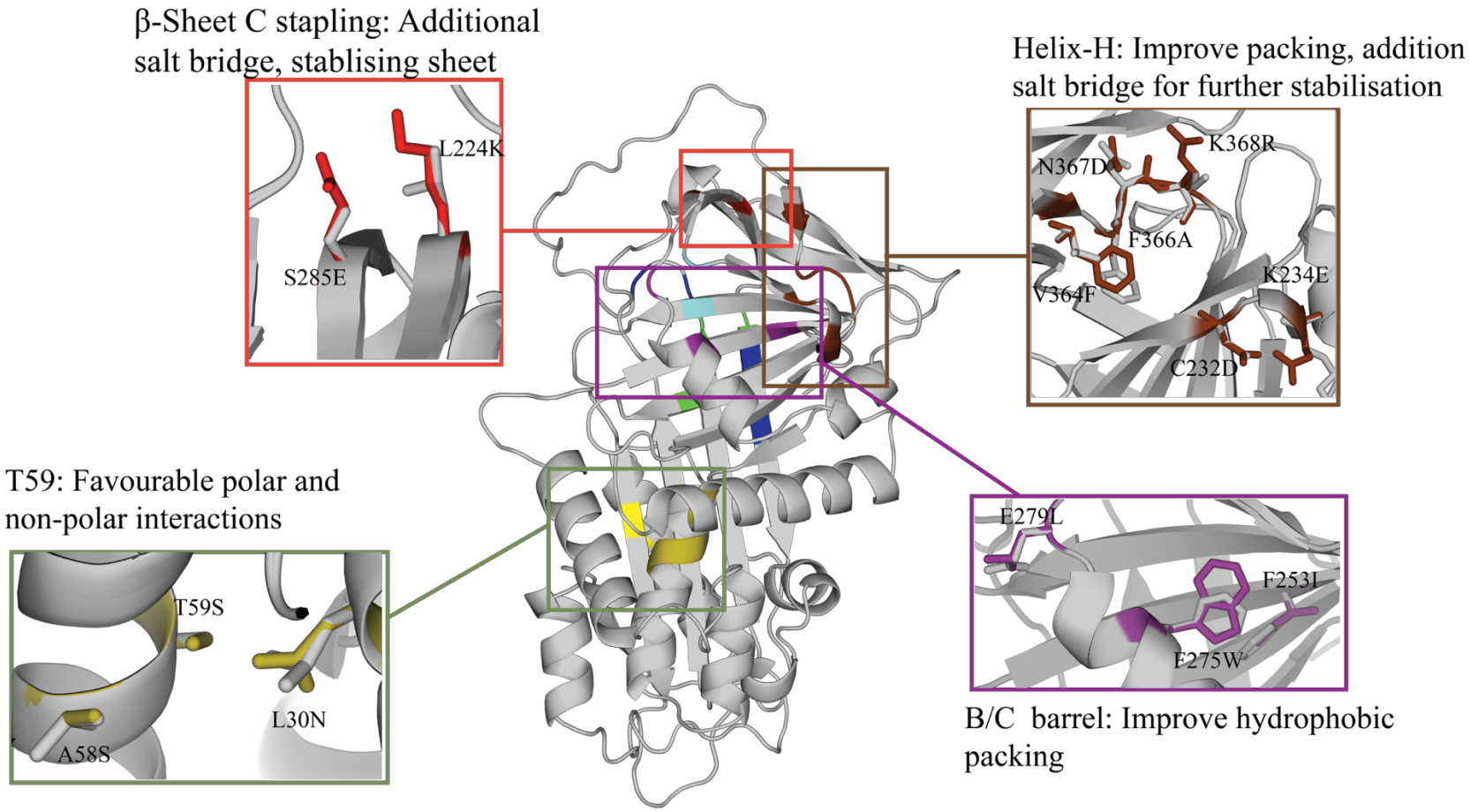


Figure 4b. Back view of which regions of conserpin were grafted onto α1-AT (PDB ID: 3NE4[269]). Each region grafted from conserpin onto α1-AT is hypothesized to contribute to conserpin's ideal biophysical properties through various favourable interactions (*grey:* WT α1-AT, *colour*: introduced mutations)

4.2.2. All α1-AT are insoluble in recombinant expression and require refolding

α1-AT is expressed in both the soluble and the insoluble fraction (inclusion bodies) during recombinant protein expression[162]. All eight grafts were synthesized and cloned into the expression vector that expresses α1-AT in the soluble fraction[91,271], with an affinity tag to aid in purification. Despite WT α1-AT expresses in the cytosol (soluble fraction), all eight grafts express in the insoluble fraction (inclusion bodies), while minute amounts of most grafts can be detected in the soluble fraction (as detected by western blot). WT α1-AT's refold protocol is well established; however, fine-tuning was required to obtain enough monomeric protein of each graft. Out of the eight grafts, six refolded into the active, native conformation, while two refolded into an inactive conformation, possibly the latent conformation (as determined by lack of inhibitory function) (Table 2)

Graft	**Insoluble**	**Soluble protein detected?**	**Refold conformation**
WT α1-AT	No	Yes	-
Breach	Yes	Yes	Native
β-sheet C stapling	Yes	No	Latent
B/C barrel	Yes	Yes	Native
Citrate-binding	Yes	Yes	Native
F51	Yes	Yes	Native
Helix-F	Yes	Yes	Native
Helix-H	Yes	No	Latent
T59S	Yes	No	Native

Table 2. Each of the eight grafts expressed a majority of the protein in the insoluble fraction during recombinant expression. All but two grafts were successfully refolded into the native, active conformation.

The use of ion exchange chromatography allows for the separation of not only monomeric serpin from a polymer/ aggregate, but also the different conformations of monomeric serpin, the native and latent conformations[162]. The increase of the salt gradient separates and elutes the different conformations (Figure 5A). The two grafts that refolded into an inactive conformation eluted at a slightly higher salt gradient compared to WT α1-AT (Figure 5A, *1, 2*), but lower than aggregate (Figure 5A, *3*). This inactive conformation was also confirmed by an assay against trypsin, with no inhibition of trypsin, even at a 2:1 serpin: protease ratio (Figure 5B). The two grafts that fold into the inactive conformation, β-sheet C stapling and Helix-H, both have the introduction of salt bridges. The new salt bridges seems to have affected the folding pathway of both grafts, reducing the energy barrier that separates the native and inactive conformations. Therefore, both grafts bypass halting the folding in the native conformation, and instead fold into an inactive, possibly latent, conformation.

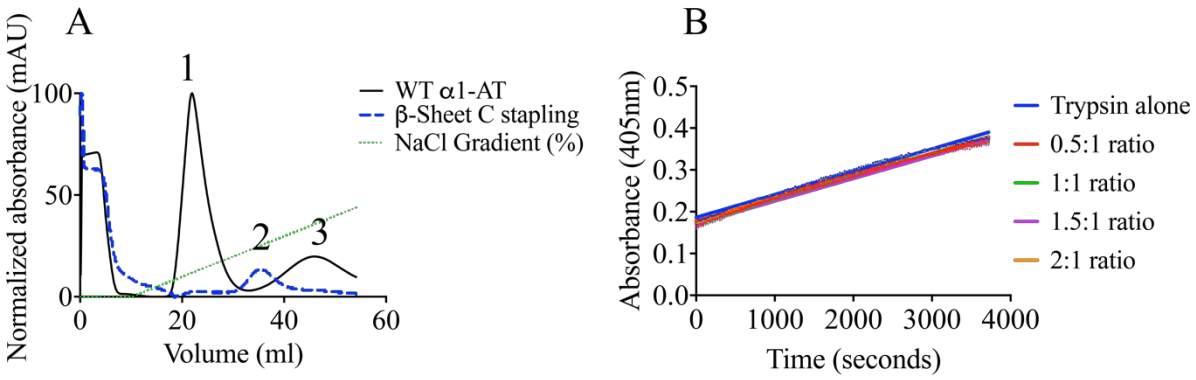


Figure 5. Confirming latency for two grafts. A. Ion exchange allows for the separation of different serpin conformations. Native protein elutes at a low salt gradient (1), followed by latent (2), and aggregate elutes at a high gradient (3). The chromatogram shows the elution profile of native WT α1-AT and the β-sheet C stapling graft, where the β-sheet C stapling graft elutes as an inactive, possibly latent serpin. **B**. Inactivity for both grafts was confirmed by an assay against the serine protease trypsin (Helix-H graft shown).

4.2.3. Three grafts improved α1-AT's thermostability

The aim for grafting regions of conserpin onto α1-AT was to increase the thermostability of the native state, while not affecting the serpins function. Conserpin is extremely thermostable, with a midpoint of thermal unfolding (*Tm*) above 100°C[150], while native α1-AT is less stable, with a *Tm* of 60°C. To pinpoint the regions of conserpin that are important in its

thermostability, each of the 8 grafts, including the latent grafts, underwent circular dichroism thermal melts (natively folded grafts: Table3, latent folded grafts: Table 4). Each sample was heated from 25-95°C and cooled back down to 25°C. Out of the six natively folded grafts, three grafts improved the thermostability of α1-AT (Breach and Helix-F grafts) (Table 3, Chapter 4 supporting information (SI) Figure 1a. b.). The F51 graft undergoes two unfolding transitions, an initial unfolding transition at 65°C, and a second transition that continues until 95°C (Chapter 4 SI Figure 1b). Compared to WT α1-AT, 3 were slightly less thermostable than WT (B/C Barrel, Citrate-binding and T59S grafts). Upon cooling from 95°C, WT α1-AT and all the grafts do not refold, instead each precipitate out of solution.

Graft	***Tm* (°C)**	**Δ*Tm* (°C)**	**$\Delta G^{(unfold)}$ (kcal/mol)**	**$\Delta\Delta G^{(unfold)}$ kcal/mol)**
WT α1-AT	60 ± 2.26	-	8.79 ± 0.85	-
Breach	67.38 ± 0.34	7.38 ± 1.92	10.33 ± 0.279	-1.54 ± 0.58
B/C Barrel	57.51 ± 0.39	-2.49 ± 1.86	6.27 ± 0.26	2.52 ± 0.59
Citrate	58.11 ± 0.17	-1.88 ± 2.1	4.672 ± 0.98	4.12 ± 0.12
Helix-F	62.08 ± 0.14	2.08 ± 2.11	12.86 ± 0.94	-4.06 ± 0.8
F51	65.48 ± 0.766	5.49 ± 1.49	-	-
T59S	56.96 ± 1.45	-3.03 ± 0.8	8.007 ± 0.415	0.719 ± 0.444
Conserpin	100+	40+	-	-

Table 3. The thermodynamics of each graft in comparison to WT α1-AT. Using the thermal denaturation curves to calculate the stability, 3 grafts are more stable than WT α1-AT, while 3 are less stable. Conserpin was added as a reference.

An increase in thermostability is correlated with an increase in thermodynamic stability[272]. The thermodynamic stability of each graft was estimated using the thermal denaturation curves and van't Hoff analysis, and compared to WT α1-AT[11,166,168,273]. The thermodynamics of each graft is suggestive, providing a rough estimate of how stable each graft is in comparison to WT α1-AT. Using the thermal denaturation curves, the Breach and Helix-F are suggested to be more thermodynamically stable than WT (-1.54 ± 0.58 and -4.06 ± 0.8, respectively). Due to the

incomplete folding transition, the thermodynamic stability of the F51 graft could not be calculated (Chapter 4 SI Figure 1B). The stability of the remaining three grafts (B/C barrel, Citrate-binding and T59S) is less than that of WT α1-AT, with the citrate-binding graft being the least thermodynamically stable than WT α1-AT.

The latent conformation of serpins is one the of the lowest energy conformations, that is, it is much more stable than the native state[101]. The thermostability of latent WT α1-AT is above 100°C (Table 4, Chapter 4 SI Figure 2). However, the two grafts that fold into an inactive conformation are not as thermostable as latent WT α1-AT (Table 4, Chapter 4 SI Figure 2). Both of these grafts contain mutations in the B/C barrel. The introduction of a salt bridge on β-sheet C (β-sheet C stapling graft) severely compromised the proteins structure and thermostability, with a *Tm* less than native α1-AT (52.97°C). The spectral scan of the β-sheet C stapling graft depicts a large β-sheet conformation, possibly a misfolded, aggregate precursor. This could explain the low *Tm* compared to the two other inactive serpins. The spectral scan of the Helix-H graft is similar to latent WT α1-AT, suggesting the Helix-H graft may fold into the latent conformation. The thermostability of the Helix-H is higher than the β-sheet C stapling graft, with incomplete unfolding transition occurring between 80-100°C (Chapter 4 SI Figure 2), further suggesting the Helix-H graft may fold into the latent conformation, rather than a misfolded conformation (as with the β-sheet C stapling graft). To conclude, the introduction of a salt bridge into the B/C barrel compromised the folding pathway of α1-AT, possibly bypassing the folding into the native conformation, and instead into inactive, possibly misfolded or latent conformations. Lastly, the introduction of a salt bridge in β-sheet C produces a misfolded serpin that is a precursor for polymerization.

Graft	**Midpoint thermal unfolding, *Tm* (°C)**
WT	100+
β-sheet C Stapling	52.97 ± 0.26
Helix-H	80+

Table 4. The midpoint of thermal unfolding (*Tm*) for the latent-folding grafts. Latent WT α1-AT does not unfold upon heating to 100°C, while the two inactive-folding grafts both unfold below 90°C.

As the desired properties of these grafts is an increase in thermostability, the three grafts that were more thermostable than WT were subjected to further investigation.

4.2.4. Increasing the thermostability did not significantly compromise function

The mechanism by which serpins inhibit proteases is unique, and the function can be compromised when the stability has increased above 13 kcal mol^{-1} [212]. To ensure increasing the thermostability did not compromise its function, the inhibitory activity of the three thermostable grafts (Breach, Helix-F and F51) was tested against α1-AT's target protease human neutrophil elastase (HNE) through a stoichiometry of inhibition (SI) assay. The stoichiometry of inhibition calculates the number moles of serpin it takes to inhibit one mole of protease. α1-AT has a stoichiometry of inhibition (SI) of 1:1 against HNE[135]. The stoichiometry of inhibition of the Breach and F51 grafts are identical to that of WT α1-AT, while the Helix-F has an increased SI (Table 5, Chapter 4 SI Figure 3).

Graft	**SI against HNE (n=3)**
WT α1-AT	1.0 ± 0.1[135]
Breach	1.09 ± 0.07
F51	1.11 ± 0.05
Helix-F	1.35 ± 0.16

Table 5. The stoichiometry of inhibition of the three thermostable grafts. All three grafts did not significantly affect the inhibitory function, compared to WT α1-AT.

4.2.5. The thermostable grafts yield more monomeric protein post-refold

α1-AT folds through at least one aggregation prone intermediate, which also increases the propensity to aggregate upon refolding[274]. As it is difficult to calculate and compare the amount of monomeric protein refolded during protein purification (due to inconsistent weight of bacterial cells), the most accurate way to determine the yield of monomeric serpin post-refold is by starting with a known protein concentration. Each protein was unfolded in 5M guanidine hydrochloride (GndHCl) for 2 hours, followed by refolding by dilution (1:10 dilution) until the

final protein concentration was 2μM. WT α1-AT largely aggregates upon refolding, yielding 11% of monomeric protein. Conserpin, however, does not aggregate and completely refolds[150] (Figure 6). The yield of each of the three thermostable grafts was higher than that of WT α1-AT (Figure 6). The F51 graft yields the most monomeric protein, followed by the Breach and lastly the Helix-F graft.

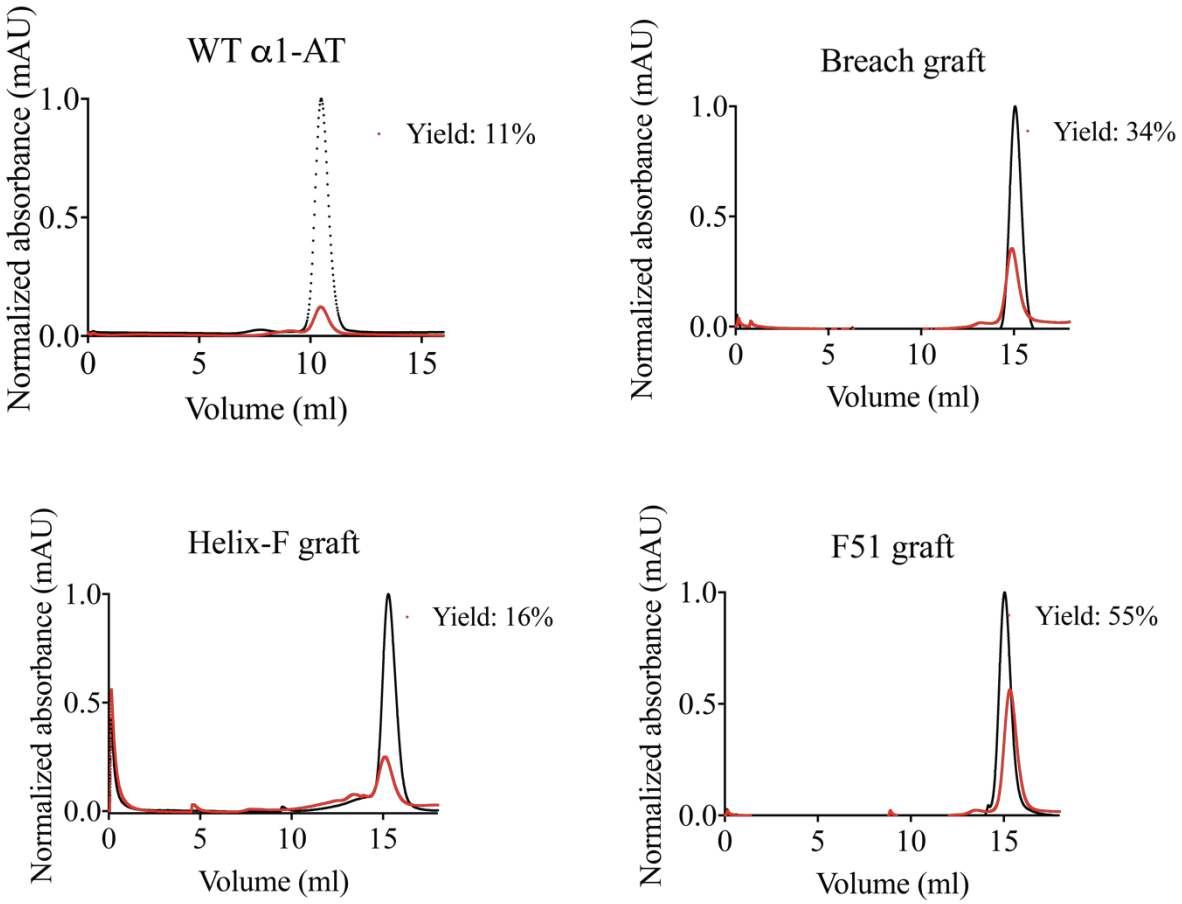


Figure 6. Analysing the yield post-refold of the 3 thermostable grafts. Each of the 3 thermal stable grafts yield higher amounts of monomeric protein post refold compared to WT α1-AT. WT α1-AT was previously performed[150]

4.2.6. The folding intermediate is populated at a higher denaturant concentration

α1-AT folds through at least one aggregation-prone folding intermediate. Introducing mutations onto the serpin could possibly affect the folding intermediate, thus affecting the folding pathway. An extrinsic fluorescent dye, Bis-ANS, binds to the hydrophobic regions of the protein and allows the detection of the folding intermediates. α1-AT's folding intermediate is populated over a wide denaturant concentration, while conserpin's intermediate is slightly

populated at a high denaturant concentration (i.e. the fluorescence is not as large as WT α1-AT) [150]. The three thermostable grafts also have an intermediate that is populated over a wide concentration and have a fluorescent profile similar to WT α1-AT (Figure 7). The difference, however, is the concentration of denaturant in which the fluorescence increases. For each of the three grafts, the maximum fluorescent intensity occurs at a higher denaturant concentration than WT α1-AT (Breach: 1.25M, F51: 1M, and Helix-F: 1.5M GndHCl concentration, compared to 0.645M of α1-AT). Therefore, each graft requires a slightly higher chemical denaturation concentration to promote unfolding.

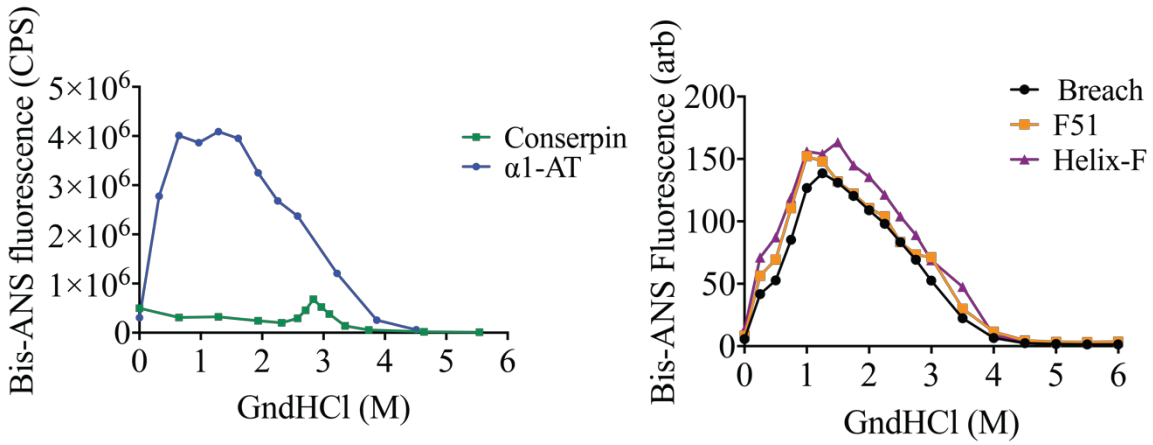


Figure 7. Extrinsic fluorescence observing the folding intermediate of each serpin. The introduction of the grafts did not affect the folding intermediate, as detected by extrinsic fluorescence. α1-AT and conserpin fluorescence profile was performed previously (a different fluorometer was used to obtain this data)[150].

4.2.7. Combination of all three thermostable grafts increases the thermostability of α1-AT

Grafting different amino acid clusters from conserpin onto α1-AT produced three grafts that contained an increase in thermostability and monomeric protein post-refold, all while not affecting the inhibitory function against HNE (only Helix-F graft has a slightly higher SI). Therefore, we hypothesized that combining these three grafts would produce an additive thermostability effect. The combined graft, dubbed "3stable" was designed, synthesized and cloned in the same way as the single grafts. Recombinant expression of the 3stable graft produced a majority of the protein in the insoluble fraction (as with all the previous grafts),

with a small amount remaining soluble. The 3stable graft successfully refolded into the active, native conformation.

The addition of the three thermostable grafts onto 3stable produced an additive midpoint of thermal denaturation, as determined by circular dichroism thermal melts (25 to 95°C, followed by cooling). Theoretically, adding the difference in *Tm* between WT and each of the three thermostable grafts would give a *Tm* of 74°C, 14°C increase from WT. The 3stable graft indeed contained an additive *Tm*, with an unfolding midpoint of 73.4°C (Figure 8B). There have been many attempts to increase the thermal stability of α1-AT, but the 3stable has the highest thermostability (as calculated by circular dichroism thermal denaturation curves) of an engineered native α1-AT to date. As with the other grafts, 3stable did not reversibly refold upon cooling from 95°C, instead, produced precipitate similarly to WT α1-AT and all the single grafts (Figure 8A).

The thermodynamic stability was estimated using the van't Hoff analysis in an identical manner to above. The thermodynamic stability of 3stable graft was 8.66kcal mol^{-1} above WT (Table 6). This is slightly above the Multi7 α1-AT, which has a thermodynamic stability of 8 kcal mol^{-1} [211].

Graft	***Tm* (°C)**	**Δ*Tm* (°C)**	**$\Delta G^{(unfold)}$ (kcal/mol)**	**$\Delta\Delta G^{(unfold)}$ (kcal/mol)**
WT α1-AT	60 ± 2.26	-	8.79 ± 0.85	-
3stable	73.4 ± 0.9	13.38 ± 1.36	17.46 ± 3.27	-8.66 ± 2.41

Table 6. The thermodynamic stability analysis of the 3stable graft. The 3stable graft has an increased midpoint of thermal denaturation (*Tm*) and Gibbs free energy (ΔG) compared to WT α1-AT.

As the stability of the 3stable graft was increased, there is a possibility of a stability-function trade-off. To ensure the 3stable graft remained as an active inhibitor against HNE, the SI was calculated, with 3stable inhibiting HNE with an SI of 1.28 ± 0.1 (Figure 8C). This SI is slightly higher than that of WT α1-AT, Breach and F51, but not to a large extent.

Each of the thermal stable grafts have an increase in yield post-refold. Under the same conditions, the 3stable graft produced a yield of 63%, confirming the hypothesis (Figure 8D). The folding intermediate was not affected, with the intermediate present over a wide concentration. The largest difference in studying the folding intermediate is the denaturant concentration of the peak fluorescence (peak fluorescence at approximately 2M GndHCl), compared to WT and each thermostable graft (Figure 8E). Therefore, a higher denaturant concentration is required to unfold the 3stable graft without affecting the intermediate.

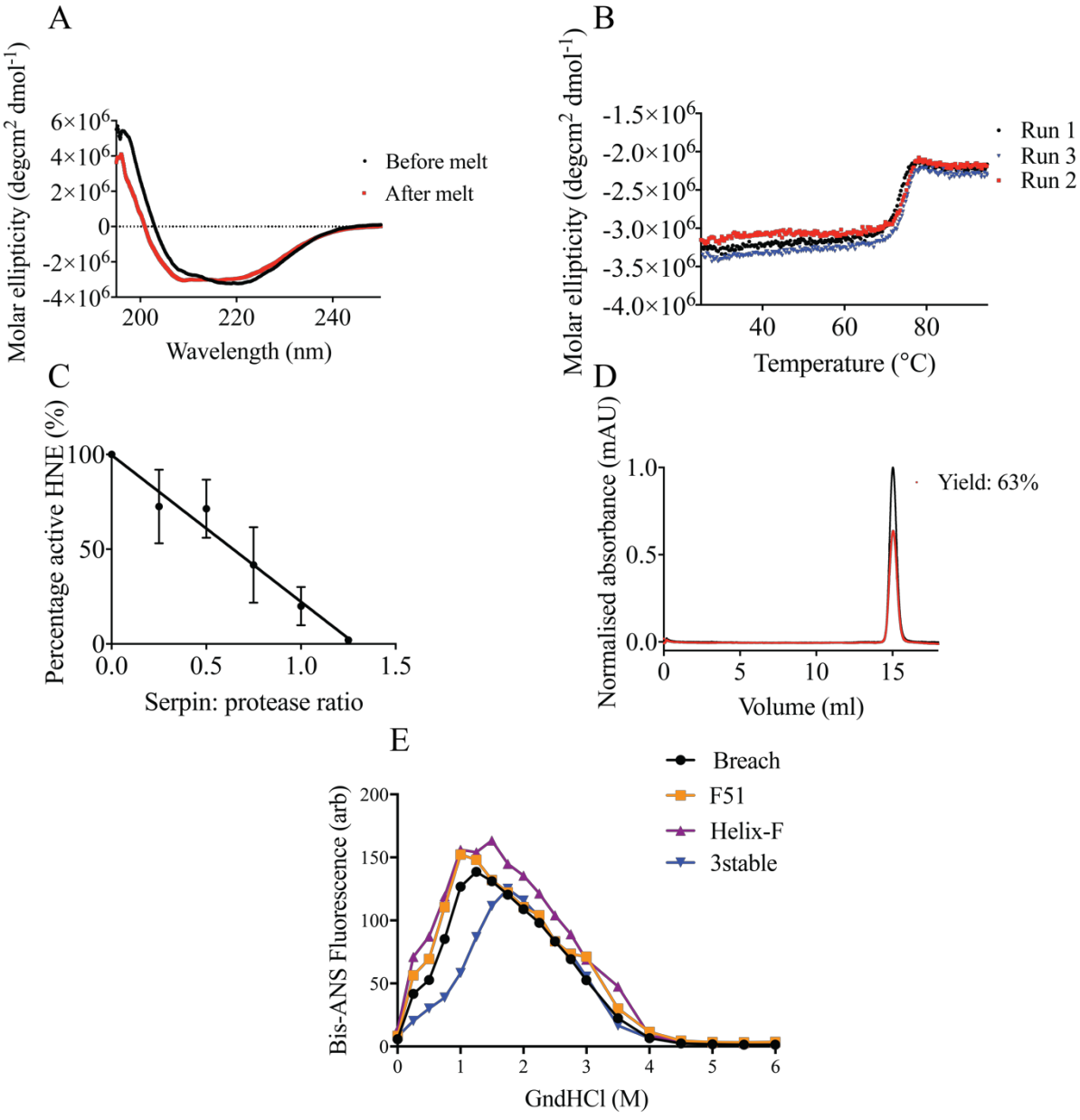


Figure 8. Biophysical analysis of the 3stable graft. A. Spectral scans before and after thermal denaturation indicate structural changes post melt. **B.** Thermal denaturation curve of each triplicate, producing a *Tm* of 73.4°C. **C**. Inhibitory activity against HNE indicates the 3stable graft is active against HNE. **D.** the yield of monomeric protein is higher than WT α1-AT and each of the individual thermal stable grafts. **E.** Extrinsic fluorescence show a folding intermediate becoming populated at a higher denaturation concentration for the 3stable graft.

4.3. Discussion

Proteins must fold into their native conformation, and remain stable in this conformation to function[275]. Serpins are one of the protein families that fold into a metastable, native conformation, that is, it is not the lowest energy conformation. The metastability is critical for the serpin's inhibitory function, which involves undergoing a large conformational change to inhibit its target protease[63]. During inhibition, a large amount of energy is released (32 kcal mol^{-1} of energy), which results in the serpin transitioning into its lowest energy, yet inactive, cleaved conformation[48,94,95]. The metastable nature of the serpin's native conformation renders it susceptible to misfolding and polymerization under various conditions (e.g. mutations and increased heat).

Previously, increasing the stability of archetypal serpin α1-AT has involved random single-point mutagenesis[66,69,212]. Many of the stabilizing mutations are present in the hydrophobic core, with no effect on α1-AT's inhibitory mechanism. However, there have been no studies to date of mutating a cluster of residues within the serpin. By mutating a cluster of residues, one destabilizing mutation could be compensated by surrounding mutated residues.

In this study, the stability of α1-AT's metastable state was engineered through studying the interactions of a consensus-designed, thermophilic serpin conserpin. To date, there are four well studied thermophilic serpins; thermopin (*Tm* 65°C[276]), tengpin (*Tm* 90°C, Chapter 4 SI: Figure 4), aeropin (*Tm* 100°C+[277]) and consensus-designed conserpin (*Tm* 100°C+[150]). Conserpin and aeropin are the most thermostable serpins known, with a thermal denaturation above 100°C. Conserpin was selected as a starting serpin for engineering stability as it has a high sequence identity to α1-AT (59%). Through studying the interactions in the native conformation of conserpin, there are various regions spread throughout the molecule that are hypothesized to contribute to thermal stability. A total of 8 regions were selected and grafted onto α1-AT.

Each graft was synthesized and cloned into a vector that has previously produced soluble WT α1-AT in high concentrations[91,162]. However, each graft expressed into inclusion bodies. Modification of the protein expression protocol was attempted, including expression in another cell line (such as BL21 (DE3) pLysS in case the protein was toxic to the cells), different inducer (IPTG) concentrations, and various induction temperatures[278,279]. Each attempt continued to result in the grafts expressing into inclusion bodies. The refold protocol for insoluble α1-AT has been

established[274,280], and out of the eight grafts expressed, six grafts fold into the metastable, native conformation, while the remaining two grafts fold into the latent, inactive conformation (Figure 5). Interestingly, the two grafts that fold into an inactive conformation, β-sheet C stapling and Helix-H, are present in the B/C barrel with the addition of a salt bridge. The B/C barrel is hypothesized to be the first region of the serpin to fold, with mutations increasing the propensity to misfold and aggregate[76,79]. The inclusion of the salt bridge in both grafts to stabilise the native conformation could not be compensated by the mutations introduced. As a result, the mutations decreased the high energy barrier between the native and inactive conformations, preventing folding to halt at the native conformation and allowing both grafts to fold into an inactive conformation.

Thermal denaturation monitored by circular dichroism determined that out of the 6 grafts that refold into the native conformation, three are more thermostable than WT α1-AT (Breach, F51 and Helix-F), while the remaining three (Citrate-binding, B/C barrel and T59S) are slightly less stable than WT. Interestingly, two grafts (F51 and T59S) have introduced mutations in the hydrophobic core, yet the F51 graft is more thermostable than T59S. The F51 graft packs against β-sheet A, where introducing this graft improves packing and favourable interactions against β-sheet A (Figure 9). This occurs through increasing the side chain of residue 291 leucine to phenylalanine, which is compensated by decreasing the side chains of residue 340 (isoleucine to valine) and residue 374 (methionine to isoleucine, also present in Multi7). Residue phenylalanine-51 (Phe-51) has been mutated previously, where decreasing the side chain to a cysteine or leucine increased the stability (3.0 kcal mol^{-1} and 2.1 kcal mol^{-1}, respectively, with Phe to Leu mutation present in the Multi7 variant)[66,281], despite producing a small cavity. Furthermore, a recent study on the Phe-51 to leucine mutation observed an approximate 10°C increase in thermal stability compared to WT α1-AT (54°C compared to 44°C[282]). However, deleting Phe-51, as present in α1-AT's Mmalton variant, leads to an increased propensity to misfold and aggregate[283–285]. Interestingly, increasing the side chain of a residue 53 in the same region can also increase the propensity to misfold and aggregate (Figure 9). The mutation of conserved serine-53 to phenylalanine in another α1-AT mutation, Siiyama, affects the ridge that allows the opening of β-sheet A, forcing β-sheet A to remain open and allowing polymerization to occur through the insertion of one serpins RCL into β-sheet A of another. The polymerization of the Siiyama α1-AT (Ser-53 to Phe) occurs at a higher degree than other α1-AT misfolding variants, indicating that this mutation has affected the stability of the serpin[286,287]. Therefore, mutations in the hydrophobic core can either increase or decrease the stability.

In the F51 graft, Phe-51 has not been mutated, instead three of the surrounding residues have been mutated (Ile-340Val, Met-374Ile and Leu-291Phe) (Figure 9). Of the three mutated residues, two have previously been mutated, isoleucine-340 (Ile-340) and methionine-374 (Met-374)[66,211]. Decreasing the side chain of Ile-340 to valine (Ile340Val) did not affect the stability of α1-AT, yet mutating methionine-374 to isoleucine (Met-374Ile) increased the thermodynamic stability (ΔΔG) by 2.3 kcal mol^{-1} [66,211]. Combining both these mutations, in addition to a mutation at residue 291 (Leu-291Phe), increased the thermostability of α1-AT.

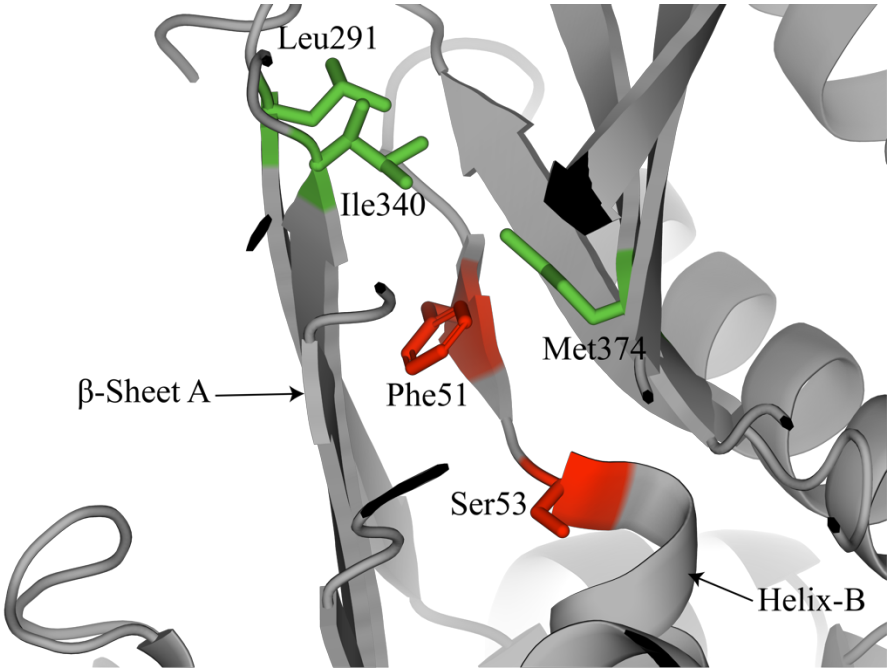


Figure 9. Mutations in the hydrophobic core that affect stability. Phe-51 and the surrounding residues can affect the stability of α1-AT. The F51 graft (*green*) has an increase in stability. Mutating Phe-51 or Ser-53 (*red*) can either increase or severely decrease the stability (PDB ID: 3NE4[269]).

The other graft in the hydrophobic core, T59S graft, faces away from β-sheet A, in close proximity to helix-A. In the T59S graft, mutating the threonine to serine (Thr-59Ser), along with residues leucine-30 to asparagine (leu-30Asp) and alanine-58 to serine (ala-58Ser), were modelled to introduce favourable polar and non-polar interactions in an attempt to stabilize the serpin. Introducing these interactions, however, did not stabilise the native state, instead, thermally destabilised this graft the most out of all the grafts. Previously threonine-59 has been mutated to an alanine (as in Multi7), producing a small pocket. Producing this small pocket led to an increase of thermodynamic stability (ΔΔG) of 1.0-1.2 kcal mol^{-1})[66,211]. Therefore, increasing the side chain of Thr-59, as in the T59S graft, decreases the stability of α1-AT, despite attempting to create

favourable interactions. This indicates that mutating stabilizing residues of one serpin onto another may not necessarily increase the stability of the acceptor serpin[213].

A total of four grafts are present in the B/C barrel (Helix-H, B/C barrel, citrate-binding and β-sheet C stapling), two fold into the native conformation (by testing the inhibitory activity against trypsin, citrate-binding and B/C barrel grafts), while the remaining two, β-sheet C stapling and Helix-H grafts, fold into an inactive conformation (no inhibitory activity against trypsin, Figure 5). One of the natively folded B/C barrel grafts is the citrate-binding graft. This graft contains mutations with increases in the side chains (asparagine-228 to tyrosine, Asn-228Tyr; and leucine-285 to glutamic acid, Leu-285Glu). Citrate stabilizes WT α1-AT from unfolding and polymerization through binding to a surface pocket of the B/C barrel[288]. The citrate-binding graft involves increasing the size of the side chains which fills pocket where citrate binds, stabilizing the native conformation. However, introducing these mutations to fill that pocket did not stabilize the native conformation, instead destabilized it (decreased the thermostability).

The last graft in the B/C barrel, the B/C barrel graft (Phe-253Ile, Phe-275Trp and Glu-279Leu) is modelled to improve the hydrophobic packing in the barrel, thus stabilizing the native conformation. Introducing a tryptophan (Phe-275Trp) would have been compensated through decreasing the size of the two other residues (Phe-253Ile and Glu-279Leu). However, this graft did not increase the thermal stability, but instead decreased it. Interestingly, Phe-275 has previously been mutated to a leucine, with the decrease in side chain leading to an increase in stability (ΔΔG) of 0.8 kcal mol^{-1} [69]. Despite the B/C barrel graft having two residues with smaller side chains to compensate the addition of a tryptophan, it seems introducing a small pocket is favourable for α1-AT's stability.

Two regions on the serpin are important for its inhibitory mechanism: the breach region and the Helix-F. The breach region is located at the top of β-sheet A and at the base of the RCL, and is the first region of β-sheet A into which the RCL inserts. Mutations in the breach leads to a decrease in kinetic stability and the propensity to misfold and polymerise during folding. This is most evident in α1-AT's Z mutation, a substitution of glutamic acid-342 to lysine. In WT α1-AT, Glu-342 forms a salt bridge with lysine-290 (lys-290)[136]. In the Z mutation, the complete charge swap from Glu-342 to a lysine produces a charge-charge repulsion with lys-290. As a result, the top of β-sheet A remains open and is more flexible than that in WT[137,138]. The opening of β-sheet A in the Z mutations leads to misfolding during synthesis through loop-sheet mechanism, where the

RCL of one Z α1-AT inserts into the β-sheet A of another[289]. In comparison to α1-AT, conserpin contains an extended salt bridge network, contributing to its thermostability by preventing the opening of β-sheet A, as observed in molecular dynamics (MD) simulations[150]. Furthermore, this salt bridge network does not make conserpin immune to the destabilizing nature of the Z mutation. By introducing the Z mutation onto conserpin, there is a decrease in thermal stability (from 72.5°C to 60°C in 2M GndHCl) and thermodynamic stability (ΔΔG of -12.8 kcal mol^{-1}, a loss of -10.04 kcal mol^{-1}). This indicates the importance of the breach region in stability of α1-AT. Individual mutations in the breach region are not sufficient to increase the thermal stability of α1-AT, as observed in a previous study[270]. Therefore, through grafting the whole breach region of conserpin onto α1-AT (Figure 10), an increase the thermal stability of α1-AT by 7°C is observed, the most of all grafts.

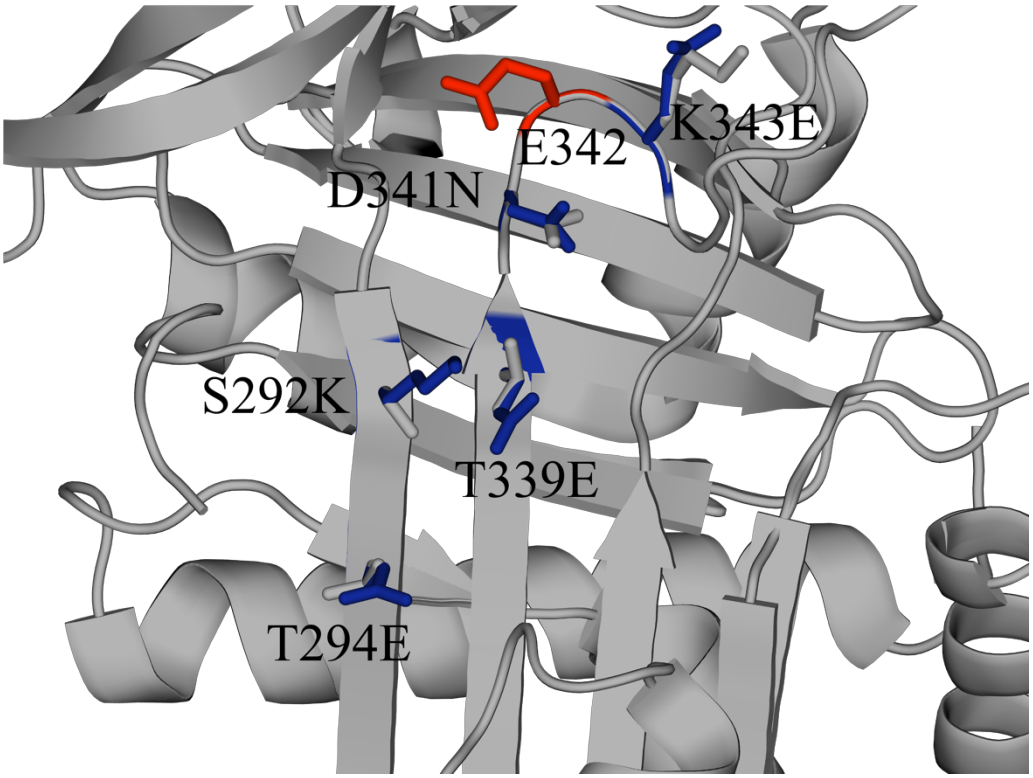


Figure 10. The breach region of α1-AT. The Z mutation, E342 (*red*) causes off-pathway misfolding. Introducing more salt bridges by grafting the breach region of conserpin onto α1-AT (*blue*), increases the thermostability more than any other graft (PDB ID: 3NE4[269]).

The second region important for a serpins inhibitory activity is Helix-F. Helix-F lies in front of β-sheet A, must unfold and shift during the opening of β-sheet A for RCL insertion, while also acting as a physical barrier to prevent polymerization[290]. Mutations in Helix-F have destabilized the native state and affected inhibitory activity[64]. Helix-F is hypothesized to contribute to conserpin's

stability by tightly packing against β-sheet A and becomes rigid, possibly acting as a physical barrier that prevents RCL insertion and polymerization[150]. Furthermore, molecular dynamics (MD) simulations of conserpin and α1-AT show the Helix-F is less flexible in conserpin, with conserpin's tryptophan-132 packed tightly against to β-sheet A. The corresponding residue in α1-AT, tyrosine-160 fluctuates in and out of a hydrophobic pocket, as observed in MD simulations[150]. The introduction of three mutations on either Helix-F or β-sheet A (Tyr-187Ala, Gly-115Ala, and Tyr-160Trp) (Figure 11) from conserpin onto α1-AT in the Helix-F graft allows the Helix-F of α1-AT to pack tightly against β-sheet A. Interestingly, increasing the side chain of glycine-115 to alanine seems to be favourable, while increasing the size even further, as present in $\alpha1\text{-AT}_{\text{Newport}}$ (Glu-115Ala) variant increases the propensity to misfold [291]. Therefore, the combination of all three mutations onto the Helix-F allowed a slight increase in thermal denaturation compared to WT α1-AT. This corresponds with previous studies, where the single Tyr-160 to Trp mutation also increased the thermal stability (by 5°C compared to WT α1-AT)[81].

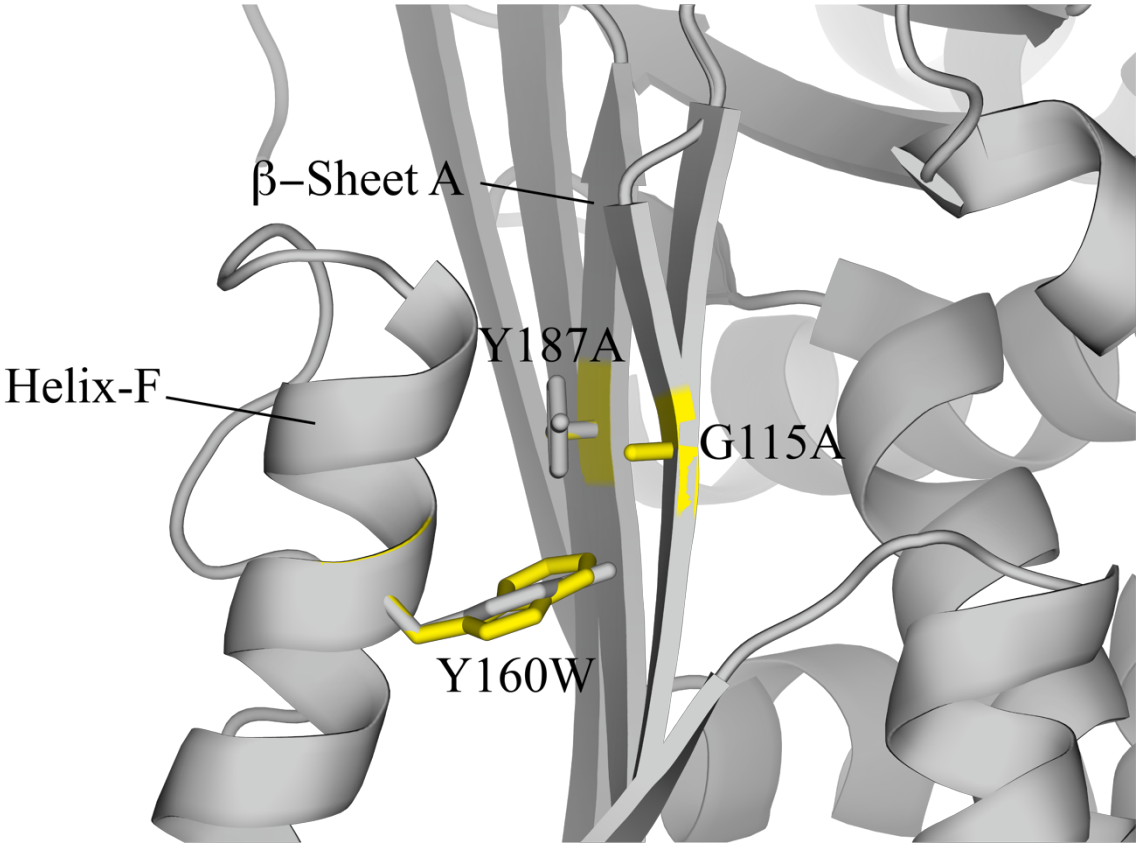


Figure 11. The Helix-F graft. The Helix-F graft (yellow) allows improved packing against β-sheet A. This increases the thermostability compared to WT α1-AT. One previously studied mutation, Y160W, also increased the thermostability[81] (PDB ID: 3NE4[269]).

Increasing the stability of proteins could possibly affect their function. This is true in some cases[262,265,266], which has led to a stability/ function trade-off hypothesis[262]. However, not all proteins have this trade-off[250,263,267]. Increasing the stability of α1-AT, through either point mutations (e.g. Gly-117[94]), or multiple mutations[212], above 13 kcal mol^{-1} (ΔΔG) leads to a decrease in inhibitory activity[212]. Furthermore, conserpin could possibly be a result of this stability/ function trade-off. With a thermodynamic stability ($\Delta G_{denatured\text{-}native}$) of -23.2 kcal mol^{-1}, yet has an SI against trypsin of 1.8[150]. This indicates that conserpin may be too stable, where β-sheet A may not open as rapidly as necessary to allow RCL insertion before trypsin's de-acylation step in the catalytic mechanism occurs. This proves that serpins are not immune to the stability/ function trade-off. However, the 3 thermostable grafts are active against α1-AT natural target, HNE. The Breach and F51 grafts have an inhibitory ratio (SI) identical to WT α1-AT, while the Helix-F requires a slightly higher ratio to inhibit HNE. The increase in the serpin: protease ratio of Helix-F could possibly be a result of the increase rigidity of the Helix-F against β-sheet A, as designed. Therefore, increasing the stability of these grafts has no large effect on their function.

Each of the grafts that increased the thermostability of α1-AT are in regions that play a role in the folding of the serpin. The F51 graft introduced mutations in the hydrophobic core, which is the first region to fold[76,79], while mutations in the breach region lead to misfolding and polymerization during folding (as observed in the Z mutation[289]). During folding, the Helix-F is highly disrupted and undergoes a conformational change during unfolding and polymerization[64,81]. During α1-AT's folding pathway, the folding intermediate is aggregation-prone, which can lead to aggregation during refolding[274]. The yield post-refold for each of the grafts increased in comparison to WT α1-AT, with the F51 graft increasing the yield the most, while the Helix-F graft only slightly increased yield. This is plausible, as the F51 graft was modelled with improved packing and introduction of favourable interactions, possibly increasing the folding rate and decreasing the time the aggregation-prone intermediate is present. This hypothesis, however, will need to be confirmed by stopped-flow kinetics experiments. Furthermore, the increase in yield of the Breach graft could be a result of the increase number of salt bridges forcing the closure of β-sheet A during the last stages of folding. The modelled improved packing in the Helix-F graft seems not contribute to the yield post-refold as much as the other two grafts. This suggests that this graft, and Helix-F does not contribute to a decrease in aggregation propensity during folding.

The increase of yield of each of the three thermal stable grafts was not a result of any changes to the folding intermediate. The folding intermediate remains populated over a wide denaturant

concentration. The difference between the grafts and WT α1-AT is the peak fluorescence of the grafts occurs at a high denaturant concentration than WT α1-AT. Therefore, if the folding intermediate has not been affected, the rate at which each serpin folds through this intermediate could possibly result in a higher post-refold yield. Further kinetic folding experiments (i.e. stopped-flow kinetics) would be beneficial to confirm this hypothesis.

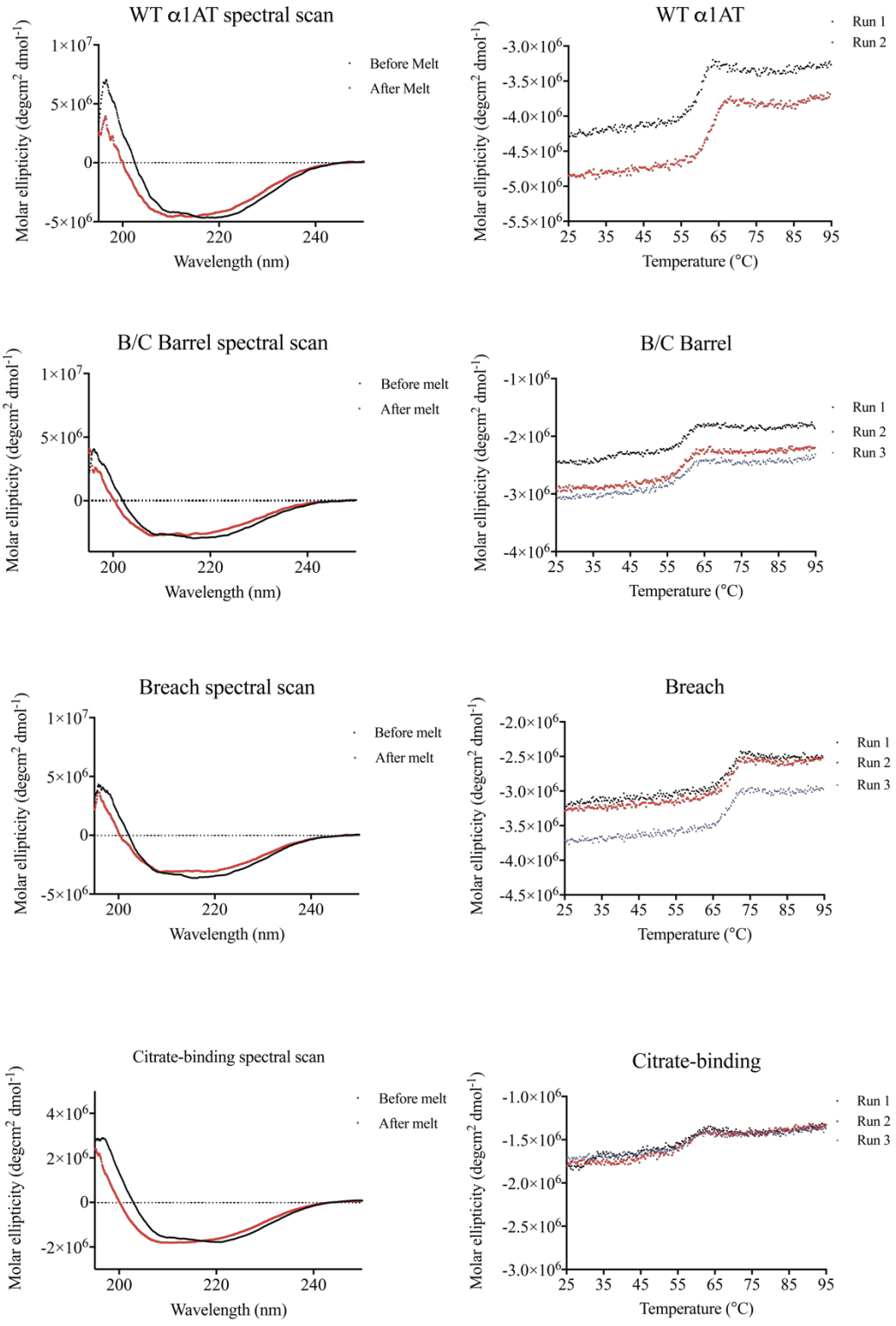
To increase the thermal stability of α1-AT even more, the three thermally stable grafts were combined to produce 3stable graft. Thermal denaturation monitored by circular dichroism determined the thermal stability of the 3stable graft to be additive of the three individual grafts. The midpoint of thermal denaturation of 3stable is the highest reported for an engineered α1-AT, and is higher than thermophilic serpin thermopin (74°C for 3stable compared to 65°C for thermopin[276]). This puts the 3stable graft on par with the Multi7 in terms of stability[211]. The increase in stability only marginally affected the inhibitory function with an SI of 1.28:1 against HNE (compared to 1:1 of WT α1-AT). Despite the increase in thermal stability, the 3stable graft expresses into inclusion bodies during recombinant expression. Nevertheless, the 3stable graft refolds and produces a higher yield than WT α1-AT and the individual grafts, while having no effect on the folding intermediate besides the intermediate being populated at a higher denaturant concentration. Therefore, it is possible to increase the stability of α1-AT without any large effect on the function or the folding of the serpin. Further stability and folding (e.g. half-life determination and folding kinetic) studies on the three individual thermal stable and the 3stable grafts can confirm the superiority of using rational design to engineer serpin stability over random mutations.

4.4. Conclusion

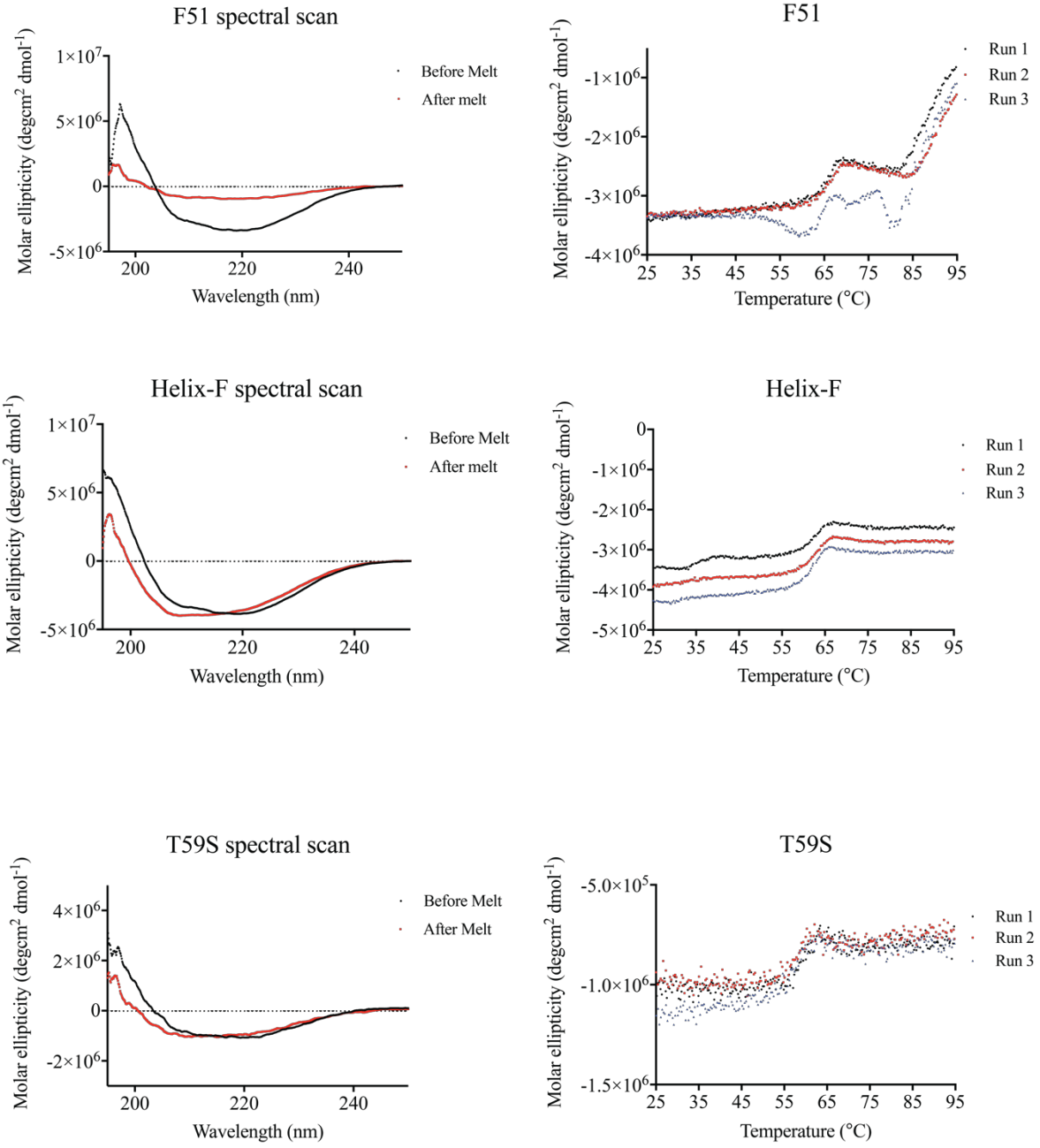
In this study, 8 regions of a thermostable serpin conserpin, were grafted onto α1-AT. Out of the eight grafts, two grafts fold into an inactive conformation, while three natively folded serpins are more thermostable than WT α1-AT. These three thermal stable grafts are active against HNE, yield more monomeric protein post refold while having a similar folding intermediate to WT α1-AT. Combining these three thermally stable grafts together produced an even more thermally stable serpin, 3stable. The 3stable graft yields more monomeric protein post refold than each of the single thermal stable grafts, while remaining active as an inhibitor.

Further studies on the folding pathway of the three thermally stable, the 3stable graft and the other three destabilized grafts can only provide information on how each region can affect the folding rates and intermediates, providing a picture of how α1-AT folds.

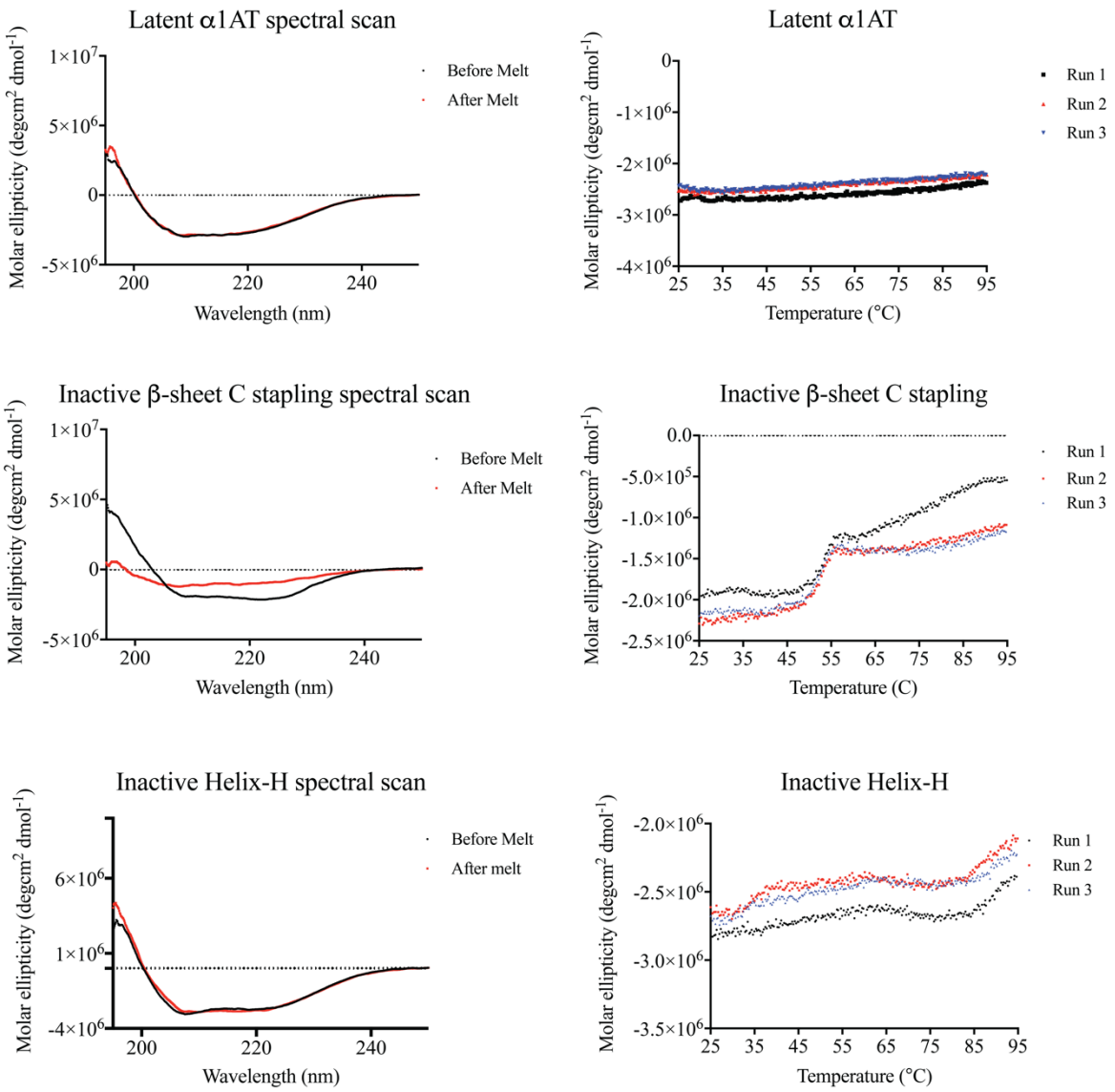
4.5. Supporting Information



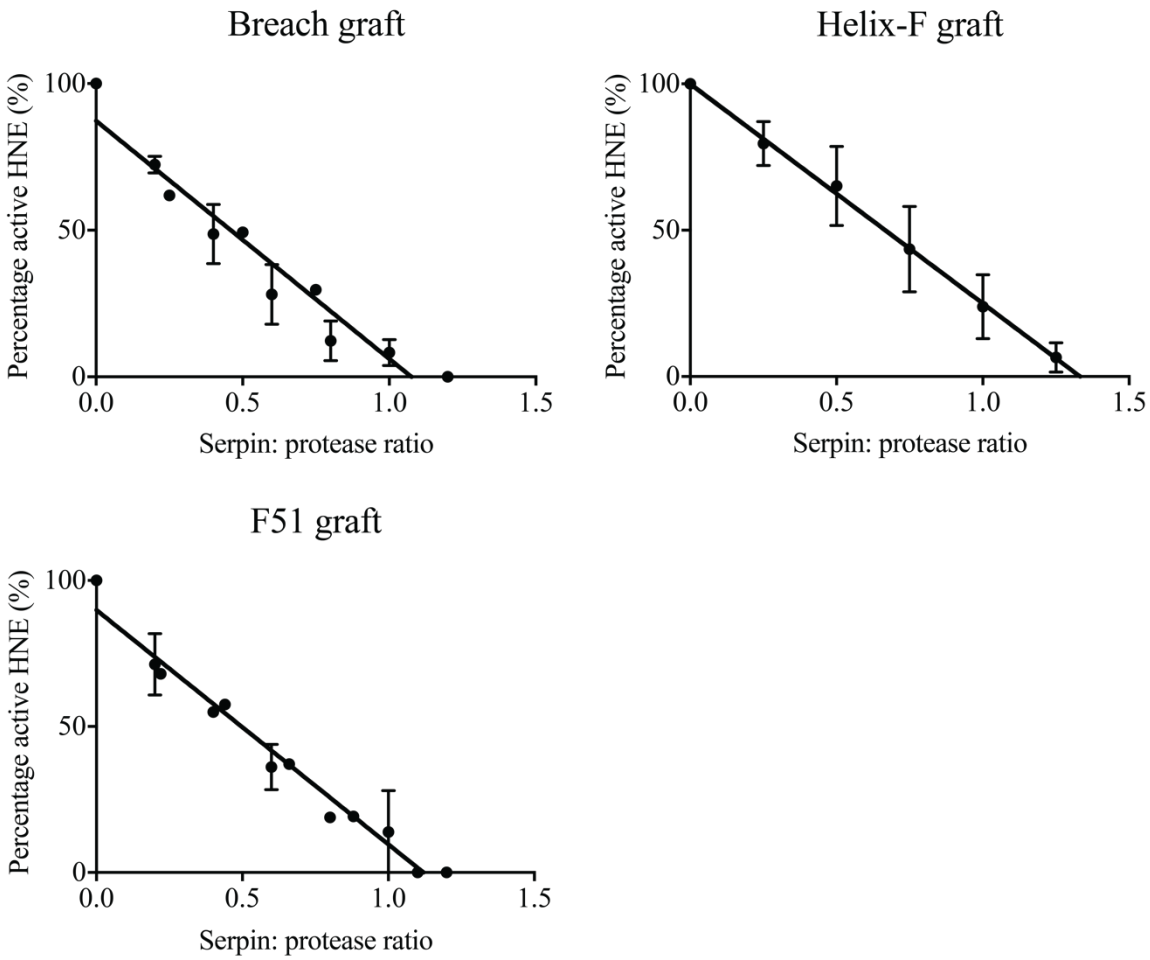
SI Figure 1a. Spectral scans (195-250nm scan, *left*) and thermal denaturation curves (measured at 222nm, *right*) of WT α1-AT and 3 of the native folding grafts.



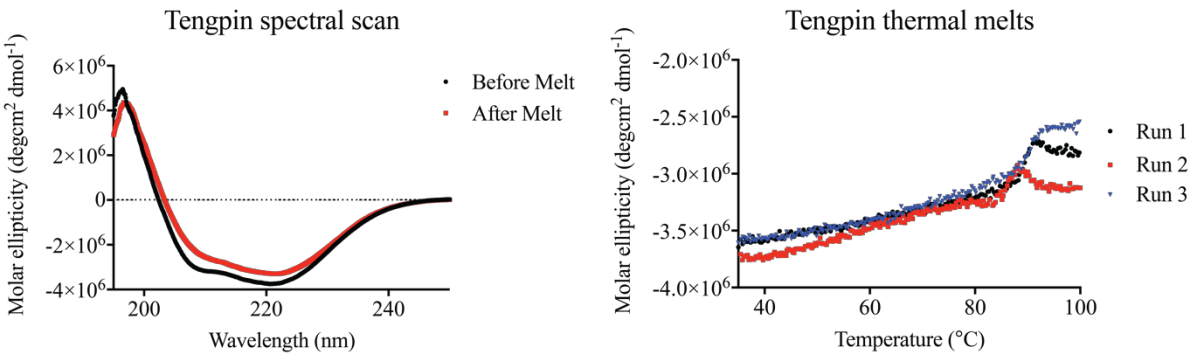
SI Figure 1b. Spectral scans (195-250nm scan, *left*) and thermal denaturation curves (measured at 222nm, *right*) of the remaining 3 of the native folding grafts.



SI Figure 2. Spectral scans (195-250nm, *left*) and thermal denaturation curves (measured at 222nm, *right*) of latent WT α1-AT and the 2 grafts that fold into an inactive conformation.



SI Figure 3. The inhibitory activity of the 3 thermal stable grafts, measured through determining the stoichiometry of inhibition (SI) against α1-AT's target enzyme human neutrophil elastase (HNE).



SI Figure 4. Spectral scans (195-250nm, *left*) and thermal denaturation curve (measured at 222nm, *right*) of thermophilic serpin, tengpin. Tengpin has a midpoint of thermal denaturation at 90ºC.

Chapter 5: Investigating the folding pathway of conserpin

5.1. Introduction

Many proteins require the adoption of a 3-dimensional structure to exhibit its function. How proteins fold into the myriad of unique structures has puzzled scientists for many years. The number of conformations that an unfolded, polypeptide chain can adopt is vast, yet, protein folding occurs on the biologically relevant timescale of microseconds[292]. How this occurs can be viewed as a funneled energy landscape[8]. Folding through a funneled energy landscape involves the denatured polypeptide chain adopting an ensemble of structures, which can proceed to push the polypeptide chain to fold through multiple pathways to reach the native state[22,293]. The landscape in which a protein can fold can be relatively smooth or rugged. The smoothness of a landscape reflects the lack of deep valleys and high energy barriers[9]. Proteins which fold through a smooth landscape are often two-state folders, where either the denatured or the native state is populated, with no intermediates observed[9]. A majority of proteins, however, fold through a rugged landscape. The ruggedness of an energy landscape arises from the formation of non-native conformations (unfolded states, transition states and intermediates) along the folding pathway[24]. To understand how a protein folds in detail, characterization of the non-native conformations is essential. New experimental techniques have allowed the observation and characterization of the non-native conformations, providing details of a proteins folding pathway.

5.1.1. Experimentally determining how protein's fold

A major challenge in understanding how a protein folds is being able to characterize the non-native species (transition states and intermediates) along the rugged folding landscape. These non-native species can be sparsely populated, therefore difficult to observe. Various experimental techniques have been developed to monitor how a protein folds, from picosecond to second timescales[23] (Table 1).

Technique	Information	Reference
Circular dichroism (CD)	The use of polarized light allows for detection of secondary and tertiary structure formation during protein refolding. Can detect structure on a sub-millisecond to millisecond timescale in conjunction with continuous or stopped-flow techniques	[294]
Fluorescence	• *Intrinsic:* utilizes fluorescent amino acids (e.g. tryptophans) to provide specific local information of secondary structure during unfolding in different denaturation concentrations • *Extrinsic:* the binding of fluorescent dye to hydrophobic regions during unfolding provides information of folding intermediates	[295]
Hydrogen-deuterium exchange (HDX)	Rapid exchange of amide hydrogens with deuterium can provide residue specific information during protein folding. During folding, amides that form hydrogen bonds are protected, while amides that do not are rapidly exchanged. Regions that have undergone exchange can be analyzed by nuclear magnetic resonance (NMR) or mass spectrometry.	[17,296]
Protein engineering (ϕ-value analysis)	Introducing mutations throughout the protein and determining its effect on the rate of folding and stability (free energy). The stability can be used as restraints for molecular dynamics simulations to generate structural models of non-native species during folding.	[16,297,298]
Single-molecule Förster Resonance Energy Transfer (smFRET)	The attachment of fluorophores to a protein and observing the change in energy transfer efficiencies, and therefore distance, between the fluorophores during the proteins folding events. Information derived from energy transfer efficiencies include detection of states not observed in other techniques, dynamics of the unfolded state and ability to determine the equilibrium constants.	[299]

Table 1. Experimental techniques to study protein folding, with observations of non-native states.

5.1.2. Studying the folding pathway of proteins using single-molecule Förster Resonance Energy Transfer (smFRET)

In recent years, a major development in observing and understanding how a protein folds has come from studying the folding pathway at a single-molecule level[299]. Using single-molecule techniques, detailed insight on the conformation of the polypeptide chain along the folding pathway can be assessed, along with the detection and separation of rare species that would otherwise be undetected in other experimental techniques[23,300].

One of the common single-molecule techniques is single-molecule Förster Resonance Energy Transfer (smFRET), where the excitation energy is transferred from the donor dye to the acceptor dye[301], with the resulting measurement being the efficiency of the energy transfer[301,302]. The two dyes can be separated along the polypeptide chain, and when the protein begins to fold, the efficiency of the energy transfer (E_{FRET}) can be observed at each stage of folding. The E_{FRET} is determined by the following equation (equation 1):

$$E_{FRET} = \frac{1}{1 + (\frac{r}{R_0})^6}$$

where r is the distance between the donor and acceptor dyes, and R_0, the Förster distance, is the distance between the dyes where E_{FRET} is 50%[301,302]. The energy transfer efficiency is strongly dependent on the distance between the two dyes. The distance between the dyes are typically between 30-70Å apart. The specific distances for transfer efficiency between the two dyes allows energy transfer to measure the distance and the dynamics within the protein[303,304].

In protein folding, smFRET can provide information on the distance between the two dyes as the protein folds. [305]. Under different denaturation concentrations (e.g. guanidinuim hydrochloride, GndHCl), the transfer efficiency between the two dyes will differ, allowing the observation of the folded, unfolded and any non-native species/ subpopulations. As the protein unfolds, the distance between the dyes increases, reducing the energy transfer between the dyes, and thus the transfer efficiency[300]. The energy transfer efficiency between the two dyes is determined as a ratio (equation 2):

$$E_{FRET} = \frac{I_A}{I_A + I_D}$$

In equation 2, I_A and I_D is the fluorescence detection for the acceptor and donor dyes, respectively[301,304]. The transfer efficiency as a function of denaturation concentrations can be

plotted as a histogram, determining the number of peaks in the transfer efficiency, which corresponds to the minimal number of non-native species/ subpopulation[299,304].

The use of smFRET in studying the folding landscape of proteins was first pioneered by Hochstrasser and colleagues[306,307], and Weiss and colleagues[305]. These studies used proteins that fold in a two-state manner (either folded or unfolded) and followed the E_{FRET} under different denaturation concentrations (similarly to equilibrium unfolding experiments). Hochstrasser and colleagues[306,307] observed the protein of interest, a two-stranded coiled-coil yeast transcription factor GCN4, fluctuate between the folded and unfolded state at the midpoint of GCN4's equilibrium. Furthermore, Weiss and colleagues[305] observed the folding of a well-studied small single-domain protein, chymotrypsin inhibitor 2, and compared the folding pathway to the folding pathway observed through protein engineering[308]. The results observed during the folding of CI2 by smFRET agreed with protein engineering results, indicating smFRET is highly comparable to other protein folding techniques.

The use of smFRET in studying the folding pathway of proteins allows for the determination of non-native species/subpopulations that cannot be observed with other protein folding techniques. Non-native species/subpopulations that could only be determined by smFRET were observed by Deniz and colleagues[309] when observing the folding of α-synuclein, an intrinsically disordered protein. Through probing the folding of α-synuclein by Far-UV circular dichroism, only 3 conformational states were detected, whereas with smFRET, a total of 5 conformational states were detected. Therefore, the use of smFRET in studying the folding pathways is reliable against other experimental techniques with the benefit of observing any non-native species/subpopulation that might not be otherwise observed using other techniques.

5.1.3. Using smFRET in studying the folding pathway of a α1-antitrypsin

As protein folding is under both kinetic and thermodynamic control, a majority of proteins within the proteome fold into the lowest free-energy minima conformation[19]. There are some exceptions, the most notable being proteins from the serine protease inhibitor (serpin) superfamily. Serpins fold into a metastable conformation which is essential for function[57,87]. This metastability, however, renders the proteins susceptible to misfolding[84].

Many studies have used a variety of experimental techniques to understand how serpins fold. Using α1-antitrypsin (α1-AT) as the archetypal serpin, these studies have revealed that serpins fold through multiple states, with at least one molten globule intermediate[82,123,310]. This folding intermediate contains approximately 80% of the secondary structural elements present in the native state, including partially folded β-sheets[82,123,310]. It is this folding intermediate that is aggregation prone and can be detrimental to human health[79,289,311]. The folding intermediate may not always be detected possibly due to rapid folding, as previously observed in two thermophilic serpins[156,277].

The overall structure of a serpin can be described as an ellipsoidal fold, characterized by a long axis (from the RCL to the bottom of β-sheet A) and a short axis[77]. How the serpin folds through a molten globule intermediate into this ellipsoidal fold was recently determined through the use of smFRET[77]. The folding of α1-AT studies by smFRET involved the binding of maleimide fluorophores to α1-AT's solvent exposed cysteine ($C232_{\alpha1\text{-}AT}$), and two introduced cysteines on the short ($S47C_{\alpha1\text{-}AT}$) and long axis of the serpin ($S313C_{\alpha1\text{-}AT}$) (Figure 1). Equilibrium unfolding and refolding (incubating the serpin under different denaturation concentrations) was performed to observe the overall change in structural dimensions. It was observed that α1-AT unfolds and refolds through two folding intermediates, with the contraction of the long axis and expansion of the short axis, creating folding intermediates with more spherical dimension compared to the native state[77]. This study further emphasizes the strength of studying the folding of proteins using smFRET.

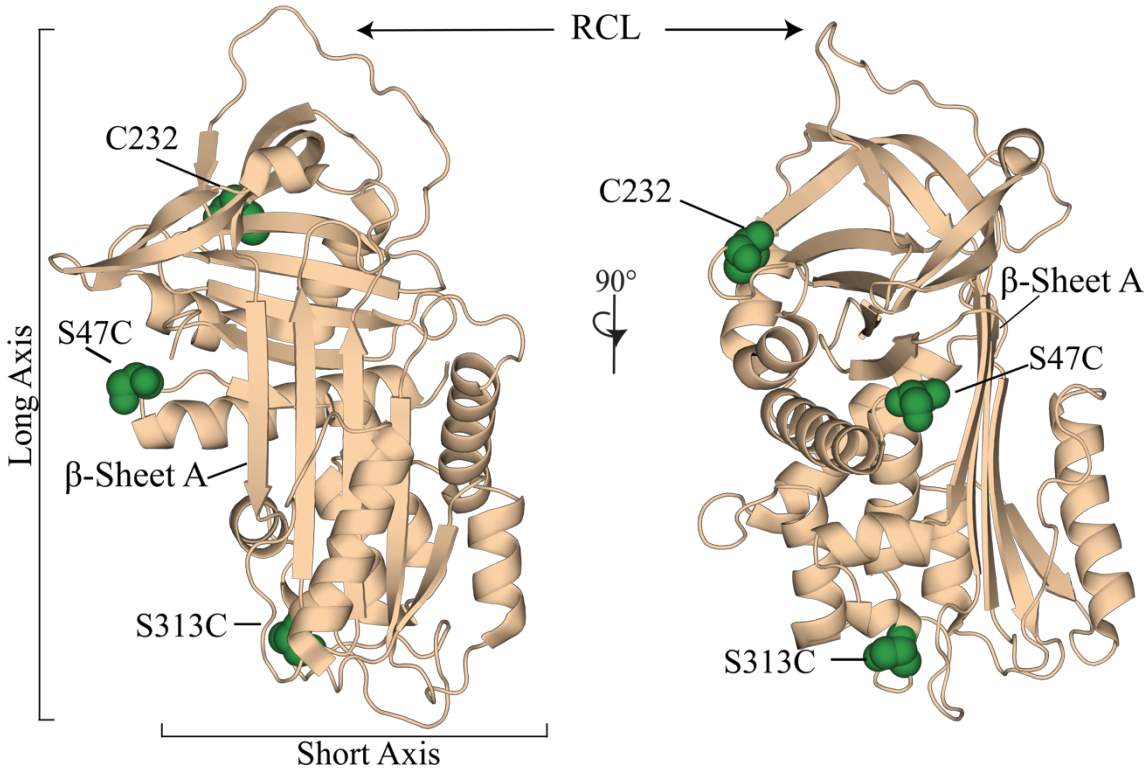


Figure 1. Solvent exposed cysteines onto α1-AT for smFRET. Two additional cysteine residues were introduced onto α1-AT (PDB ID: 3NE4[269]) to study the folding by smFRET. The long axis and short axis of the serpin have been noted.

In this chapter, the folding pathway of conserpin was studied. Previous biophysical experiments on the folding pathway of conserpin show different results, depending on the technique used[150]. Chemical denaturation shows conserpin unfolds in a two-state manner, with one transition from folded to unfolded[312]. However, stopped-flow kinetics observed a three-state unfolding mechanism, with the presence of two folding intermediates[150]. The folding intermediates observed with stopped-flow kinetics is poorly populated, that is, is only present for a short period of time and under a small range of denaturant concentrations. The uncertainty in the folding intermediates provides difficulty in understanding conserpin's folding pathway. Therefore, the use of smFRET can provide a clearer understanding of how conserpin folds. Using conserpin as a model would also provide an explanation on how extremely thermostable, naturally occurring serpins fold at a rapid rate. This can lead to using thermophilic serpins to engineer mesophilic serpins, like α1-AT, to avoid folding through an aggregation-prone folding intermediate, which can be detrimental to human health.

5.1.4. Aim of this study

Single-molecule Förster Resonance Energy Transfer (smFRET) can provide detailed insight into the folding pathways of proteins, detecting non-native species that would otherwise be undetected by other experimental techniques. The study of α1-AT using smFRET detected two folding intermediates, an additional intermediate that was not detected using other techniques. The aim of this study is to use smFRET in studying the folding pathway of a consensus-designed serpin, conserpin, and compare the pathway to that of α1-AT. This will provide insight into how conserpin, an extremely thermostable and aggregation-resistant serpin folds, providing insight on how naturally occurring thermostable serpins fold. In this study, mutation pairs (producing 2 variants) were introduced onto conserpin, with each variant assessed for changes in thermal stability and activity.

5.2. Results

5.2.1. Introducing solvent exposed cysteine

In order to study and compare the folding between conserpin and α1-AT, the same maleimide fluorophore dyes and solvent-exposed cysteine residues were introduced onto conserpin using site-directed mutagenesis[77] (Figure 2A). Unlike α1-AT, conserpin contains one cysteine residue (C209) buried in the B/C barrel (strand 1 of β-sheet B). In this folded/native position, the fluorophore could not access the cysteine residue for binding and may form aberrant disulfide bonds when going from the unfolded to folded conformation. Therefore, this cysteine was mutated to a serine (C209S) to avoid any potential disulfide-driven aggregation during folding. A new cysteine was mutated into the region corresponding to α1-AT's single cysteine (on strand s3C, mutating D205C). Residue D205 is next to the residues that correspond to α1-AT's cysteine ($C232_{\alpha1\text{-}AT}$, $D204_{conserpin}$). The exact corresponding residue in conserpin was not mutated, as it is hypothesized to stabilize the hydrophobic core of the B/C barrel through an introduction of a salt bridge network[150]. The two additional cysteine residues were successfully introduced to the corresponding cysteine pair residues in α1-AT (short axis: D25C, long axis: S285C). The results of the site-directed mutagenesis produced two conserpin variants that each contain two solvent-exposed cysteines: D25C (D25C, D205C, C209S) and S285C (S285C, D205C, C209S). Each mutation was confirmed by DNA sequencing (Chapter 5 Supporting Information: Figure 1)

5.2.2. The double cysteine variants remain soluble during expression

To ensure the double-cysteine conserpin variants continued to express in the soluble fraction during recombinant expression, pre- and post- induction samples were taken with the soluble and insoluble fractions checked for expression (Figure 2B). As with WT, both double-cysteine variants continued to express in the soluble fraction of the bacterial cells, with some leaky expression (Figure 2B). The purification of both variants occurred as previous[150], with the addition of a reducing agent (either 2-mercaptoethanol or DTT) to ensure prevention of disulfide-driven aggregation. Through two purification techniques (affinity chromatography and size-exclusion chromatography), the two double-cysteine conserpin variants were successfully purified.

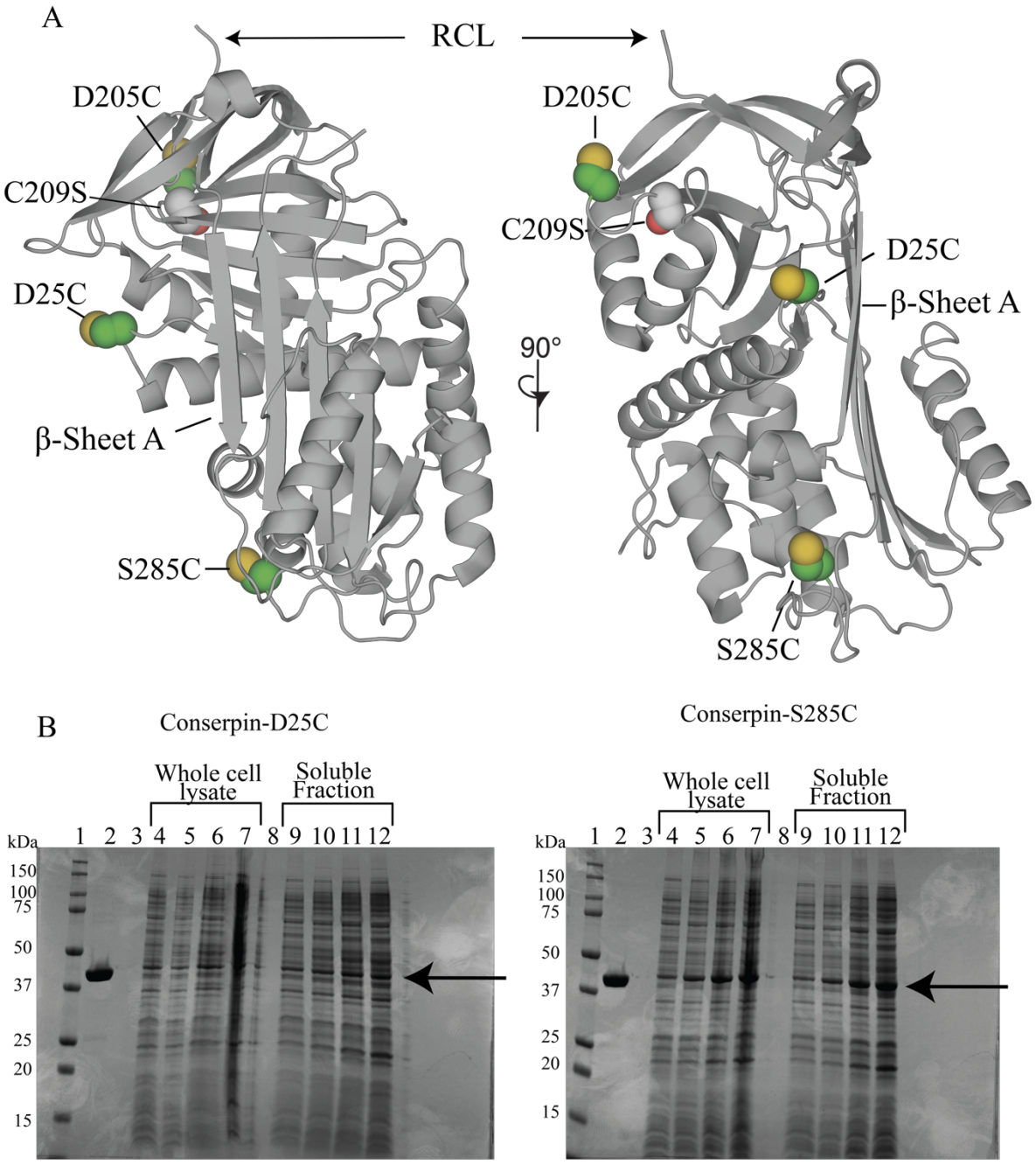


Figure 2. (A) Three solvent exposed cysteines were introduced onto conserpin in corresponding regions to that of α1-AT (PDB: 5CDX[150]). **(B)** Both double-cysteine variants were recombinantly expressed in the soluble fraction, as shown with pre- and post-induction gels: *1.* Molecular weight ladder, *2.* WT conserpin, *3. blank*, *4-7.* whole cell lysate of pre, 1 hour, 2 hour, and 3 hour post induction, respectively, *8. blank, 9-12.* Soluble fraction of pre, 1 hour, 2 hour, and 3 hour post induction, respectively. The conserpin variants are indicated with the arrow. Leaky expression is occurring, as conserpin variants are present in the pre-induction sample.

5.2.3. The thermostability of the conserpin variants were not affected

One of conserpin's characteristics is its high thermostability, where there was no unfolding profile when heating to 110°C, as determined by circular dichroism thermal melts (at 222nm)[150]. Heating the protein to 110°C did not change its structure either, with far-UV spectral scans of before and after heating being identical. To observe any affect introducing solvent-exposed cysteines has on the thermostability, both variants underwent spectral scans to observe any structural changes to the secondary structure, and circular dichroism thermal melts to determine the thermal midpoint of unfolding (*Tm*), by increasing the temperature from 25°C -100°C. Reducing agent DTT was present in each sample to minimize any potential disulfide-driven aggregation during heating. Spectral scans of conserpin and the two variants show D25C variant has a slight structural change in the α-helical segment (195-200nm) compared to conserpin, while S285C variant is identical (Figure 3a). Upon heating to 100°C, both variants exhibited slight structural changes, with a minute increase in molar ellipticity at 90°C. This was further confirmed with spectral scans of before and after heating, where the scans do not overlap (Figure 3b, c).

The slight change in molar ellipticity of both variants, however, is not sufficient to determine a midpoint of thermal unfolding (*Tm*). Therefore, guanidine hydrochloride (GndHCl) was added to the sample at a final concentration of 2M, to help unfold the variants. Conserpin was also subjected to thermal unfolding under the same conditions as a comparison (Table 2).

Protein	**Thermal midpoint of unfolding (*Tm*) (°C) (n=3)**
Conserpin	71.3 ± 1.4
D25C	67.1 ± 0.5
S285C	67.9 ± 3

Table 2: The thermal midpoint of unfolding, measured by circular dichroism, of conserpin and the double-cysteine variants in the presence of 2M GndHCl and DTT.

In 2M GndHCl, Conserpin has a *Tm* of 71.3 ± 1.4°C (n=3). The double cysteine variants have a reduced thermostability compared to conserpin (Table 2), however, the reduced thermostability is not significant enough to compromise the protein (Figure 3D).

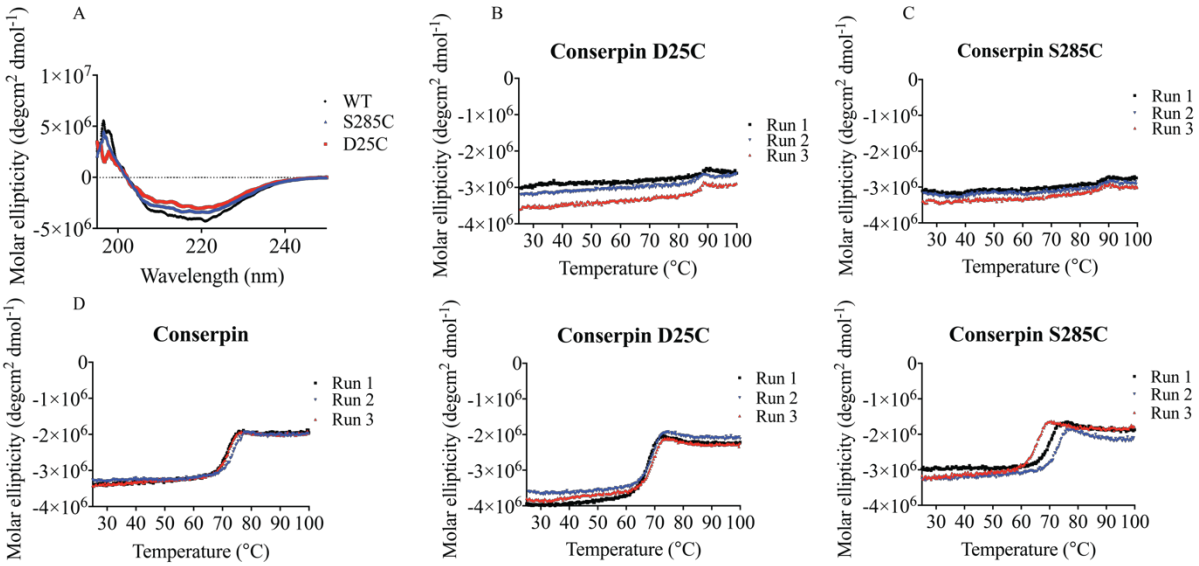


Figure 3. Circular dichroism (CD) thermal melts of conserpin variants. (A) Spectral scans show a slight change in structure of the D25C variant, while the S285C variant has a structure similar to conserpin. **(B, C)** The D25C and S285C did not undergo thermal unfolding when heated from 25 to 100°C. **(D)** Thermal unfolding was only achieved in 2M GndHCl, with midpoint of unfolding (*Tm)* for both variants similar to conserpin.

5.2.4. The activity of the variants was not compromised

Conserpin inhibits trypsin to a stoichiometry of inhibition (SI, the number of moles for a serpin to inhibit one mole of protease) of 1.8[150]. To ensure the variants inhibitory function was not compromised, the SI was calculated for each variant against trypsin (Table 3, Figure 4). The S285C variant has an SI against trypsin identical to conserpin, while the D25C variant has a slightly increased SI. This slight increase is negligible, given conserpin is a poor inhibitor to trypsin compared to other serpins against trypsin (such as α1-AT, where the SI against trypsin is 1:1[313]).

Protein	**Stoichiometry of Inhibition (SI) against trypsin (n=3)**
Conserpin	1.8[150]
D25C	2.1 ± 0.1
S285C	1.85 ± 0.15
α1-AT	1[313]

Table 3. The stoichiometry of inhibition against trypsin for conserpin and the double-cysteine conserpin variants. α1-AT was included as an archetypal serpin.

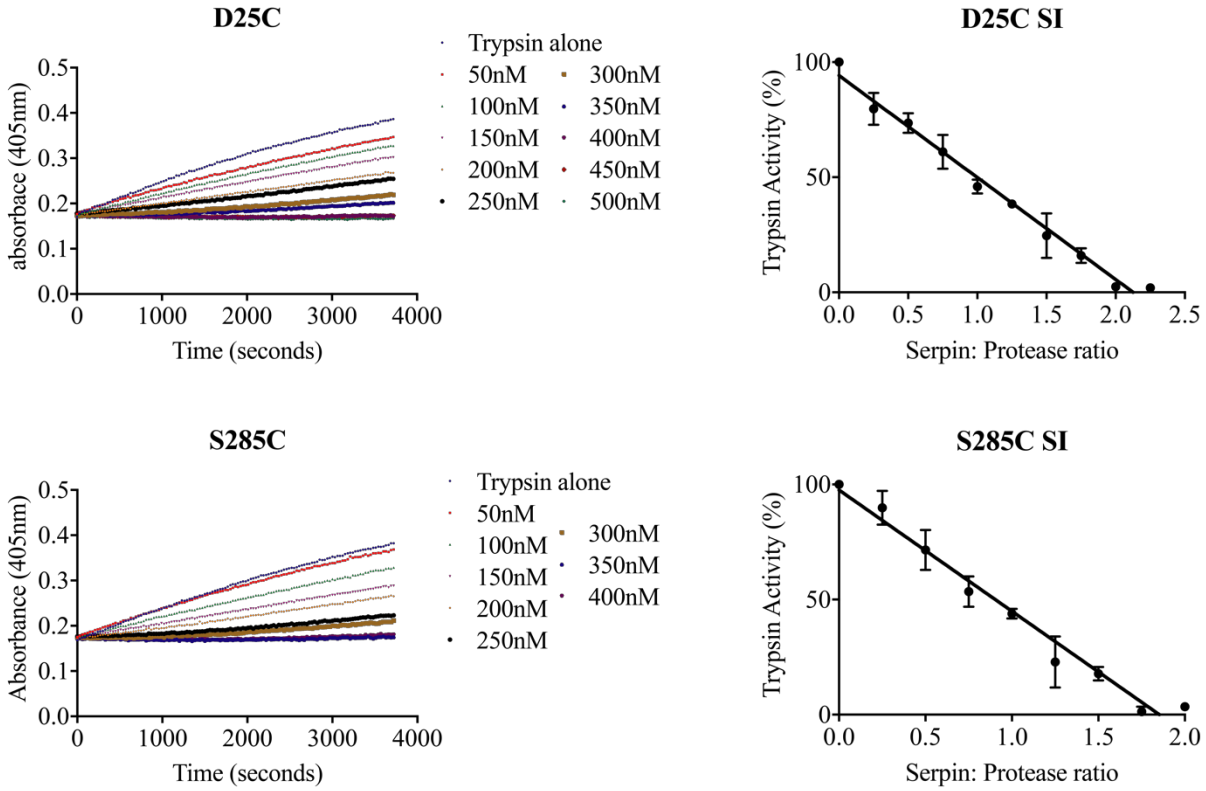


Figure 4. The inhibitory function of each conserpin variant. The D25C variant has a slight increase in SI, while the S285C variant has an SI that is identical to conserpin.

5.3. Discussion

Single-molecule techniques, such as smFRET, are a major development in allowing us to understand how proteins fold. The detection of non-native species that otherwise could not be detected through other experimental techniques, provides detailed understanding on the folding pathway of a protein of interest[23,300]. smFRET provided further insight of the folding of α1-AT, an archetypal member of the serpin superfamily that fold into a metastable conformation. The use of smFRET in studying the folding pathway of α1-AT allowed the detection of an additional folding intermediate that has not been detected by other experimental techniques[77]. This highlights the power of smFRET when studying the folding pathway of proteins.

Conserpin, consensus-designed serpin is extremely thermostable and aggregation-resistant compared to α1-AT. When studying the folding of conserpin, equilibrium unfolding depicts conserpin folds through a two-state mechanism. However, extrinsic fluorescence and kinetic folding detects folding intermediates that are poorly populated[150,312]. Therefore, smFRET will provide insight into the folding intermediates along conserpin's pathway. This will also allow a

comparison between conserpin and archetypal serpin α1-AT and the folding pathway of this unique protein family.

Studying α1-AT's folding pathway involved the introduction of two additional cysteine residues (to the single cysteine present, producing a pair of cysteines on each serpin molecule), and site-specific labeling with the cysteine residues being chemically modified with a FRET dye via the sulfur atom[301]. To compare the folding pathways between conserpin and α1-AT, corresponding cysteine residues were introduced onto conserpin. Conserpin contains one buried cysteine residue (C209) on strand 1 of β-sheet B (s1B), inaccessible to a maleimide dye. Therefore, this cysteine was replaced with a serine, and a cysteine was placed in a solvent exposed area that sits next to the corresponding solvent exposed cysteine in α1-AT (D205C). The removal of conserpin's buried cysteine was to ensure there was no disulfide bond formation with the newly introduced solved exposed cysteine (D205C).

The removal of conserpin's buried cysteine to a serine could potentially be problematic as the buried cysteine is in the B/C barrel. The B/C barrel of serpins is hypothesized to be the first region to fold[76,79,87]. By mutating conserpin's buried cysteine to a serine (C209S), the serpin's propensity to aggregate could increase, as mutations in the hydrophobic core increase aggregation propensity[107]. The cysteine amino acid has previously been observed to act as a hydrophobic amino acid, where the sulfhydryl group cannot form a hydrogen bond with the solvent[314,315]. By mutating the cysteine to a serine, despite both amino acids containing the same side chain length, the folding could be compromised. To ensure the double-cysteine conserpin variants remained soluble during recombinant expression, pre-and post- induction with samples were taken. Both variants remained soluble, ensuring that the cysteine-to- serine mutation did not affect the recombinant expression conserpin.

To ensure the introduction of cysteine residues onto conserpin did not affect its thermostability and inhibitory mechanism, the midpoint of thermal unfolding was determined by circular dichrosim, and the variants inhibitory mechanism studied by a stoichiometry of inhibition assays against trypsin. The midpoint of thermal unfolding for the variants is slightly lower than that of conserpin, but unfolding was only observed in the presence of denaturant (2M guanidine hydrochloride). Given that heating each sample to 100°C provided no significant change in elliptical signal, the double cysteine variants remain highly thermostable like conserpin. Furthermore, the inhibitory activity of the variants was not compromised. The increase in SI for the D25C variant is negligible, as conserpin is already a poor inhibitor of trypsin. Therefore, the

mutation of the buried cysteine to a serine, and introduction of three solvent exposed cysteine residues does not affect the conserpin molecule.

The investigation of the folding pathway of conserpin by smFRET is critical in understanding how thermostable serpins fold, as conserpin is not the only serpin to fold through a poorly populated intermediate. The folding pathway of an extremely thermostable serpin from archaeon *Pyrobaculum aerophilum* (optimum growth temperature 100°C), aeropin, seems to occur through a two-state mechanism, as determined with the use of fluorescence and circular dichroism equilibrium unfolding studies[277]. However, stopped-flow kinetics detected rapidly forming, poorly populated intermediates. Furthermore, the folding of another thermophilic serpin, tengpin, from bacteria *Thermoanaerobacter tengcongenesis* (optimal growth temperature 75°C), occurs as a two-state mechanism according to the same experimental techniques[156]. The formation of a poorly populated intermediate (present over a short denaturation concentration range and a short timeframe) in tengpin and aeropin may reflect the environment in which the serpin functions at. As the serpin's folding intermediate is aggregation prone and aggregation propensity increases with heat[79,115,311], these thermophilic serpins use a greatly increased rate of folding through this intermediate in their heated environments to avoid heat-induced aggregate[316]. Understanding how thermophilic serpins fold can provide insight in engineering mesophilic serpins to avoid this aggregation prone-folding intermediates.

5.4. Conclusion

In this study, smFRET capable conserpin variants were engineered to study its folding pathway. Two double-cysteine variants were produced to study the short and long axis of the serpin. These variants were successfully expressed, purified and tested for activity and thermal stability. Future smFRET studies using these proteins will provide insight into the folding pathway of conserpin and other naturally occurring thermophilic serpins, and comparing the folding pathways to observe the difference from mesophilic counterparts, such as α1-AT.

5.5. Supporting Information

```
                                                                25
Conserpin       MHHHHHHENLYFQGAASSHKLAEANTDFAFSLYRELAKSSPDKNIFFSPVSISSALAMLS
D25C            MHHHHHHENLYFQGAASSHKLAEANTDFAFSLYRELAKSSPCKNIFFSPVSISSALAMLS
S285C           MHHHHHHENLYFQGAASSHKLAEANTDFAFSLYRELAKSSPDKNIFFSPVSISSALAMLS
                ************************************** ******************

Conserpin       LGAKGDTHTQILEGLGFNSEADIHQGFQHLLQTLNRPKGLQLKTANGLFVDKSLKLLDSF
D25C            LGAKGDTHTQILEGLGFNSEADIHQGFQHLLQTLNRPKGLQLKTANGLFVDKSLKLLDSF
S285C           LGAKGDTHTQILEGLGFNSEADIHQGFQHLLQTLNRPKGLQLKTANGLFVDKSLKLLDSF
                ************************************************************

Conserpin       LEDSKKLYQAEAFSVDFDPEEAKKQINDWVEKQTNGKIKDLLKDLDSDTVLVLVNAIYFK
D25C            LEDSKKLYQAEAFSVDFDPEEAKKQINDWVEKQTNGKIKDLLKDLDSDTVLVLVNAIYFK
S285C           LEDSKKLYQAEAFSVDFDPEEAKKQINDWVEKQTNGKIKDLLKDLDSDTVLVLVNAIYFK
                ************************************************************

                                                        205  209
Conserpin       GKWKKPFDPENTKEEDFHVDEKTTVKVPMMSQKGKFYYYHDDELSCKVLELPYKGNASML
D25C            GKWKKPFDPENTKEEDFHVDEKTTVKVPMMSQKGKFYYYHDCELSSKVLELPYKGNASML
S285C           GKWKKPFDPENTKEEDFHVDEKTTVKVPMMSQKGKFYYYHDCELSSKVLELPYKGNASML
                ************************************** ***.**************

Conserpin       IILPDEGGLQHLEQSLTPETLSKWLKSLTRRSVELYLPKFKIEGTYDLKEVLSNLGITDL
D25C            IILPDEGGLQHLEQSLTPETLSKWLKSLTRRSVELYLPKFKIEGTYDLKEVLSNLGITDL
S285C           IILPDEGGLQHLEQSLTPETLSKWLKSLTRRSVELYLPKFKIEGTYDLKEVLSNLGITDL
                ************************************************************

                285
Conserpin       FSPGADLSGITEEKLYVSKAVHKAVLEVNEEGTEAAAATGVEIVPRSPPEFKADRPFLFL
D25C            FSPGADLSGITEEKLYVSKAVHKAVLEVNEEGTEAAAATGVEIVPRSPPEFKADRPFLFL
S285C           FCPGADLSGITEEKLYVSKAVHKAVLEVNEEGTEAAAATGVEIVPRSPPEFKADRPFLFL
                *.************************************************************

Conserpin       IRENKTGSILFMGKVVNP
D25C            IRENKTGSILFMGKVVNP
S285C           IRENKTGSILFMGKVVNP
                ******************
```

SI Figure 1. Sequence alignment of Conserpin and the two cysteine variants showing the incorporation of cysteine residues by site-directed mutagenesis.

Chapter 6: General Discussion

6.1. Overview

Serpins are one of the few protein families that do not fold into the thermodynamically, lowest energy conformation[56,57]. Instead, serpins fold into a metastable conformation that is critical for inhibitory function. This inhibitory mechanism is unique and viewed as a 'molecular mousetrap'; the binding and cleavage of the reactive centre loop (RCL) by the target protease triggers a large, rapid conformational change where the RCL inserts into the central β-sheet (β-sheet A)[63]. During this process, the protease is still covalently attached (prior to the deacylation step in the protease catalytic mechanism) and is translocated to the opposite pole of the serpin (71 Å). The translocation crushes the protease against the serpin body and the catalytic serine in the active site is displaced. The end result renders both the serpin and protease inactive[63].

The metastable state of serpins renders the proteins susceptible to misfolding and polymerization under various conditions, including mutations[84]. This misfolding and polymerization is the bases of many diseases termed serpinopathies. Serpinopathies lead to a loss-of-function disease, caused by low circulating levels of a particular serpin, and gain-of-toxicity from polymer formation at the site of serpin synthesis. One of the most-studied serpinopathies is α1-AT deficiency caused by the Z variant. The Z variant contains one amino acid substitution, glutamic acid-342 to lysine[136], resulting in a loss of a salt bridge with lysine-290, and charge repulsion between the two lysine residues[289]. This creates an open and highly dynamic breach region (top of β-sheet A/ base of the RCL), as detected with hydrogen-deuterium exchange and molecular dynamics (MD) simulations, and observed in the x-ray crystal structure of native Z α1-AT [78,137,138]. During folding, the folding intermediate folds off-pathway, leading to misfolding and aggregation at the site of α1-AT synthesis (hepatocytes)[129,132]. As a result, there are low circulating levels of α1-AT, leading to an imbalance between α1-AT and HNE in the lungs. As a consequence, early onset of emphysema occurs[127].

Currently, the only treatment for α1-AT deficiency is augmentation therapy, weekly intravenous injections of plasma purified α1-AT. This aims to restore the circulating levels within the lungs, and to correct the serpin: protease imbalance[126,133]. However, inactive, latent α1-AT has been detected in the purified sample[140], it is expensive and large quantities of purified serpin are required[141,142]. Other approaches to treating α1-AT have been investigated, all of which aim to reduce polymerization by targeting the open β-sheet A. These include chemical chaperones[143–145], small compounds[142] and intrabody treatment[148]. However, these approaches result in a non-functioning serpin. Therefore, another approach would be to engineer a serpin that mimics α1-AT

in its function but is also stable and does not convert into inactive serpin as easily as α1-AT. This would allow the use of recombinant DNA technology to produce large quantities of the engineered serpin while also being cost effective.

Protein therapeutics can be classified into two approaches: replacement therapeutics (replacing deficient or non-functioning proteins) or antagonist therapies (inhibiting a target proteins function)[317]. The replacement therapy approach, used as a treatment for α1-AT deficiency, first involved purifying the deficient/ dysfunctional protein from human and animal sources[149,318,319]. These proteins are often not sufficient to be therapeutics, as they are deficient in many important therapeutic properties, such as stability (e.g. latent α1-AT has been detected in the purified sample for augmentation therapy[140]), biological half-life and immunogenicity risk[317]. The rise of recombinant DNA technology has led to the production of protein therapeutics with several benefits over source-derived proteins. These benefits include larger production of a given protein therapeutic and cost efficacy, reduction of possible exposure to human or animal diseases, and the ability to modify the protein to improve its function[149].

The use of recombinant DNA technology to produce a protein therapeutic allows for protein engineering to modify a native protein for ideal and customized therapeutic qualities. One important quality is the increase in serum half-life. Increasing the serum half-life can lead to greater target localization and efficacy, along with longer intervals between doses[320]. Protein engineering can involve three principle strategies: linking the therapeutic protein to another protein or polymer through genetic or chemical linkage, or introduction of mutations (one or several) or deletions to the primary sequence of the protein[318].

One of the best-known protein therapeutics that was developed through protein engineering is the production of recombinant insulin. Insulin previously obtained from the bovine and porcine pancreas was cost-effective and produces an immunological response[321]. The use of protein engineering lead to two variations of insulin that could be used, a short acting[322–324] and long-acting[325–328]. Therefore, protein engineering is essential in the development of protein therapeutics.

Protein engineering can not only be used to improve a native protein but can also create a completely synthetic protein. One of these approaches is consensus-engineering[252]. Consensus-engineering was used to produce a synthetic serpin, conserpin, that is extremely thermostable and aggregation resistant, while also being functional as an inhibitor[150]. The thermal stability of conserpin can be attributed to various regions within the body that contains favourable interactions.

The thermostability, aggregation resistance and ability to remain robust with mutations (Chapter 5) of conserpin makes it a great candidate as a starting molecule for engineering a serpin with a high stability, while retaining function. In this thesis, conserpin was used to engineer function (Chapter 3) and stability (Chapter 4), in an attempt to create a serpin that mimics α1-AT, yet remains stable, a desired property for a protein therapeutic. Furthermore, the folding pathway of conserpin was also studied using single-molecule Förster Resonance Energy Transfer (smFRET) (Chapter 5), to understand how it is aggregation resistant during folding.

6.2. Engineering serpins for function

There are a number of families of inhibitors that target serine proteases. Some of the most well-studied are the Kazal, Kunitz, Bowman-Birk and serpins[329]. What is common to these inhibitors is the reactive site, an exposed loop (RCL, in serpins) that is stabilized in a canonical conformation through intramolecular interactions between the loop and the core of the inhibitor[330,331]. This exposed loop acts similarly to a substrate, binding to the active site of the protease[331,332]. It is an exposed loop that contains the sequences that dictates specificity[333], with residue P1 hypothesized to be the primary determinant of specificity[330]. Mutations in P1 can change the specificity, as observed in the soybean trypsin inhibitor (Figure 1). The mutation of P1 arginine to tryptophan changed the specificity of the inhibitor from trypsin to chymotrypsin[334]. Moreover, the P1 residue also contributes to a serpins specificity; mutating P1 residue of archetypal serpin α1-AT from methionine to arginine converts specificity from human neutrophil elastase (HNE) to thrombin[184]. Exchanging residues within the loop with another inhibitor has also proven successful in changing specificity. Residues from α1-AT's RCL was introduced onto sunflower-trypsin inhibitor, changing the specificity from trypsin to kallikrein-related peptidase 5 and 7 (KLK)[335].

The ability to mutate the reactive site of a protease inhibitor to change specificity makes these proteins good candidates to target proteases involved in disease[336]. Successful inhibitors produced as therapeutics include inhibition of the angiotensin converting enzyme for hypertension treatment[337,338] and inhibitors of HIV aspartyl protease inhibitors to prevent development of AIDS[339,340]. However, one conundrum when engineering specific inhibitors is the possibility of off-target effects. Serine proteases often have near-identical geometry in the active site, making it difficult for the inhibitor to distinguish between closely related serine proteases. An explanation for this is that there may not be enough contact between the protease and the inhibitor for the inhibitor to distinguish which protease to target[341]. This is evident in an attempt to engineer an

inhibitor to mesotrypin (PRSS3), which is upregulated in promoting tumor progression in cancers[342,343]. Engineering two Kunitz- domain proteins, bovine pancreatic trypsin inhibitor and the Kunitz-domain of amyloid precursor protein, led to an increase in specificity and inhibition against mesotrypsin[342,343]. However, there were off-target inhibition against other trypsins[342,343]. Therefore, a large inhibitor, such as a serpin, may be more beneficial in engineering an inhibitor to a target protease due to the potential of producing more interactions between the inhibitor and protease.

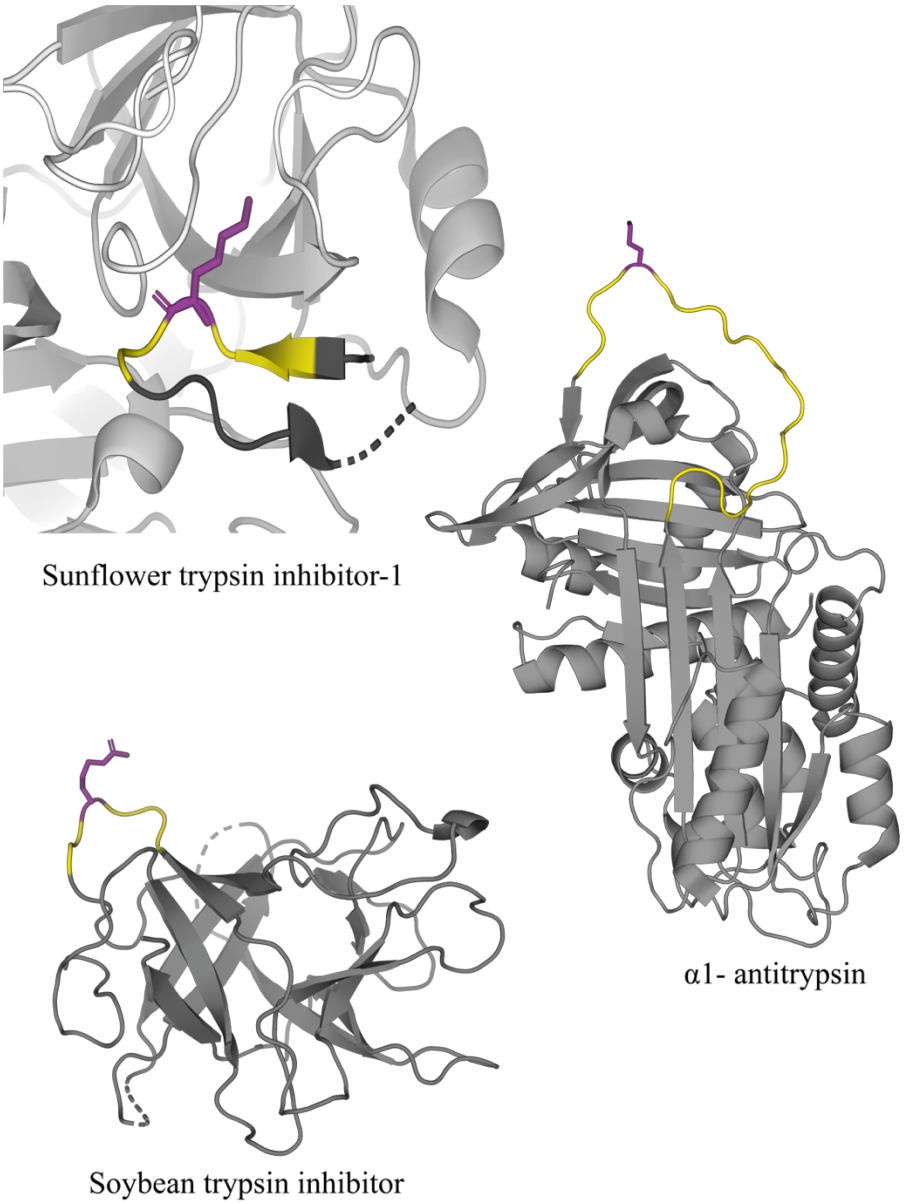


Figure 1. Different protease inhibitors. Three types of protease inhibitors: Bowman-Birk (e.g. sunflower trypsin inhibitor-1 (in the active site of KLK4), PDB ID: 4K8Y[344]), Kunitz (e.g. soybean trypsin inhibitor, PDB ID: 1AVU[345]) and serpin (e.g. α1-AT, PDB ID: 3NE4[197]). Each inhibitor has a reactive loop (*yellow*), containing a P1 residue (*purple*) that is targeted by a protease.

As discussed above, the current treatment for α1-AT deficiency is augmentation therapy to restore the serpin: protease balance in the lungs[126,133]. However, inactive α1-AT arises and this therapeutic is not cost effective. Therefore, the development of serpin with specificity towards HNE is needed, while remaining stable in the native conformation and reducing its ability to become inactive. Engineering a serpin for specificity is more difficult than other inhibitors. This is due to the difference in the inhibitory mechanism of the inhibitors. Kazal, Kunitz and Bowman-Birk inhibitors bind within the active site in a tight mechanism (as a lock- and key or induced fit mechanism)[334,341]. The P1-P1' bond is hydrolyzed slowly, but the products are not released, possibly allowing the amine bond to re-ligate[329]. Serpins, however, differ in that the cleavage of the P1-P1' peptide bond triggers a large conformational change, where the protease is crushed against the body of the serpin, and the catalytic serine is displaced out of the active site[63]. Therefore, mutations in the RCL to change specificity must be compatible with the residues in β-sheet A in order to inhibit still the target protease.

The P1 residue within the RCL is hypothesized to play a role in determining the specificity of a serpin. This is most prominent in the naturally occurring mutation of α1-AT, where P1 methionine becomes arginine. This changes specificity from human neutrophil elastase (HNE) to thrombin, resulting in thrombosis[184]. The rate of which α1-AT (P1 Arg) inhibits thrombin is greater than antithrombin-III (thrombin's serpin inhibitor) without co-factor heparin. Therefore, it is hypothesized that the specificity of a serpin could be changed by mutating residues within its RCL. Multiple studies have attempted to change a serpins specificity, with many creating an α1-antichymotrypsin (ACT)/ α1-AT chimera that targets HNE through mutating various residues within the RCL. ACT has a sequence identity of 45% to α1-AT, one of the highest sequence identities in the serpin family[346]. This makes ACT an appropriate starting serpin to change specificity. ACT contains a hydrophobic leucine residue at P1. HNE, which favours hydrophobic residues, recognises ACT's P1 but results in rapid cleavage by HNE, with ACT being a substrate rather than an inhibitor of HNE[191]. Mutating ACT's P1 to methionine did convert specificity towards HNE, but to a lesser extent than α1-AT (SI of 5 and a rate of inhibition on an order of three magnitudes slower). Therefore, to understand what dictates a serpins specificity, many of these studies have created ACT/ α1-AT chimeras with different amounts of mutated residues.

It is hypothesized that residues P6-P3' of the RCL may interact with HNE upon protease binding to the RCL. Therefore, ACT/α1-AT chimeras were developed that cover these residues in an attempt to increase the chimera's inhibitory rate against HNE. Some of the chimeras produced

include ACT/α1-AT P6-P3', P4-P3', P3-P3' and a full loop swap (P10-P4'). The most effective chimera was the P3-P3', with an SI of 1.4 and a rate of inhibition of 10^5 M^{-1} s^{-1} [180]. However, despite this success, this chimera inhibits HNE with a slightly higher SI than α1-AT (SI of 1:1 with HNE), and a rate of inhibition on an order of two magnitudes slower. Furthermore, increasing the number of mutated residues in the ACT/α1-AT chimeras leads to a decrease in rate of inhibition (as observed in P6-P3' chimera[191]), with a full RCL swap from α1-AT onto ACT resulting in rate of inhibition four orders of magnitude less than α1-AT (10^3) and an SI of 111:1[190]. Therefore, it was hypothesized that other factors must be responsible in determining specificity, including secondary binding sites on the serpin, and possibly favourable interactions between the two proteins.

The engineering of conserpin has led to a serpin with a higher sequence identity to α1-AT than ACT (59% compared to 45%). This makes conserpin a more attractive serpin, along with its high thermostability, to engineer to target HNE. In this thesis (Chapter 3), a total of nine residues of α1-AT's RCL (P7-P2') were mutated onto conserpin. This chimera covers the residues that are hypothesized to interact with HNE[191], along with the addition of a P2' residue, isoleucine (absent in conserpin). The resulting chimera was noted as conserpin-AAT_{RCL}.

Mutations in the RCL of conserpin-AAT_{RCL} did not affect the thermal stability of the serpin, remaining as thermal stable as conserpin (72°C in 2M GndHCl). This is not surprising, as mutations in exposed loops generally have a small effect on thermal stability[330]. However, similarly to the ACT/ α1-AT chimeras, introducing the RCL sequence of α1-AT did not change the specificity of conserpin-AAT_{RCL}. α1-AT inhibits trypsin and HNE at a stoichiometry ratio of 1:1 (serpin: protease), while conserpin inhibits only trypsin at a higher ratio (1.8:1). The introduction of α1-AT's RCL does not improve this ratio against trypsin (1.46:1), while there was no inhibition observed while attempting to calculate the SI against HNE. Understanding the reaction between HNE and conserpin-AAT_{RCL} could only be determined when observing any potential serpin: protease complex on SDS-PAGE. Only a minute amount of conserpin-AAT_{RCL} formed a serpin: protease complex with HNE, with a majority being cleaved, indicating that conserpin-AAT_{RCL} was undergoing the substrate pathway more than the inhibitory pathway. This is consistent with previous chimeras[180]. It should be noted that mutations in the RCL does not prevent the RCL from insertion into β-sheet A, as conserpin-AAT_{RCL} inhibits trypsin.

The determination of a serpin chimera's structure by x-ray crystallography could possibly provide an explanation of why serpin chimeras fail in inhibiting a new target protease. The x-ray crystal structure of an ACT/ α1-AT P3-P3' chimera revealed the RCL formed a helical conformation, possibly affecting its ability to inhibit HNE to the rate of α1-AT[201]. Therefore, the x-ray structure of native conserpin-AAT_{RCL} was determined (PDB ID: 6EE5). Unfortunately, due to its flexibility, electron density was not present for the RCL, so the structure of the RCL could not be observed. Nevertheless, the x-ray crystal structure provided information on the surface electrostatics of conserpin-AAT_{RCL}. Given that there needs to be an encounter between the serpin and protease to form a stable serpin: protease Michaelis-Menton complex, there must be favourable interactions between the two proteins. This hypothesis was previously stated, where the inability of the ACT/ α1-AT chimeras to inhibit HNE could be a result of the absence of appropriate body-body interactions, or the presence of unfavourable interactions[191]. The surface electrostatics between conserpin-AAT_{RCL} and α1-AT differs, in that conserpin-AAT_{RCL} contains a large positive potential underneath and surrounding the RCL (Chapter 3, Figure 2). Furthermore, modeling a serpin: protease complex of conserpin-AAT_{RCL} with trypsin and HNE strengthen this hypothesis of the importance of surface electrostatics in a serpin: protease complex. Conserpin-AAT_{RCL} has better surface electrostatic complementarity to trypsin than to HNE, providing an explanation on why conserpin-AAT_{RCL} inhibits trypsin but not HNE. The hypothesis that the surface potential between a serpin and protease must be complemented is further strengthened through observing the difference in inhibition rates of serpin plasminogen activator inhibitor-1 (PAI-1) and its two targets, tissue-type plasminogen activator (tPA) and urokinase-type plasminogen activator (uPA). PAI-1 inhibits tPA at a faster rate than uPA[207,208]. An explanation for this is that tPA and PAI-1 have more complementary electrostatic interactions than uPA. Therefore, this confirms what was previously stated, there must be favourable interactions between the serpin and protease to allow for a stable complex formation and protease inhibition.

Another possible explanation for conserpin-AAT_{RCL} failure to inhibit HNE is the dynamics of the RCL. Just as for small inhibitors, the RCL must be rigid enough to bind into the active site of the protease[334,341], yet dynamics is needed for the RCL to insert into β-sheet A in a rapid movement[93]. The rate of which the RCL inserts into β-sheet A must be faster than the deacylation step of the proteases catalytic mechanism. The rate at which HNE catalyzes the P1-P1' peptide bond is faster than loop insertion, while chymotrypsin's catalysis is on par with loop insertion. This is why α1-AT inhibits HNE on an order of 10^7 M^1 s^1 [180]. Through modeling the RCL into the x-ray crystal structure of conserpin-$AAT_{RCL,}$ and subsequent performance of molecular dynamics (MD)

simulations, the dynamics of the RCL can be analyzed. This study is the first to use molecular dynamics (MD) simulations in understanding how the dynamics of the RCL could dictate a serpin's specificity. The hinge region conserpin-AAT$_{RCL}$'s RCL, residues P12-P9, are the first to insert into the breach region of β-sheet A, with simulations showing a high degree of flexibility compared to α1-AT. The low flexibility of the hinge region of α1-AT allows it to be primed into position for insertion, that is, it is poised between strands 3 and 5 of β-sheet A (s3A and s5A). Conserpin-AAT$_{RCL}$ contains a mostly extended hinge region, away from β-sheet A. This hinge region, however, does sample one primed conformation, which may explain why there is inhibition of trypsin. Therefore, the conformation and dynamics of the RCL can provide an explanation as to why chimeric serpin's fail at transferring specificity.

To conclude, Chapter 3 looked at engineering the specificity of α1-AT onto conserpin. Despite conserpin having a high sequence identity to α1-AT, specificity to HNE could not be transferred onto conserpin. The determination of the x-ray crystal structure of native conserpin-AAT$_{RCL}$ aided in providing two possible reasons for this failure: incompatible surface electrostatics between conserpin-AAT$_{RCL}$ and HNE, and/or the dynamics of the RCL. Taken together, we have confirmed a previous hypothesis that compatible body-body interactions are necessary, while also adding that the dynamics of the RCL could also play a role in determining specificity of a serpin.

6.3. Engineering serpin stability

One important property that a protein therapeutic needs is high stability. Increasing the stability can improve the expression yields during manufacturing, increase half-life (shelf and biological retention time), and become more robust under difficult conditions (change in temperature, pH and solution conditions)[317,319,347,348]. The stability of a protein therapeutic is important as it is critical for large-scale production. For a protein therapeutic to be effective, it must retain its physical and chemical nature (remain potent) during both production and storage, minimizing any potential for aggregation or degradation. The protein must retain its folded conformation, as partial unfolding can lead to aggregation, potentially eliciting an immune response in a patient[349]. Currently, this is a problem with purified α1-AT for augmentation therapy, where latent, inactive α1-AT has been detected post processing and storage[140].

Increasing the stability can be obtained through introducing various intramolecular interactions, including salt bridges and disulfide bonds[38,40,41,47,255–258], and improving the hydrophobic packing

in the core of the protein[29,33,259,260]. Introducing these interactions to stabilize α1-AT could create a stable α1-AT variant that could possibly be a replacement for the current augmentation therapy. Engineering α1-AT for stability is a delicate task, as introducing mutations could possibly lower the kinetic barrier between the active, native conformation and the inactive, latent conformations. If this occurs during folding of α1-AT could force the serpin to bypass the native conformation and fold into the more stable, latent conformation.

There have been many attempts in stabilizing α1-AT. Many of these studies involved random single-point mutations throughout the serpin conformation[66,68,69,94,211,281]. Many of the random mutations successfully increased the stability of α1-AT (as determined by either calculating the Gibbs free energy, ΔG, or loss of activity at a set temperature over time), but with mixed results on what effect this has on its function. Random mutations found in regions that are important for a serpins function (e.g. β-sheet A) often correlated with activity loss, while mutations in other regions, such as the hydrophobic core, had no effect on function. Therefore, it was stated that introducing stabilizing mutations in functional regions affect the inhibitory activity of the serpin[68,69,213,281]. The single-mutations within the hydrophobic core were combined to produce a stable α1-AT, noted as Multi7[211]. This Multi7 α1-AT is as stable as non-inhibitory serpin ovalbumin ($\Delta\Delta G_{WT\text{-}Mutant}$ of 8 kcal mol-[1]) with no compromise on function (SI of 1.1 for human elastase). It has not been stated in the literature if the Multi7 is undergoing clinical trials as a replacement for augmentation therapy. Unlike only targeting residues in the hydrophobic core like in Multi7 α1-AT, is it possible to use rational design and to target regions of a serpin that are central to its inhibitory mechanism (e.g. β-sheet A) to increase stability without compromising function?

Chapter 4 investigates increasing the stability of α1-AT by using rational design to introduce clusters of amino acids from a thermostable serpin, conserpin, onto α1-AT. Conserpin was selected with an identical reasoning as in Chapter 3, it has a high sequence identity to α1-AT (59%), while also being aggregation resistant and extremely thermostable (*Tm* 100°C+). Importantly, conserpin is also functional as an inhibitor against trypsin[150]. What contributes to conserpin's high thermostability is various regions that contain favourable polar and non-polar interactions, an increased number of salt bridges (compared to α1-AT) and improved packing in the hydrophobic core[150]. Each of these regions were identified by analysing the x-ray crystal structure of native conserpin (PDB ID: 5CDX[150]). As random mutagenesis can introduce amino acids in regions where they may not fit (e.g. a large, hydrophobic side chain in a tightly packed region), rational

design allows for modelling amino acids into appropriate regions. This approach was critical for grafting clusters from conserpin onto α1-AT. While the residues selected for grafting may be favourable in the conserpin structure, not all residues could be grafted onto α1-AT. A few residues were unselected due to large side chains that could not be accommodated in tight regions. Nevertheless, a total of eight grafts were produced, varying from different regions of the hydrophobic core and B/C barrel, to Helices F and H, and the breach region.

Out of eight grafts, two fold into an inactive conformation, three are less thermostable and three are more thermostable than WT α1-AT. These results also determined that different regions of the hydrophobic core and B/C barrel contribute to thermostability. A previous study that mutated stabilizing residues of α1-AT onto ACT and antithrombin-III observed that while some of α1-AT's stabilizing mutations did stabilize ACT and antithrombin-III, other mutations decreased the stability of these two serpins[213]. The authors concluded that each serpin contains its own stabilizing strain. The results from this grafting study support this study, as certain residues may contribute to conserpin's thermostability, but grafting these residues onto α1-AT results in a decrease in stability.

While previous studies state that mutations in the hydrophobic core increase the thermal stability[213], Chapter 4 shows that there are different regions in the core that contribute to stability. The F51 graft, containing mutations on β-sheet A and Helix-B, increased the thermal stability of α1-AT, while the T59S graft (mutations on Helix-B and Helix-A, facing away from β-sheet A) has a decrease the thermal stability. Therefore, introducing modelled favourable mutations in the hydrophobic core does not necessarily lead to an increase in stability.

The F51 graft in the hydrophobic core increased the thermostability while not compromising function (SI of 1.1, identical to that of WT α1-AT). This is in good agreement with the successful single-point mutations in the hydrophobic core, and the Multi7 α1-AT [211,213]. Furthermore, previous studies with stabilizing mutations in functional regions (e.g. β-sheet A) led to a decrease in inhibitory function [64,68,69,212,213,270]. Yet, introducing a cluster of mutations into one of these critical regions, as in the Breach graft (increased number of salt bridges) did not compromise inhibitory function of this graft (SI of 1.1:1, identical to WT α1-AT). The increased number of salt bridges in the breach continued to allow β-sheet A to open and accept the RCL. While two of the thermostable grafts did not have a compromise in function, the Helix-F graft has an increase in SI (1.33). This was expected, as Helix-F lies in front of β-sheet A and must shift to allow RCL

insertion. The aim of the Helix-F graft was to rigidify the helix, possibly slowing down its ability to shift away from β-sheet A, increasing the thermostability. This result shows the benefit of using ration design over random mutagenesis when engineering stability without compromising function.

The increase in stability can increase the yields during large-scale protein production[350]. While this may be true for some proteins, this is not the case for serpins. Each of the three thermostable grafts engineered, along with the previous α1-AT stabilizing mutations (including the Multi7[211]) involved purifying recombinant α1-AT variants from *E. coli* inclusion bodies. Nevertheless, each of the thermostable grafts yielded more monomeric protein than WT α1-AT post-refold. Furthermore, a higher concentration of chemical denaturant (guanidinium hydrochloride, GndHCl) was required to unfold and force each graft to populate the folding intermediate. A comparison between these thermostable grafts and the previous stabilized α1-AT mutations cannot be performed, as the other studies did not know what effect these mutations have on the folding. What is now known is introducing these three grafts (F51, Breach and Helix-F) onto α1-AT has led to variants that require a higher denaturant concentration to force it to populate the folding intermediate and increasing the yield of monomeric protein post refold.

The only α1-AT variant that confers high stability is the Multi7, and as discussed earlier, this stability is a result of mutations in the hydrophobic core[211] (Figure 2). As stated above, functional regions were targeted for mutation with no significant effect on function. Would it be possible to combine the three thermostable grafts (F51, Breach and Helix-F) together to increase the stability of α1-AT even more? The resulting graft was termed 3stable (Figure 2). By engineering 3stable, we increased the thermal stability by 14°C compared to WT α1-AT (T_m of 73°C). This is the most thermostable (T_m) α1-AT variant, and is more thermostable than thermophilic serpin thermopin (*Tm* 65°C[276]). This thermostability correlates with an increase in Gibbs free energy ($\Delta\Delta G_{WT\text{-}Mutant}$ of 8.66 kcal mol-[1]), on par with the Multi7 α1-AT (8 kcal mol^{-1}). The increase of thermostability did not significantly compromise function (SI 1.28), concluding 3stable is still a decent inhibitor against HNE. In comparison to Multi7 α1-AT, this is a higher SI (Multi7 SI of 1.1), but here two functional regions were targeted, both of which have previously been targeted for mutagenesis and resulting in a decrease in inhibitory activity [64,69]. Furthermore, like the Multi7 α1-AT and the three thermally stable grafts, the 3stable is expressed in inclusion bodies, yet yields significantly more monomeric protein post-refold compared to WT α1-AT and the three thermally stable grafts (over 60% yield, while WT α1-AT yields 11%[150]). Lastly, the folding intermediate for the 3stable

graft was populated at a higher chemical denaturant concentration compared to all the thermal stable grafts and WT α1-AT, indicating 3stable more stable than WT α1-AT.

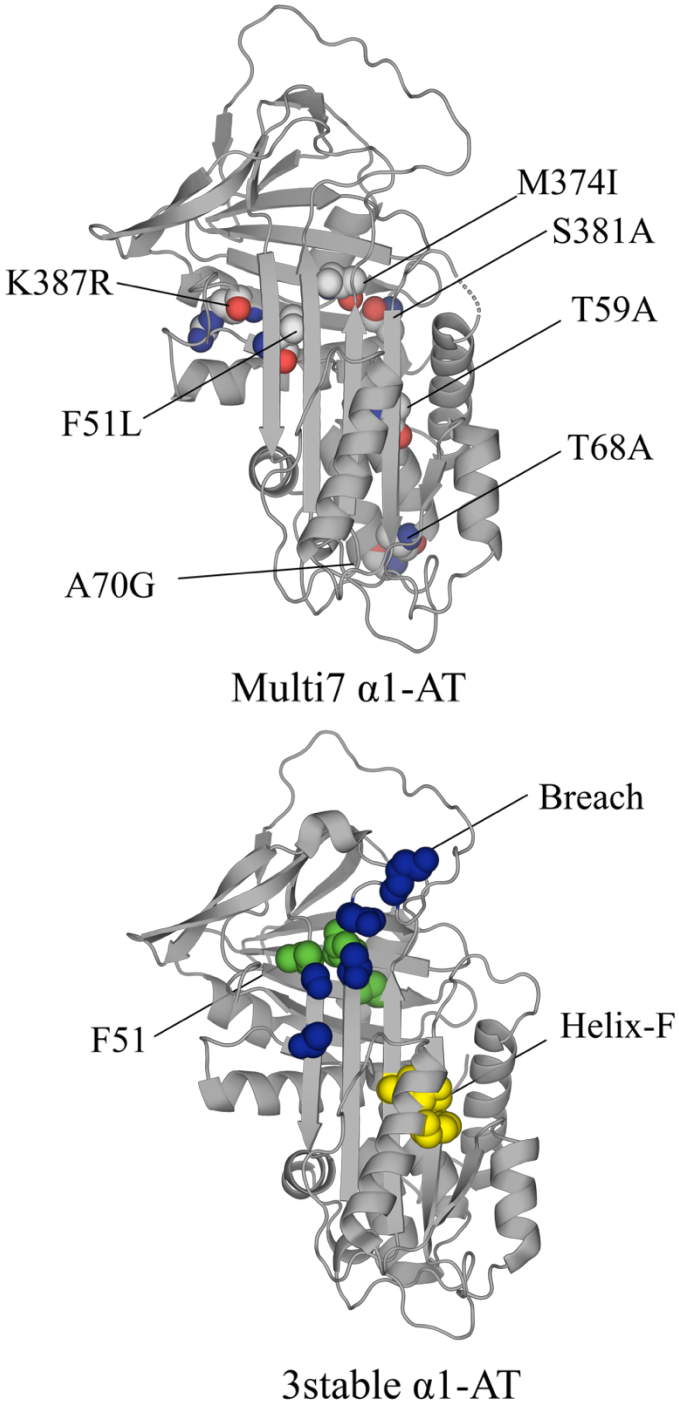


Figure 2. The two stable α1-AT variants. The Multi7 α1-AT (PDB ID: 5NBV, *unpublished*) (left) contains mutations within the hydrophobic core (*spheres*), while the 3stable α1-AT (modelled onto PDB ID: 3NE4[197]) (right) contains mutations in the core (*green*), but also the breach (*blue*) and Helix-F (*yellow*) regions.

Taken together, Chapter 4 explores the use of rational design and a thermal stable serpin to engineer thermostability onto α1-AT. Unlike the previous α1-AT stabilizing studies, rational

design successfully introduced mutations in functional regions without significantly compromising the serpins function. Furthermore, combining 3 thermally stable grafts together produced an α1-AT variant that can rival the Multi7 α1-AT variant. This chapter is the beginning of developing an α1-AT variant that could possibly replace purified α1-AT for augmentation therapy. Lastly, the success of this engineering is the beginning of possibly using the same process for engineering other serpins, such as antithrombin-III.

6.4. Conclusion

In conclusion, the work presented in this thesis focused on using conserpin, a synthetic thermostable serpin, to engineer function and stability, that may be able to replace plasma-purified α1-AT as a treatment for augmentation therapy. Studying the folding pathway of conserpin using smFRET also began, concluding that conserpin retains activity and thermostability in the presence of mutations. While it was unsuccessful in converting conserpin into an inhibitor of HNE (Chapter 3), two possible explanations were derived for this failure: the surface electrostatics between the serpin and protease must be compatible to allow for protease binding and inhibition, and the dynamics of the RCL is important for rapid insertion before the deacylation mechanism of the protease's catalytic mechanism. Furthermore, it was shown that engineering stability by using rational design and a thermostable serpin can be successful when targeting functional regions of the serpin (Chapter 4). The substantial increase in thermostability did not significantly compromise function, while improving the yield obtained post refold. Combining the thermostable grafts together produced an α1-AT variant that is more stable than a naturally occurring thermophilic serpin, all while retaining close to WT-like function. Therefore, the success of increasing the stability of α1-AT is the beginning of developing a recombinant variant of α1-AT that could possibly replace purified α1-AT sample. This technique could also be used for other serpins.

References

References

(1) Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci. 25*, 331–9.

(2) Rose, G. D., Fleming, P. J., Banavar, J. R., and Maritan, A. (2006) A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. 103*, 16623–16633.

(3) Sela, M., White, F. H., and Anfinsen, C. B. (1957) Reductive Cleavage of Disulfide Bridges in Ribonuclease. *Science (80-.). 125*, 691–692.

(4) Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A. 47*, 1309–1314.

(5) Anfinsen, C. B. ., and Scheraga, H. A. (1975) Experimental and Theoretical Aspects of Protein Folding. *Adv. Protein Chem.*

(6) Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science (80-.). 181*, 223–230.

(7) Baker, D., and Agard, D. a. (1994) Kinetics versus thermodynamics in protein folding. *Biochemistry 33*, 7505–9.

(8) Dill, K. a, and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol. 4*, 10–19.

(9) Makhatadze, G. I., and Privalov, P. L. (1995) Energetics of Protein Structure. *Adv. Protein Chem. Volume 47*, 307–425.

(10) Lattman, E. E., and Rose, G. D. (1993) Protein folding--what's the question? *Proc. Natl. Acad. Sci. 90*, 439–441.

(11) Fersht, A. (1999) Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. *Lavoisier.Fr*.

(12) Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol. 224*, 771–82.

(13) Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys. Physico-Chimie Biol. 65*, 44–45.

(14) Levinthal, C. (1969) How to fold graciously. *Mössbauer Spectrosc. Biol. Syst. Proc. 24*, 22–24.

(15) Šali, A., Shakhnovich, E., and Karplus, M. (1994) Kinetics of Protein Folding. *J. Mol. Biol. 235*, 1614–1636.

(16) Fersht, A. R. (1995) Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol. 5*, 79–84.

(17) Englander, S. W. (2000) Protein Folding Intermediates and Pathways Studied by Hydrogen Exchange. *Annu. Rev. Biophys. Biomol. Struct 29*, 213–238.

(18) Brockwell, D. J., and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol. 17*, 30–37.

(19) Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A. 89*, 8721–8725.

(20) Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol. 15*, 3–16.

(21) Naeem, A., and Fazili, N. A. (2011) Defective Protein Folding and Aggregation as the Basis of Neurodegenerative Diseases: The Darker Aspect of Proteins. *Cell Biochem. Biophys. 61*, 237–250.

(22) Dill, K. a, and MacCallum, J. L. (2012) The protein-folding problem, 50 years on. *Science 338*, 1042–1046.

(23) Bartlett, A. I., and Radford, S. E. (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol. 16*, 582–588.

(24) Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct. Funct. Bioinforma. 21*, 167–195.

(25) Sanchez-Ruiz, J. M. (2010) Protein kinetic stability. *Biophys. Chem. 148*, 1–15.

(26) Tokuriki, N., and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol. 19*, 596–604.

(27) Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem. 14*, 1–63.

(28) Tanford, C. (1997) How protein chemists learned about the hydrophobic factor. *Protein Sci. 6*, 1358–1366.

(29) Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M. N., Iimura, S., Gajiwala, K., Scholtz, J. M., and Grimsley, G. R. (2011) Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol. 408*, 514–528.

(30) Nick Pace, C., Scholtz, J. M., and Grimsley, G. R. (2014) Forces stabilizing proteins. *FEBS Lett. 588*, 2177–2184.

(31) Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A., and Grimsley, G. R. (2014) Contribution of hydrogen bonds to protein stability. *Protein Sci. 23*, 652–661.

(32) Fersht, A. R., and Serrano, L. (1993) Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol. 3*, 75–83.

(33) Pace, C. N. (1992) Contribution of the Hydrophobic Effect to Globular Protein Stability 29–35.

(34) Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J. 10*, 75–83.

(35) Shirley, B. A., Stanssens, P., Hahn, U., and Pace, C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry 31*, 725–732.

(36) Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry 29*, 2403–2408.

(37) Fersht, A. R. (1972) Conformational equilibria in a- and d-chymotrypsin. The energetics and importance of the salt bridge. *J. Mol. Biol. 64*, 497–509.

(38) Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol. 216*, 1031–1044.

(39) Roth, C. M., Neal, B. L., and Lenhoff, A. M. (1996) Van der Waals interactions involving proteins. *Biophys. J. 70*, 977–987.

(40) Clarke, J., and Fersht, A. R. (1993) Engineered Disulfide Bonds as Probes of the Folding Pathway of Barnase: Increasing the Stability of Proteins against the Rate of Denaturation. *Biochemistry 32*, 4322–4329.

(41) Matsumura, M., Signor, G., and Matthews, B. W. (1989) Substantial increase of protein stability by multiple disulphide bonds. *Nature 342*, 291–3.

(42) c. Nick, P. (1990) Measuring and increasing protein stability. *Tibtech 8*, 93–98.

(43) Lattman, E. E., and Rose, G. D. (1993) Protein folding--what's the question? *Proc. Natl. Acad. Sci. U. S. A. 90*, 439–441.

(44) Cabrita, L. D., and Bottomley, S. P. (2004) How do proteins avoid becoming too stable? Biophysical studies into metastable proteins. *Eur. Biophys. J. 33*, 83–88.

(45) Baker, D., Sohl, J. L., and Agard, D. A. (1992) A protein-folding reaction under kinetic control. *Nature*.

(46) Carr, C. M., Chaudhry, C., and Kim, P. S. (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proc. Natl. Acad. Sci. U. S. A. 94*, 14306–14313.

(47) Clark, A. C., Sinclair, J. F., and Baldwin, T. O. (1993) Folding of bacterial luciferase involves a non-native heterodimeric intermediate in equilibrium with the native enzyme and the unfolded subunits. *J. Biol. Chem. 268*, 10773–10779.

(48) Whisstock, J. C., and Bottomley, S. P. (2006) Molecular gymnastics: serpin structure, folding and misfolding. *Curr. Opin. Struct. Biol. 16*, 761–768.

(49) Law, R. H. P., Zhang, Q., McGowan, S., Buckle, A. M., Silverman, G. a, Wong, W., Rosado, C. J., Langendorf, C. G., Pike, R. N., Bird, P. I., and Whisstock, J. C. (2006) An overview of the serpin superfamily. *Genome Biol. 7*, 216.

(50) Silverman, G. A., Whisstock, J. C., Bottomley, S. P., Huntington, J. A., Kaiserman, D., Luke, C. J., Pak, S. C., Reichhart, J., and Bird, P. I. (2010) Serpins Flex Their Muscle. *J. Biol. Chem. 285*, 24299–24305.

(51) Hunt, L. T., and Dayhoff, M. O. (1980) A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha1-proteinase inhibitor. *Top. Catal. 95*, 864–871.

(52) Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G. W., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J., and Whisstock, J. C. (2001) The Serpins Are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins. *J. Biol. Chem. 276*, 33293–33296.

(53) Zhou, A., Wei, Z., Read, R. J., and Carrell, R. W. (2006) Structural mechanism for the carriage and release of thyroxine in the blood. *Proc. Natl. Acad. Sci. 103*, 13321–13326.

(54) Liu, T., Pemberton, P. A., and Robertson, A. D. (1999) Three-state unfolding and self-association of maspin, a tumor-suppressing serpin. *J. Biol. Chem. 274*, 29628–32.

(55) Ito, S., and Nagata, K. (2017) Biology of Hsp47 (Serpin H1), a collagen-specific molecular chaperone. *Semin. Cell Dev. Biol. 62*, 142–151.

(56) Tsutsui, Y., and Wintrode, P. L. (2007) Cooperative Unfolding of a Metastable Serpin to a Molten Globule Suggests a Link Between Functional and Folding Energy Landscapes. *J. Mol. Biol. 371*, 245–255.

(57) Irving, J. A., Haq, I., Dickens, J. A., Faull, S. V., and Lomas, D. A. (2014) Altered native stability is the dominant basis for susceptibility of α 1 -antitrypsin mutants to polymerization. *Biochem. J. 460*, 103–119.

(58) Loebermann, H., Tokuoka, R., Deisenhofer, J., and Huber, R. (1984) Human α1-proteinase inhibitor. *J. Mol. Biol. 177*, 531–557.

(59) Langdown, J., Belzar, K. J., Savory, W. J., Baglin, T. P., and Huntington, J. A. (2009) The Critical Role of Hinge-Region Expulsion in the Induced-Fit Heparin Binding Mechanism of Antithrombin. *J. Mol. Biol. 386*, 1278–1289.

(60) Yamasaki, M., Sendall, T. J., Harris, L. E., Lewis, G. M. W., and Huntington, J. A. (2010) Loop-sheet mechanism of serpin polymerization tested by reactive center loop mutations. *J. Biol. Chem. 285*, 30752–30758.

(61) Engh, R. A., Huber, R., Bode, W., and Schulze, A. J. (1995) Divining the serpin inhibition mechanism: a suicide substrate “springe”? *Trends Biotechnol. 13*, 503–510.

(62) Janciauskiene, S. (2001) Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles. *Biochim. Biophys. Acta - Mol. Basis Dis. 1535*, 221–235.

(63) Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature 407*, 923–926.

(64) Cabrita, L. D., Dai, W., and Bottomley, S. P. (2004) Different conformational changes within the F-helix occur during serpin folding, polymerization, and proteinase inhibition. *Biochemistry 43*, 9834–9839.

(65) Knaupp, A. S., and Bottomley, S. P. (2011) Structural change in β-sheet A of Z α1-antitrypsin is responsible for accelerated polymerization and disease. *J. Mol. Biol. 413*, 888–98.

(66) Lee, K. N., Park, S. D., and Yu, M.-H. (1996) Probing the native strain in α1-antitrypsin. *Nat. Struct. Mol. Biol. 3*, 497–500.

(67) Ryu, S., Choi, H., Kwon, K., Lee, K. N., and Yu, M. (1996) The native strains in the hydrophobic core and flexible reactive loop of a serine protease inhibitor: crystal structure of an uncleaved α1-antitrypsin at 2.7 å. *Structure 4*, 1181–1192.

(68) Im, H., Seo, E. J., and Yu, M. H. (1999) Metastability in the inhibitory mechanism of human a1-antitrypsin. *J. Biol. Chem. 274*, 11072–11077.

(69) Seo, E. J., Im, H., Maeng, J. S., Kim, K. E., and Yu, M. H. (2000) Distribution of the native strain in human α1-antitrypsin and its association with protease inhibitor function. *J. Biol. Chem. 275*, 16904–16909.

(70) Bottomley, S. P., and Stone, S. R. (1998) Protein engineering of chimeric Serpins: an investigation into effects of the serpin scaffold and reactive centre loop length. *Protein Eng. Des. Sel. 11*, 1243–1247.

(71) Whisstock, J. C., Skinner, R., Carrell, R. W., and Lesk, a M. (2000) Conformational changes in serpins: I. The native and cleaved conformations of alpha(1)-antitrypsin. *J. Mol. Biol. 296*, 685–699.

(72) Huntington, J. A. (2011) Serpin structure, function and dysfunction. *J. Thromb. Haemost. 9*, 26–34.

(73) Elliott, P. R., Pei, X. Y., Dafforn, T. R., and Lomas, D. A. (2000) Topography of a 2.0 A structure of alpha1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci. 9*, 1274–81.

(74) Villanueva, G. B., and Allen, N. (1983) Demonstration of a two-domain structure of antithrombin III during its denaturation in guanidinium chloride. *J. Biol. Chem. 258*, 11010–3.

(75) James, E. L., Whisstock, J. C., Gore, M. G., and Bottomley, S. P. (1999) Probing the Unfolding Pathway of α 1 -Antitrypsin. *J. Biol. Chem. 274*, 9482–9488.

(76) Tew, D. J., and Bottomley, S. P. (2001) Probing the equilibrium denaturation of the serpin alpha(1)-antitrypsin with single tryptophan mutants; evidence for structure in the urea unfolded state. *J. Mol. Biol. 313*, 1161–1169.

(77) Liu, L., Werner, M., and Gershenson, A. (2014) Collapse of a Long Axis: Single-Molecule Förster Resonance Energy Transfer and Serpin Equilibrium Unfolding. *Biochemistry 53*, 2903–2914.

(78) Huang, X., Zheng, Y., Zhang, F., Wei, Z., Wang, Y., Carrell, R. W., Read, R. J., Chen, G. Q., and Zhou, A. (2016) Molecular mechanism of Z α1-antitrypsin deficiency. *J. Biol. Chem. 291*, 15674–15686.

(79) Krishnan, B., and Gierasch, L. M. (2011) Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat. Struct. Mol. Biol. 18*, 222–6.

(80) Stocks, B. B., Sarkar, A., Wintrode, P. L., and Konermann, L. (2012) Early Hydrophobic Collapse of α1-Antitrypsin Facilitates Formation of a Metastable State: Insights from Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol. 423*, 789–799.

(81) Cabrita, L. D., Whisstock, J. C., and Bottomley, S. P. (2002) Probing the role of the F-helix in serpin stability through a single tryptophan substitution. *Biochemistry 41*, 4575–81.

(82) Kim, D., and Yu, M. (1996) Folding Pathway of Human α1-Antitrypsin: Characterization of an Intermediate That Is Active but Prone to Aggregation. *Biochem. Biophys. Res. Commun. 226*, 378–384.

(83) Pearce, M. C., Cabrita, L. D., Rubin, H., Gore, M. G., and Bottomley, S. P. (2004) Identification of residual structure within denatured antichymotrypsin: implications for serpin folding and misfolding. *Biochem. Biophys. Res. Commun. 324*, 729–735.

(84) Chow, M. K. M., Lomas, D. a, and Bottomley, S. P. (2004) Promiscuous beta-strand interactions and the conformational diseases. *Curr. Med. Chem. 11*, 491–499.

(85) Gooptu, B., Hazes, B., Chang, W. S., Dafforn, T. R., Carrell, R. W., Read, R. J., and Lomas, D. A. (2000) Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. *Proc. Natl. Acad. Sci. U. S. A. 97*, 67–72.

(86) Dolmer, K., and Gettins, P. G. W. (2012) How the Serpin α 1 -Proteinase Inhibitor Folds. *J. Biol. Chem. 287*, 12425–12432.

(87) Tsutsui, Y., Dela Cruz, R., and Wintrode, P. L. (2012) Folding mechanism of the metastable serpin α1-antitrypsin. *Proc. Natl. Acad. Sci. U. S. A. 109*, 4467–72.

(88) Aulak, K. S., Eldering, E., Hack, C. E., Lubbers, Y. P., Harrison, R. A., Mast, A., Cicardi, M., and Davis, A. E. (1993) A hinge region mutation in C1-inhibitor (Ala436-->Thr) results in nonsubstrate-like behavior and in polymerization of the molecule. *J. Biol. Chem. 268*, 18088–94.

(89) Wilczynska, M., Fa, M., Ohlsson, P.-I., and Ny, T. (1995) The Inhibition mechanism of Serpins. *J. Biol. Chem. 1*, 29652–29655.

(90) Stratikos, E., and Gettins, P. G. W. (1999) Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 A and full insertion of the reactive center loop into -sheet A. *Proc. Natl. Acad. Sci. 96*, 4808–4813.

(91) Zhou, A., Carrell, R. W., and Huntington, J. A. (2001) The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. *J. Biol. Chem. 276*, 27541–7.

(92) Gettins, P. G. W. (2002) Serpin structure, mechanism, and function. *Chem. Rev. 102*, 4751–804.

(93) Lawrence, D. A., Olson, S. T., Muhammad, S., Day, D. E., Kvassman, J. O., Ginsburg, D., and Shore, J. D. (2000) Partitioning of serpin-proteinase reactions between stable inhibition and substrate cleavage is regulated by the rate of serpin reactive center loop insertion into β-sheet A. *J. Biol. Chem. 275*, 5839–5844.

(94) Lee, C., Park, S. H., Lee, M. Y., and Yu, M. H. (2000) Regulation of protein function by native metastability. *Proc. Natl. Acad. Sci. U. S. A. 97*, 7727–31.

(95) Yamasaki, M., Li, W., Johnson, D. J. D., and Huntington, J. A. (2008) Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature 455*, 1255–1258.

(96) Zheng, X., Wintrode, P. L., and Chance, M. R. (2008) Complementary structural mass spectrometry techniques reveal local dynamics in functionally important regions of a metastable serpin. *Structure 16*, 38–51.

(97) Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000) Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. *Genome Res. 10*, 1845–1864.

(98) Blouse, G. E., Perron, M. J., Kvassman, J.-O., Yunus, S., Thompson, J. H., Betts, R. L., Lutter, L. C., and Shore, J. D. (2003) Mutation of the highly conserved tryptophan in the serpin breach region alters the inhibitory mechanism of plasminogen activator inhibitor-1. *Biochemistry 42*, 12260–72.

(99) Tsutsui, Y., Liu, L., Gershenson, A., and Wintrode, P. L. (2006) The Conformational Dynamics of a Metastable Serpin Studied by Hydrogen Exchange and Mass Spectrometry. *Biochemistry 45*, 6561–6569.

(100) Baek, J.-H., Yang, W. S., Lee, C., and Yu, M.-H. (2009) Functional unfolding of alpha1-antitrypsin probed by hydrogen-deuterium exchange coupled with mass spectrometry. *Mol. Cell.*

Proteomics 8, 1072–81.

(101) Cazzolli, G., Wang, F., a Beccara, S., Gershenson, A., Faccioli, P., and Wintrode, P. L. (2014) Serpin latency transition at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A. 111*, 15414–9.

(102) Lomas, D. A., Elliott, P. R., Chang, W. S., Wardell, M. R., and Carrell, R. W. (1995) Preparation and characterization of latent alpha 1-antitrypsin. *J. Biol. Chem. 270*, 5282–5288.

(103) Stein, P. E., and Carrell, R. W. (1995) What do dysfunctional serpins tell us about molecular mobility and disease? *Nature 1*, 1–6.

(104) Devlin, G. L., Chow, M. K. M., Howlett, G. J., and Bottomley, S. P. (2002) Acid Denaturation of alpha1-antitrypsin: characterization of a novel mechanism of serpin polymerization. *J. Mol. Biol. 324*, 859–70.

(105) Chang, W.-S. W., and Lomas, D. A. (1998) Latent α 1 -Antichymotrypsin. *J. Biol. Chem. 273*, 3695–3701.

(106) Ye, S., Cech, A. L., Belmares, R., Bergstrom, R. C., Tong, Y., Corey, D. R., Kanost, M. R., and Goldsmith, E. J. (2001) The structure of a Michaelis serpin-protease complex. *Nat. Struct. Biol. 8*, 979–983.

(107) Im, H., Woo, M. S., Hwang, K. Y., and Yu, M. H. (2002) Interactions causing the kinetic trap in serpin protein folding. *J. Biol. Chem. 277*, 46347–46354.

(108) Huntington, J. a, Sendall, T. J., and Yamasaki, M. (2009) New insight into serpin polymerization and aggregation. *Prion 3*, 12–4.

(109) Belorgey, D., Irving, J. A., Ekeowa, U. I., Freeke, J., Roussel, B. D., Miranda, E., Pérez, J., Robinson, C. V., Marciniak, S. J., Crowther, D. C., Michel, C. H., and Lomas, D. A. (2011) Characterisation of serpin polymers in vitro and in vivo. *Methods 53*, 255–66.

(110) Lomas, D. A., and Carrell, R. W. (2002) Serpinopathies and the conformational dementias. *Nat. Rev. Genet. 3*, 759–68.

(111) Poller, W., Faber, J.-P., Weidinger, S., Tief, K., Scholz, S., Fischer, M., Olek, K., Kirchgesser, M., and Heidtmann, H.-H. (1993) A Leucine-to-Proline Substitution Causes a Defective α1-Antichymotrypsin Allele Associated with Familial Obstructive Lung Disease. *Genomics 17*, 740–743.

(112) Davis, R. L., Shrimpton, a E., Holohan, P. D., Bradshaw, C., Feiglin, D., Collins, G. H., Sonderegger, P., Kinter, J., Becker, L. M., Lacbawan, F., Krasnewich, D., Muenke, M., Lawrence, D. a, Yerby, M. S., Shaw, C. M., Gooptu, B., Elliott, P. R., Finch, J. T., Carrell, R. W., and Lomas, D. A. (1999) Familial dementia caused by polymerization of mutant neuroserpin. *Nature 401*, 376–9.

(113) Picard, V., Dautzenberg, M. D., Villoutreix, B. O., Orliaguet, G., Alhenc-Gelas, M., and

Aiach, M. (2003) Antithrombin Phe229Leu: A new homozygous variant leading to spontaneous antithrombin polymerization in vivo associated with severe childhood thrombosis. *Blood 102*, 919–925.

(114) Zhou, A., Faint, R., Charlton, P., Dafforn, T. R., Carrells, R. W., and Lomas, D. A. (2001) Polymerization of Plasminogen Activator Inhibitor-1. *J. Biol. Chem. 276*, 9115–9122.

(115) Purkayastha, P., Klemke, J. W., Lavender, S., Oyola, R., Cooperman, B. S., and Gai, F. (2005) α1-antitrypsin polymerization: A fluorescence correlation spectroscopic study. *Biochemistry 44*, 2642–2649.

(116) Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C., and Huntington, J. a. (2011) Molecular basis of α1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep. 12*, 1011–1017.

(117) Dunstone, M. a, Dai, W., Whisstock, J. C., Rossjohn, J., Pike, R. N., Feil, S. C., Le Bonniec, B. F., Parker, M. W., and Bottomley, S. P. (2000) Cleaved antitrypsin polymers at atomic resolution. *Protein Sci. 9*, 417–420.

(118) Mast, a E., Enghild, J. J., and Salvesen, G. (1992) Conformation of the reactive site loop of alpha 1-proteinase inhibitor probed by limited proteolysis. *Biochemistry 31*, 2720–8.

(119) Zhou, A., Stein, P. E., Huntington, J. a., and Carrell, R. W. (2003) Serpin polymerization is prevented by a hydrogen bond network that is centered on his-334 and stabilized by glycerol. *J. Biol. Chem. 278*, 15116–22.

(120) Devlin, G. L., and Bottomley, S. P. (2005) A protein family under "stress" - serpin stability, folding and misfolding. *Front. Biosci. 10*, 288–99.

(121) Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P., and Lomas, D. A. (1999) A Kinetic Mechanism for the Polymerization of α 1 -Antitrypsin. *J. Biol. Chem. 274*, 9548–9555.

(122) Crowther, D. C., Serpell, L. C., Dafforn, T. R., Gooptu, B., and Lomas, D. A. (2003) Nucleation of α1 -Antichymotrypsin Polymerization. *Biochemistry 42*, 2355–2363.

(123) James, E. L., and Bottomley, S. P. (1998) The mechanism of alpha 1-antitrypsin polymerization probed by fluorescence spectroscopy. *Arch. Biochem. Biophys. 356*, 296–300.

(124) Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C., and Bottomley, S. P. (2010) Kinetic instability of the serpin Z alpha1-antitrypsin promotes aggregation. *J. Mol. Biol. 396*, 375–83.

(125) Pearce, M. C., Rubin, H., and Bottomley, S. P. (2000) Conformational Change and Intermediates in the Unfolding of α 1 -Antichymotrypsin. *J. Biol. Chem. 275*, 28513–28518.

(126) Stoller, J. K., and Aboussouan, L. S. (2012) A review of α1-antitrypsin deficiency. *Am. J. Respir. Crit. Care Med. 185*, 246–59.

(127) Lee, K. J., Lee, S. M., Gil, J. Y., Kwon, O., Kim, J. Y., Park, S. J., Chung, H.-S., and Oh,

D.-B. (2013) N-glycan analysis of human α1-antitrypsin produced in Chinese hamster ovary cells. *Glycoconj. J. 30*, 537–47.
(128) Zaimidou, S., van Baal, S., Smith, T. D., Mitropoulos, K., Ljujic, M., Radojkovic, D., Cotton, R. G., and Patrinos, G. P. (2009) A1ATVar: a relational database of human SERPINA1 gene variants leading to alpha1-antitrypsin deficiency and application of the VariVis software. *Hum. Mutat. 30*, 308–13.
(129) McCracken, A. A., Kruse, K. B., and Brown, J. L. (1989) Molecular basis for defective secretion of the Z variant of human alpha-1-proteinase inhibitor: secretion of variants having altered potential for salt bridge formation between amino acids 290 and 342. *Mol. Cell. Biol. 9*, 1406–14.
(130) Mallya, M., Phillips, R. L., Saldanha, S. A., Gooptu, B., Leigh Brown, S. C., Termine, D. J., Shirvani, A. M., Wu, Y., Sifers, R. N., Abagyan, R., and Lomas, D. A. (2007) Small Molecules Block the Polymerization of Z α 1 -Antitrypsin and Increase the Clearance of Intracellular Aggregates. *J. Med. Chem. 50*, 5357–5363.
(131) Marciniak, S. J., and Lomas, D. A. (2010) Alpha 1 -Antitrypsin Deficiency and Autophagy. *N. Engl. J. Med. 363*, 1863–1864.
(132) Matamala, N., Martínez, M. T., Lara, B., Pérez, L., Vázquez, I., Jimenez, A., Barquín, M., Ferrarotti, I., Blanco, I., Janciauskiene, S., and Delgado, B. M. (2015) Alternative transcripts of the SERPINA1 gene in alpha - 1 antitrypsin deficiency. *J. Transl. Med.* 1–11.
(133) Gooptu, B., Dickens, J. A., and Lomas, D. A. (2014) The molecular and cellular pathology of α1-antitrypsin deficiency. *Trends Mol. Med. 20*, 116–127.
(134) Elliott, P. R., Bilton, D., and Lomas, D. A. (1998) Lung polymers in Z α1-antitrypsin deficiency-related emphysema. *Am. J. Respir. Cell Mol. Biol. 18*, 670–674.
(135) Levina, V., Dai, W., Knaupp, A. S., Kaiserman, D., Pearce, M. C., Cabrita, L. D., Bird, P. I., and Bottomley, S. P. (2009) Expression, purification and characterization of recombinant Z α1-Antitrypsin-The most common cause of α1-Antitrypsin deficiency. *Protein Expr. Purif. 68*, 226–232.
(136) Jeppsson, J.-O. (1976) Amino acid substitution Glu→Lys in α 1 -antitrypsin PiZ. *FEBS Lett. 65*, 195–197.
(137) Kass, I., Knaupp, A. S., Bottomley, S. P., and Buckle, A. M. (2012) Conformational properties of the disease-causing Z variant of α1-antitrypsin revealed by theory and experiment. *Biophys. J. 102*, 2856–65.
(138) Hughes, V. A., Meklemburg, R., Bottomley, S. P., and Wintrode, P. L. (2014) The Z mutation alters the global structural dynamics of α1-antitrypsin. *PLoS One 9*, e102617.
(139) Yu, M. H., Lee, K. N., and Kim, J. (1995) The Z type variation of human alpha 1-

antitrypsin causes a protein folding defect. *Nat. Struct. Biol. 2*, 363–7.
(140) Lomas, D. A., Elliott, P. R., and Carrell, R. W. (1997) Commercial plasma α1-antitrypsin (Prolastin®) contains a conformationally inactive, latent component. *Eur. Respir. J. 10*, 672–675.
(141) Silverman, E. K., and Sandhaus, R. A. (2009) Alpha 1 -Antitrypsin Deficiency. *N. Engl. J. Med. 360*, 2749–2757.
(142) Berthelier, V., Harris, J. B., Estenson, K. N., and Baudry, J. (2015) Discovery of an inhibitor of Z-alpha1 antitrypsin polymerization. *PLoS One 10*, 1–18.
(143) Burrows, J. a, Willis, L. K., and Perlmutter, D. H. (2000) Chemical chaperones mediate increased secretion of mutant alpha 1-antitrypsin (alpha 1-AT) Z: A potential pharmacological strategy for prevention of liver injury and emphysema in alpha 1-AT deficiency. *Proc. Natl. Acad. Sci. U. S. A. 97*, 1796–801.
(144) Devlin, G. L., Parfrey, H., Tew, D. J., Lomas, D. A., and Bottomley, S. P. (2001) Prevention of polymerization of M and Z alpha1-Antitrypsin (alpha1-AT) with trimethylamine N-oxide. Implications for the treatment of alpha1-at deficiency. *Am. J. Respir. Cell Mol. Biol. 24*, 727–32.
(145) Singh, P., Khan, M. S., Naseem, A., and Jairajpuri, M. A. (2012) Analysis of surface cavity in serpin family reveals potential binding sites for chemical chaperone to reduce polymerization. *J. Mol. Model. 18*, 1143–1151.
(146) Mahadeva, R., Dafforn, T. R., Carrell, R. W., and Lomas, D. A. (2002) 6-mer Peptide Selectively Anneals to a Pathogenic Serpin Conformation and Blocks Polymerization: IMPLICATIONS FOR THE PREVENTION OF Z 1-ANTITRYPSIN-RELATED CIRRHOSIS. *J. Biol. Chem. 277*, 6771–6774.
(147) Zhou, A., Stein, P. E., Huntington, J. a., Sivasothy, P., Lomas, D. a., and Carrell, R. W. (2004) How small peptides block and reverse serpin polymerisation. *J. Mol. Biol. 342*, 931–41.
(148) Ordóñez, A., Pérez, J., Tan, L., Dickens, J. A., Motamedi-Shad, N., Irving, J. A., Haq, I., Ekeowa, U., Marciniak, S. J., Miranda, E., and Lomas, D. A. (2015) A single-chain variable fragment intrabody prevents intracellular polymerization of Z α1-antitrypsin while allowing its antiproteinase activity. *FASEB J. 29*, 2667–78.
(149) Leader, B., Baca, Q. J., and Golan, D. E. (2008) Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov. 7*, 21–39.
(150) Porebski, B. T., Keleher, S., Hollins, J. J., Nickson, A. A., Marijanovic, E. M., Borg, N. A., Costa, M. G. S., Pearce, M. A., Dai, W., Zhu, L., Irving, J. A., Hoke, D. E., Kass, I., Whisstock, J. C., Bottomley, S. P., Webb, G. I., McGowan, S., and Buckle, A. M. (2016) Smoothing a rugged protein folding landscape by sequence-based redesign. *Sci. Rep. 6*, 33958.

(151) Lehmann, M., and Wyss, M. (2001) Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol. 12*, 371–375.
(152) Kazlauskas, R. (2018) Engineering more stable proteins. *Chem. Soc. Rev.*
(153) Dai, M., Fisher, H. E., Temirov, J., Kiss, C., Phipps, M. E., Pavlik, P., Werner, J. H., and Bradbury, A. R. M. (2007) The creation of a novel fluorescent protein by guided consensus engineering. *Protein Eng. Des. Sel. 20*, 69–79.
(154) Porebski, B. T., Nickson, A. A., Hoke, D. E., Hunter, M. R., Zhu, L., McGowan, S., Webb, G. I., and Buckle, A. M. (2015) Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Eng. Des. Sel. 28*, 67–78.
(155) Fulton, K. F., Buckle, A. M., Cabrita, L. D., Irving, J. A., Butcher, R. E., Smith, I., Reeve, S., Lesk, A. M., Bottomley, S. P., Rossjohn, J., and Whisstock, J. C. (2005) The high resolution crystal structure of a native thermostable serpin reveals the complex mechanism underpinning the stressed to relaxed transition. *J. Biol. Chem. 280*, 8435–8442.
(156) Zhang, Q., Buckle, A. M., Law, R. H. P., Pearce, M. C., Cabrita, L. D., Lloyd, G. J., Irving, J. a, Smith, a I., Ruzyla, K., Rossjohn, J., Bottomley, S. P., and Whisstock, J. C. (2007) The N terminus of the serpin, tengpin, functions to trap the metastable native state. *EMBO Rep. 8*, 658–63.
(157) Grant, S. G. N., Jesseet, J., Bloomt, F. R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants *87*, 4645–4649.
(158) Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. 81*, 2035–2039.
(159) Studier, F. W., and Moffattf, B. A. (1986) Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes 113–130.
(160) Gottesman, S., Halpern, E., and Trisler, P. (1981) Role of sulA and sulB in Filamentation by Lon Mutants of Escherichia coli K-12 *148*, 265–273.
(161) Cabrita, L. D., Dai, W., and Bottomley, S. P. (2006) A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol. 6*, 12.
(162) Pearce, M. C., and Cabrita, L. D. (2011) Production of Recombinant Serpins in Escherichia coli, in *Methods in Enzymology* 1st ed., pp 13–28. Elsevier Inc.
(163) Qiagen. (2003) The QIA expressionist Handbook, 5th Ed. *Diagen GmbH, Diisseldorf, Ger.*
(164) Bachman, J. (2013) Site-directed mutagenesis. *Methods Enzymol.* 1st ed. Elsevier Inc.
(165) Horvath, A. J., Lu, B. G. C., Pike, R. N., and Bottomley, S. P. (2011) Methods to measure

the kinetics of protease inhibition by serpins. *Methods Enzymol.* 1st ed. Elsevier Inc.
(166) John, D. M., and Weeks, K. M. (2000) Van't Hoff enthalpies without baselines. *Protein Sci. 9*, 1416–1419.
(167) Greenfield, N. J. (2009) Using circular dichroism collected as a funcion of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc. 1*, 2527–2535.
(168) Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. U. S. A. 84*, 6663–7.
(169) Battye, T. G. G., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM : a new graphical interface for diffraction- image processing with MOSFLM research papers. *Acta Crystallogr. D Biol. Crystallogr 67*, 271–281.
(170) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr. 62*, 72–82.
(171) Winn, M. D., Ballard, C. C., Cowtan, K. D., and Dodson, E. J. (2011) Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 235–242.
(172) McCoy, A. J. (2007) Phaser crystallographic software. *J Appl Crystallogr 40*, 658–674.
(173) Adams, P. D. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 213–221.
(174) Emsley, P., and Cowtan, K. (2004) COOT: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr 60*, 2126–2132.
(175) Schrodinger, L. The Pymol Molecular Graphics System, Version 2.0.4.
(176) Elliott, P. R., Lomas, D. A., Carrell, R. W., and Abrahams, J. P. (1996) Inhibitory conformation of the reactive loop of alpha 1-antitrypsin. *Nat. Struct. Biol. 3*, 676–681.
(177) Dupont, D. M., Madsen, J. B., Kristensen, T., Bodker, J. S., Blouse, G. E., Wind, T., and Andreasen, P. A. (2009) Biochemical properties of plasminogen activator inhibitor-1. *Front. Biosci. (Landmark Ed. 14*, 1337–61.
(178) Mushunje, A., Evans, G., Brennan, S. O., Carrell, R. W., and Zhou, A. (2004) Latent antithrombin and its detection, formation and turnover in the circulation. *J. Thromb. Haemost. 2*, 2170–2177.
(179) Padrines, M., Schneider-Pozzer, M., and Bieth, J. G. (1989) Inhibition of neutrophil elastase by alpha-1-proteinase inhibitor oxidized by activated neutrophils. *Am. Rev. Respir. Dis. 139*, 783–90.
(180) Rubin, H., Plotnick, M., Wang, Z. M., Liu, X., Zhong, Q., Schechter, N. M., and Cooperman, B. S. (1994) Conversion of alpha 1-antichymotrypsin into a human neutrophil

elastase inhibitor: demonstration of variants with different association rate constants, stoichiometries of inhibition, and complex stabilities. *Biochemistry 33*, 7627–7633.
(181) Hopkins, P. C. R., Carrell, R. W., and Stone, S. R. (1993) Effects of mutations in the hinge region of serpins. *Biochemistry 32*, 7650–7657.
(182) Rau, J. C., Beaulieu, L. M., Huntington, J. A., and Church, F. C. (2007) Serpins in thrombosis, hemostasis and fibrinolysis. *J. Thromb. Haemost. 5 Suppl 1*, 102–15.
(183) Polderdijk, S. G. I., Adams, T. E., Ivanciu, L., Camire, R. M., Baglin, T. P., and Huntington, J. A. (2017) Design and characterization of an APC-specific serpin for the treatment of hemophilia. *Blood 129*, 105–113.
(184) Owen, M. C., Brennan, S. O., Lewis, J. H., and Carrell, R. W. (1983) Mutation of Antitrypsin to Antithrombin. *N. Engl. J. Med. 309*, 694–698.
(185) Ehrlich, H. J., Gebbink, R. K., Keijer, J., Linders, M., Preissner, K. T., and Pannekoek, H. (1990) Alteration of serpin specificity by a protein cofactor: Vitronectin endows plasminogen activator inhibitor 1 with thrombin inhibitory properties. *J. Biol. Chem. 265*, 13029–13035.
(186) Lawrence, D. a, Strandberg, L., Ericson, J., and Ny, T. (1990) Structure-Function Inhibitor Type 1 Studies of the SERPIN Plasminogen Activator. *J. Biol. Chem. 265*, 20293–20301.
(187) Patston, P. A., Roodi, N., Schifferli, J. A., Bischoff, R., Courtney, M., and Schapira, M. (1990) Reactivity of α1-antitrypsin mutants against proteolytic enzymes of the kallikrein-kinin, complement, and fibrinolytic systems. *J. Biol. Chem. 265*, 10786–10791.
(188) Hopkins, P. C. R., Crowther, D. C., Carrell, R. W., and Stone, S. R. (1995) Development of a novel recombinant serpin with potential antithrombotic properties. *J. Biol. Chem. 270*, 11866–11871.
(189) Chaillan-Huntington, C. E., Gettins, P. G. W., Huntington, J. A., and Patston, P. A. (1997) The P6-P2 region of serpins is critical for proteinase inhibition and complex stability. *Biochemistry 36*, 9562–9570.
(190) Djie, M. Z., Stone, S. R., and Le Bonniec, B. F. (1997) Intrinsic specificity of the reactive site loop of α1-antitrypsin, α1-antichymotrypsin, antithrombin III, and protease nexin I. *J. Biol. Chem. 272*, 16268–16273.
(191) Plotnick, M. I., Schechter, N. M., Wang, Z. M., Liu, X., and Rubin, H. (1997) Role of the P6-P3' region of the serpin reactive loop in the formation and breakdown of the inhibitory complex. *Biochemistry 36*, 14601–14608.
(192) Whisstock, J. C., Silverman, G. A., Bird, P. I., Bottomley, S. P., Kaiserman, D., Luke, C. J., Pak, S. C., Reichhart, J. M., and Huntington, J. A. (2010) Serpins flex their muscle: II. Structural insights into target peptidase recognition, polymerization, and transport functions. *J. Biol. Chem. 285*, 24307–24312.

(193) Qureshi, T., Goswami, S., McClintock, C. S., Ramsey, M. T., and Peterson, C. B. (2016) Distinct encounter complexes of PAI-1 with plasminogen activators and vitronectin revealed by changes in the conformation and dynamics of the reactive center loop. *Protein Sci. 25*, 499–510.
(194) Gettins, P. G. W., and Olson, S. T. (2009) Exosite determinants of serpin specificity. *J. Biol. Chem. 284*, 20441–20445.
(195) Yang, L., Irving, J. A., Dai, W., Aguilar, M., and Bottomley, S. P. (2018) Probing the folding pathway of a consensus serpin using single tryptophan mutants. *Sci. Rep.* 1–15.
(196) Belorgey, D., Hägglöf, P., Onda, M., and Lomas, D. A. (2010) pH-dependent stability of neuroserpin is mediated by histidines 119 and 138; Implications for the control of β-sheet a and polymerization. *Protein Sci. 19*, 220–228.
(197) Patschull, A. O. M., Segu, L., Nyon, M. P., Lomas, D. A., Nobeli, I., Barrett, T. E., and Gooptu, B. (2011) Therapeutic target-site variability in α 1-antitrypsin characterized at high resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67*, 1492–1497.
(198) Buck, M. J., and Atchley, W. R. (2005) Networks of Coevolving Sites in Structural and Functional Domains of Serpin Proteins. *Mol. Biol. Evol. 22*, 1627–1634.
(199) Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007) Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A. 104*, 19819–24.
(200) Jenik, M., Parra, R. G., Radusky, L. G., Turjanski, A., Wolynes, P. G., and Ferreiro, D. U. (2012) Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res. 40*, W348–W351.
(201) Wei, A., Rubin, H., Cooperman, B. S., and Christianson, D. W. (1994) Crystal structure of an uncleaved serpin reveals the conformation of an inhibitory reactive loop. *Nat. Struct. Mol. Biol. 1*, 251–258.
(202) Schreiber, G., and Fersht, A. R. (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol. 3*, 427–431.
(203) Schreiber, G., and Fersht, A. R. (1993) Interaction of Barnase with Its Polypeptide Inhibitor Barstar Studied by Protein Engineering. *Biochemistry 32*, 5145–5150.
(204) Plotnick, M. I., Rubin, H., and Schechter, N. M. (2002) The Effects of Reactive Site Location on the Inhibitory Properties of the Serpin α 1 -Antichymotrypsin. *J. Biol. Chem. 277*, 29927–29935.
(205) Liu, L., Mushero, N., Hedstrom, L., and Gershenson, A. (2007) Short-lived protease serpin complexes: partial disruption of the rat trypsin active site. *Protein Sci. 16*, 2403–2411.
(206) Liu, L., Mushero, N., Hedstrom, L., and Gershenson, A. (2006) Conformational distributions of protease-serpin complexes: A partially translocated complex. *Biochemistry 45*,

10865–10872.

(207) Gong, L., Liu, M., Zeng, T., Shi, X., Yuan, C., Andreasen, P. A., and Huang, M. (2015) Crystal structure of the Michaelis complex between tissue-type plasminogen activator and plasminogen activators inhibitor-1. *J. Biol. Chem. 290*, 25795–25804.

(208) Lin, Z., Jiang, L., Yuan, C., Jensen, J. K., Zhang, X., Luo, Z., Furie, B. C., Furie, B., Andreasen, P. a, and Huang, M. (2011) Structural basis for recognition of urokinase-type plasminogen activator by plasminogen activator inhibitor-1. *J. Biol. Chem. 286*, 7027–32.

(209) Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M. J., Sambrook, J. F., and Bassel-Duby, R. S. (1990) Amino acid residues that affect interaction of tissue-type plasminogen activator with plasminogen activator inhibitor 1. *Proc. Natl. Acad. Sci. U. S. A. 87*, 3530–3.

(210) Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M. J., and Sambrook, J. F. (1989) Serpin-resistant mutants of human tissue-type plasminogen activator. *Nature 339*, 721–724.

(211) Lee, K. N., Im, H., Kang, S. W., and Yu, M. H. (1998) Characterization of a human alpha1-antitrypsin variant that is as stable as ovalbumin. *J. Biol. Chem. 273*, 2509–2516.

(212) Seo, E. J., Lee, C., and Yu, M. H. (2002) Concerted regulation of inhibitory activity of α1-antitrypsin by the native strain distributed throughout the molecule. *J. Biol. Chem. 277*, 14216–14220.

(213) Im, H., Ryu, M. J., and Yu, M. H. (2004) Engineering thermostability in serine protease inhibitors. *Protein Eng. Des. Sel. 17*, 325–331.

(214) Shin, J.-S., and Yu, M.-H. (2006) Viscous drag as the source of active site perturbation during protease translocation: insights into how inhibitory processes are controlled by serpin metastability. *J. Mol. Biol. 359*, 378–89.

(215) Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. M., Pieper, U., Sali, A., and Marti-Renom, M. A. (2002) Comparative protein structure modeling using Modeller, in *Current Protocols in Bioinformatics*, pp 5–6.

(216) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.

(217) Mcgibbon, R. T., Beauchamp, K. A., Schwantes, C. R., Wang, L., Hern, C. X., Herrigan, M. P., Lane, T. J., Swails, J. M., and Pande, V. S. (2014) MDTraj : a modern , open library for the analysis of molecular dynamics trajectories MDTraj : a modern , open library for the analysis of molecular dynamics trajectories. *Biorxiv.Org 109*, 0–2.

(218) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. 98*, 10037–10041.

(219) Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H.,

Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci. 27*, 112–128.

(220) Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res. 32*, W665–W667.

(221) Søndergaard, C. R., Olsson, M. H. M., and Rostkowski Michałand Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput 7*, 2284–2295.

(222) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (2001) Comparison of simple potential functions for simulating liquid water Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 926*, 926–935.

(223) Suk Joung, I., and Cheatham, T. E. (2008) Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B 112*, 9020–9041.

(224) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput. 11*, 3696–3713.

(225) Li, P., Roberts, B. P., Chakravorty, D. K., and Merz, K. M. (2013) Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput. 9*, 2733–2748.

(226) Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.

(227) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089–10092.

(228) Case DA, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE, et. al. (2014) Amber 14.

(229) Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.

(230) Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol. 30*, 772–780.

(231) Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein

structures. *Nucleic Acids Res. 33*, 299–302.
(232) Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res. 44*, W344–W350.
(233) Kass, I., and Horovitz, A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins Struct. Funct. Genet. 48*, 611–617.
(234) Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science 286*, 295–9.
(235) Dekker, J. P., Fodor, A., Aldrich, R. W., and Gary, Y. (2004) A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics 20*, 1565–1572.
(236) McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J. 109*, 1528–32.
(237) Hunter, J. D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng. 9*, 99–104.
(238) Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013) Density-Based Clustering Based on Hierarchical Density Estimates, in *Antimicrobial agents and chemotherapy*, pp 160–172.
(239) McInnes, L., Healy, J., and Astels, S. (2017) hdbscan: Hierarchical density based clustering. *J. Open Source Softw. 2*, 11–12.
(240) Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M. Y., Guzovsky, A. B., Wolynes, P. G., and Ferreiro, D. U. (2016) Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res. 44*, W356–W360.
(241) Dobson, C. M. (2002) Protein Misfolding Diseases: Getting out of shape. *Nature 418*, 729–30.
(242) Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A. 103*, 5869–74.
(243) Magliery, T. J. (2015) Protein stability: computation, sequence statistics, and new experimental methods. *Curr. Opin. Struct. Biol. 33*, 161–8.
(244) Sanchez-ruiz, J. M. (2010) Biophysical Chemistry Protein kinetic stability *148*, 1–15.
(245) Matthews, B. W. (1993) Structural and genetic analysis of protein stability. *Annu. Rev. Biochem. 62*, 139–60.
(246) Dill, K. A. (1990) Dominant forces in protein folding 7133–7155.
(247) Miyazaki, K., and Arnold, F. H. (1999) Exploring Nonnatural Evolutionary Pathways by Saturation Mutagenesis : Rapid Improvement of Protein Function 716–720.
(248) Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000)

Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol. 297*, 1015–1026.
(249) Giver, L., Gershenson, A., Freskgard, P.-O., and Arnold, F. H. (1998) Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. 95*, 12809–12813.
(250) Arnold, F. H., Wintrode, P. L., Miyazaki, K., Gershenson, A., Arnold, F. H., and Wintrode, P. L. (2001) How enzymes adapt : lessons from directed evolution *26*, 100–106.
(251) Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet. 16*, 379–394.
(252) Lehmann, M., Pasamontes, L., Lassen, S. F., and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta 1543*, 408–415.
(253) Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L., and Bryan, P. N. (1989) Large Increases in General Stability for Subtilisin BPN' through Incremental Changes in the Free Energy of Unfolding. *Biochemistry 28*, 7205–7213.
(254) Fu, H., Grimsley, G. R., Razvi, A., Scholtz, J. M., and Pace, C. N. (2009) Increasing protein stability by improving beta-turns. *Proteins 77*, 491–8.
(255) Fersht, A. R. (1972) Conformational equilibria in α- and δ-chymotrypsin. *J. Mol. Biol. 64*, 497–509.
(256) Strop, P., and Mayo, S. L. (2000) Contribution of surface salt bridges to protein stability. *Biochemistry 39*, 1251–1255.
(257) Lee, C. W., Wang, H. J., Hwang, J. K., and Tseng, C. P. (2014) Protein thermal stability enhancement by designing salt bridges: A combined computational and experimental study. *PLoS One 9*.
(258) Chaillan-huntington, C. E., Gettins, P. G. W., Huntington, J. A., and Patston, P. A. (1997) The P 6 - P 2 Region of Serpins Is Critical for Proteinase Inhibition and Complex *2960*, 9562–9570.
(259) Anderson, D. E., Hurley, J. H., Nicholson, H., Matthews, B. W., and Baase, W. A. (1993) Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant *0*, 1285–1290.
(260) Desjarlais, J. R., and Handel, T. M. (2018) De novo design of the hydrophobic cores of proteins 2006–2018.
(261) Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins. *Protein Sci. 15*, 1569–78.
(262) Schreiber, G., Buckle, A. M., and Fersht, A. R. (1994) Stability and function: two constraints in the evolution of barstar and other proteins. *Structure 2*, 945–951.
(263) Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. H. (1998)

Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. 95*, 2056–2060.

(264) Imanaka, T., Shibazaki, M., and Takagi, M. (1986) A new way of enhancing the thermostability of proteases. *Nature 324*, 695–7.

(265) Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. 92*, 452–456.

(266) Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol. 4*, 35–37.

(267) Sánchez, I. E., Tejero, J., Medina, M., and Serrano, L. (2006) Point Mutations in Protein Globular Domains : Contributions from Function , Stability and Misfolding 422–432.

(268) Im, H., Seo, E. J., and Yu, M. (1999) Metastability in the Inhibitory Mechanism of Human α 1 -Antitrypsin. *J. Biol. Chem. 274*, 11072–11077.

(269) Patschull, A. O. M., Segu, L., Nyon, M. P., Lomas, D. A., Nobeli, I., Barrett, T. E., and Gooptu, B. (2011) Therapeutic target-site variability in α 1 -antitrypsin characterized at high resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67*, 1492–1497.

(270) Knaupp, A. S., Keleher, S., Yang, L., Dai, W., Bottomley, S. P., and Pearce, M. C. (2013) The Roles of Helix I and Strand 5A in the Folding, Function and Misfolding of a1-Antitrypsin. *PLoS One* (Crowther, D. C., Ed.) *8*, e54766.

(271) Parfrey, H., Mahadeva, R., Ravenhill, N. A., Zhou, A., Dafforn, T. R., Foreman, R. C., and Lomas, D. A. (2003) Targeting a surface cavity of α1-antitrypsin to prevent conformational disease. *J. Biol. Chem. 278*, 33060–33066.

(272) Maxwell, K. L., and Davidson, A. R. (1998) Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides the best prediction of stability effects. *Biochemistry 37*, 16172–16182.

(273) Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc. 1*, 2527–35.

(274) Kwon, K. S., Lee, S., and Yu, M. H. (1995) Refolding of alpha 1-antitrypsin expressed as inclusion bodies in Escherichia coli: characterization of aggregation. *Biochim. Biophys. Acta 1247*, 179–84.

(275) Taverna, D. M., and Goldstein, R. A. (2002) Why are proteins marginally stable? *Proteins Struct. Funct. Genet. 46*, 105–109.

(276) Irving, J. A., Cabrita, L. D., Rossjohn, J., Pike, R. N., Bottomley, S. P., and Whisstock, J. C. (2003) The 1.5 A crystal structure of a prokaryote serpin: controlling conformational change in a heated environment. *Structure 11*, 387–97.

(277) Cabrita, L. D., Irving, J. A., Pearce, M. C., Whisstock, J. C., and Bottomley, S. P. (2007)

Aeropin from the extremophile Pyrobaculum aerophilum bypasses the serpin misfolding trap. *J. Biol. Chem. 282*, 26802–9.

(278) Francis, D. M., and Page, R. (2010) Strategies to Optimize Protein Expression in E . coli 1–29.

(279) Deller, M. C., Kong, L., and Rupp, B. (2016) Protein stability: A crystallographer's perspective. *Acta Crystallogr. Sect. Struct. Biol. Commun. 72*, 72–95.

(280) Bird, P. I., Pak, S. C., Worrall, D. M., and Bottomley, S. P. (2004) Production of recombinant serpins in Escherichia coli. *Methods 32*, 169–76.

(281) Kwon, K. S., Kim, J., Shin, H. S., and Yu, M. H. (1994) Single amino acid substitutions of alpha 1-antitrypsin that confer enhancement in thermal stability. *J. Biol. Chem. 269*, 9627–31.

(282) Zhu, W., Li, L., Deng, M., Wang, B., Li, M., Ding, G., Yang, Z., Medynski, D., Lin, X., Ouyang, Y., Lin, J., Li, L., and Lin, X. (2018) Oxidation-resistant and thermostable forms of alpha-1 antitrypsin from Escherichia coli inclusion bodies *8*, 1711–1721.

(283) Lomas, D. A., Elliott, P. R., Sidhar, S. K., Foreman, R. C., Finch, J. T., Cox, D. W., Whisstock, J. C., and Carrell, R. W. (1995) alpha 1-Antitrypsin Mmalton (Phe52-deleted) forms loop-sheet polymers in vivo. Evidence for the C sheet mechanism of polymerization. *J. Biol. Chem. 270*, 16864–70.

(284) Fraizer, G. C., Harrold, T. R., Hofker, M. H., and Cox, D. W. (1989) In-frame single codon deletion in the Mmalton deficiency allele of alpha 1-antitrypsin. *Am. J. Hum. Genet. 44*, 894–902.

(285) Graham, A., Kalsheker, N., Newton, C., Bamforth, F., Powell, S., and Markham, A. (1989) Molecular characterisation of three alpha-1-antitrypsin deficiency variants: proteinase inhibitor (Pi) nullcardiff (Asp256 Val); Pi Mmalton (Phe51 deletion) and Pi I (Arg39 Cys). *Hum. Genet. 84*, 55–58.

(286) Lomas, D. A., Finch, J. T., Seyama, K., Nukiwa, T., and Carrell, R. W. (1993) Alpha 1-antitrypsin Siiyama (Ser53-->Phe). Further evidence for intracellular loop-sheet polymerization. *J. Biol. Chem. 268*, 15333–5.

(287) Kang, H. A., Lee, K. N., and Yu, M. H. (1997) Folding and stability of the Z and S(iiyama) genetic variants of human alpha1-antitrypsin. *J. Biol. Chem. 272*, 510–6.

(288) Pearce, M. C., Morton, C. J., Feil, S. C., Hansen, G., Adams, J. J., Parker, M. W., and Bottomley, S. P. (2008) Preventing serpin aggregation: the molecular mechanism of citrate action upon antitrypsin unfolding. *Protein Sci. 17*, 2127–33.

(289) Lomas, D. A., Evans, D. L., Finch, J. T., and Carrell, R. W. (1992) The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature 357*, 605–7.

(290) Gettins, P. G. W. (2002) The F-helix of serpins plays an essential, active role in the

proteinase inhibition mechanism. *FEBS Lett. 523*, 2–6.
(291) Graham, A., Kalsheker, N. A., Bamforth, F. J., Newton, C. R., Markham, A. F., Pharmaceuticals, I., Park, A., and Sk, C. (1990) Molecular characterisation of two alpha-l-antitrypsin deficiency variants :Gly----ser, Null Newport Ser-----leu, Pi Z Wrexham Leu, Pi Z Wrexham Ser *c*, 537–540.
(292) Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding "speed limit." *Curr. Opin. Struct. Biol. 14*, 76–88.
(293) Onuchic, J. N., and Wolynes, P. G. (2004) Theory of protein folding. *Curr. Opin. Struct. Biol. 14*, 70–75.
(294) Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics 1751*, 119–139.
(295) Royer, C. a. (2006) Probing Protein Folding and Conformational Transitions with Fluorescence Probing Protein Folding and Conformational Transitions with Fluorescence *106*, 1769–1784.
(296) Krishna, M. M. G., Hoang, L., Lin, Y., and Englander, S. W. (2004) Hydrogen exchange methods to study protein folding. *Methods 34*, 51–64.
(297) Salvatella, X., Dobson, C. M., Fersht, A. R., and Vendruscolo, M. (2005) Determination of the folding transition states of barnase by using I-value-restrained simulations validated by double mutant IJ-values. *Proc. Natl. Acad. Sci. 102*, 12389–12394.
(298) Fersht, A. R., and Sato, S. (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A. 101*, 7976–81.
(299) Schuler, B., and Eaton, W. A. (2008) Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol. 18*, 16–26.
(300) Hoffmann, A., Kane, A., Nettels, D., Hertzog, D. E., Baumgartel, P., Lengefeld, J., Reichardt, G., Horsley, D. A., Seckler, R., Bakajin, O., and Schuler, B. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. 104*, 105–110.
(301) Banerjee, P. R., and Deniz, A. A. (2014) Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem. Soc. Rev. 43*, 1172–88.
(302) Sasmal, D. K., Pulido, L., Kasal, S., and Huang, J. (2017) Single-Molecule Fluorescence Resonance Energy Transfer in Molecular Biology *8*, 19928–19944.
(303) Nettels, D., Hoffmann, A., and Schuler, B. (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J. Phys. Chem. B 112*, 6137–6146.
(304) Gambin, Y., and Deniz, A. A. (2010) Multicolor single-molecule FRET to explore protein

folding and binding. *Mol. Biosyst. 6*, 1540–7.
(305) Deniz, A. A., Laurence, T. A., Beligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S., Dawson, P. E., Schultz, P. G., and Weiss, S. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. 97*, 5179–5184.
(306) Jia, Y., Talaga, D. S., Lau, W. L., Lu, H. S. M., DeGrado, W. F., and Hochstrasser, R. M. (1999) Folding dynamics of single GCN-4 peptides by fluorescence resonant energy transfer confocal microscopy. *Chem. Phys. 247*, 69–83.
(307) Talaga, D. S., Lau, W. L., Roder, H., Tang, J., Jia, Y., DeGrado, W. F., and Hochstrasser, R. M. (2000) Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proc. Natl. Acad. Sci. 97*, 13021–13026.
(308) Itzhaki, L. S., Otzen, D. E., and Fersht, a R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol. 254*, 260–288.
(309) Ferreon, A. C. M., Gambin, Y., Lemke, E. A., and Deniz, A. A. (2009) Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl. Acad. Sci. U. S. A. 106*, 5645–50.
(310) Tsutsui, Y., and Wintrode, P. L. (2007) Cooperative Unfolding of a Metastable Serpin to a Molten Globule Suggests a Link Between Functional and Folding Energy Landscapes 245–255.
(311) Lomas, D. A., Evans, D. L., Stone, S. R., Chang, W. S., and Carrell, R. W. (1993) Effect of the Z mutation on the physical and inhibitory properties of alpha 1-antitrypsin. *Biochemistry 32*, 500–508.
(312) Yang, L., Irving, J. A., Dai, W., Aguilar, M. I., and Bottomley, S. P. (2018) Probing the folding pathway of a consensus serpin using single tryptophan mutants. *Sci. Rep. 8*, 1–15.
(313) Moroi, M., and Yamasaki, M. (1974) Mechanism of interaction of bovine trypsin with human alpha1-antitrypsin. *Biochim. Biophys. Acta 359*, 130–41.
(314) Nagano, N., Ota, M., and Nishikawa, K. (1999) Strong hydrophobic nature of cysteine residues in proteins. *FEBS Lett. 458*, 69–71.
(315) Marino, S. M., and Gladyshev, V. N. (2010) Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J. Mol. Biol. 404*, 902–16.
(316) Feller, G. (2017) Protein folding at extreme temperatures: Current issues. *Semin. Cell Dev. Biol.*
(317) Szymkowski, D. E. (2005) Creating the next generation of protein therapeutics through rational drug design. *Curr. Opin. Drug Discov. Devel. 8*, 590–600.
(318) Kimchi-Sarfaty, C., Schiller, T., Hamasaki-Katagiri, N., Khan, M. A., Yanover, C., and

Sauna, Z. E. (2013) Building better drugs: Developing and regulating engineered therapeutic proteins. *Trends Pharmacol. Sci. 34*, 534–548.

(319) Marshall, S. A., Lazar, G. A., Chirino, A. J., and Desjarlais, J. R. (2003) Rational design and engineering of therapeutic proteins. *Drug Discov. Today 8*, 212–21.

(320) Carter, P. J. (2011) Introduction to current and future protein therapeutics: A protein engineering perspective. *Exp. Cell Res. 317*, 1261–1269.

(321) Wilson, C. J. (2015) Rational protein design: Developing next-generation biological therapeutics and nanobiotechnological tools. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology 7*, 330–341.

(322) Reynolds, N. A., and Wagstaff, A. J. (2004) Insulin aspart: a review of its use in the management of type 1 or 2 diabetes mellitus. *Drugs 64*, 1957–74.

(323) Gammeltoft, S., Hansen, B. F., Dideriksen, L., Lindholm, A., Schäffer, L., Trüb, T., Dayan, A., Kurtzhals, P., Gammeltoft, S., Hansen, B. F., Dideriksen, L., Schäffer, L., Trüb, T., Dayan, A., and Kurtzhals, P. (2005) Expert Opinion on Investigational Drugs Insulin aspart : a novel rapid-acting human insulin analogue Expert Opinion on Investigational Drugs Insulin aspart : a novel rapid-acting human insulin analogue *3784*.

(324) Haycox, A. (2004) Insulin aspart : an evidence-based medicine review. *Clin. Drug Investig. 24*, 695–717.

(325) Gerich, J. E., and Gerich, J. E. (2008) Insulin glargine : long-acting basal insulin analog for improved metabolic control Insulin glargine : long-acting basal insulin analog for improved metabolic control *7995*.

(326) Lepore, M., Pampanelli, S., Fanelli, C., Porcellati, F., Bartocci, L., Vincenzo, A. Di, Cordoni, C., Costa, E., Brunetti, P., and Bolli, G. B. (2000) Pharmacokinetics and Pharmacodynamics of Subcutaneous Injection of Long-Acting Human Insulin Analog Glargine , NPH Insulin , and Ultralente Human Insulin and Continuous Subcutaneous Infusion of Insulin Lispro *49*.

(327) Bolli, G. B., and Owens, D. R. (2000) Insulin glargine. *Lancet (London, England) 356*, 443–5.

(328) Rosenstock, J., Schwartz, S. L., Clark, C. M., Park, G. D., Donley, D. W., and Edwards, M. B. (2001) Basal insulin therapy in type 2 diabetes: 28-week comparison of insulin glargine (HOE 901) and NPH insulin. *Diabetes Care 24*, 631–6.

(329) Farady, C. J., and Craik, C. S. (2010) Mechanisms of macromolecular protease inhibitors. *Chembiochem 11*, 2341–6.

(330) Bode, W., and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem. 204*, 433–51.

(331) McBride, J. D., Brauer, A. B., Nievo, M., and Leatherbarrow, R. J. (1998) The role of threonine in the P2 position of Bowman-Birk proteinase inhibitors: studies on P2 variation in cyclic peptides encompassing the reactive site loop. *J. Mol. Biol. 282*, 447–58.
(332) Stoop, A. A., and Craik, C. S. (2003) Engineering of a macromolecular scaffold to develop specific protease inhibitors. *Nat. Biotechnol. 21*, 1063–1068.
(333) Brauer, A. B. E., Kelly, G., Matthews, S. J., and Leatherbarrow, R. J. (2002) The1H-NMR solution structure of the antitryptic core peptide of Bowman-Birk inhibitor proteins: A minimal "canonical loop." *J. Biomol. Struct. Dyn. 20*, 59–70.
(334) Laskowski, M., and Kato, I. (1980) Protein inhibitors of proteinases. *Annu. Rev. Biochem. 49*, 593–626.
(335) Jendrny, C., and Beck-Sickinger, A. G. (2016) Inhibition of Kallikrein-Related Peptidases 7 and 5 by Grafting Serpin Reactive-Center Loop Sequences onto Sunflower Trypsin Inhibitor-1 (SFTI-1). *ChemBioChem 17*, 719–726.
(336) Scott, C. J., and Taggart, C. C. (2010) Biologic protease inhibitors as novel therapeutic agents. *Biochimie 92*, 1681–1688.
(337) Ondetti, M. A., Rubin, B., and Cushman, D. W. (1977) Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. *Science 196*, 441–4.
(338) Abbenante, G., and Fairlie, D. P. (2005) Protease inhibitors in the clinic. *Med. Chem. 1*, 71–104.
(339) Collier, A. C., Coombs, R. W., Schoenfeld, D. A., Bassett, R. L., Timpone, J., Baruch, A., Jones, M., Facey, K., Whitacre, C., McAuliffe, V. J., Friedman, H. M., Merigan, T. C., Reichman, R. C., Hooper, C., and Corey, L. (1996) Treatment of Human Immunodeficiency Virus Infection with Saquinavir, Zidovudine, and Zalcitabine. *N. Engl. J. Med. 334*, 1011–1018.
(340) Gulick, R. M., Mellors, J. W., Havlir, D., Eron, J. J., Gonzalez, C., McMahon, D., Richman, D. D., Valentine, F. T., Jonas, L., Meibohm, A., Emini, E. A., and Chodakewitz, J. A. (1997) Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N. Engl. J. Med. 337*, 734–9.
(341) Viswanathan, M., Comeau, S. R., and Ladner, R. C. (2009) Engineered Protein Protease Inhibitors 87–98.
(342) Cohen, I., Kayode, O., Hockla, A., Sankaran, B., Radisky, D. C., Radisky, E. S., and Papo, N. (2016) Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem. J. 473*, 1329–1341.
(343) Salameh, M. A., Soares, A. S., Hockla, A., Radisky, D. C., and Radisky, E. S. (2011) The P(2)' residue is a key determinant of mesotrypsin specificity: engineering a high-affinity

inhibitor with anticancer activity. *Biochem. J. 440*, 95–105.
(344) Riley, B. T., Ilyichova, O., Costa, M. G. S., Porebski, B. T., De Veer, S. J., Swedberg, J. E., Kass, I., Harris, J. M., Hoke, D. E., and Buckle, A. M. (2016) Direct and indirect mechanisms of KLK4 inhibition revealed by structure and dynamics. *Sci. Rep. 6*, 1–14.
(345) Song, H. K., and Suh, S. W. (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J. Mol. Biol. 275*, 347–63.
(346) Rubin, H., Wang, Z. M., Nickbarg, E. B., McLarney, S., Naidoo, N., Schoenberger, O. L., Johnson, J. L., and Cooperman, B. S. (1990) Cloning, Expression, Purification and Biological Activity of Recombinant Native and Variant Human a1-Antichymotrypsin. *Journal 265*, 1199–1207.
(347) Ferdjani, S., Ionita, M., Roy, B., Dion, M., Djeghaba, Z., Rabiller, C., and Tellier, C. (2011) Correlation between thermostability and stability of glycosidases in ionic liquid. *Biotechnol. Lett. 33*, 1215–1219.
(348) Gao, D., Narasimhan, D. L., Macdonald, J., Brim, R., Ko, M.-C., Landry, D. W., Woods, J. H., Sunahara, R. K., and Zhan, C.-G. (2009) Thermostable variants of cocaine esterase for long-time protection against cocaine toxicity. *Mol. Pharmacol. 75*, 318–23.
(349) Kintzing, J. R., Filsinger Interrante, M. V., and Cochran, J. R. (2016) Emerging Strategies for Developing Next-Generation Protein Therapeutics for Cancer Treatment. *Trends Pharmacol. Sci. 37*, 993–1008.
(350) Marshall, S. A., Lazar, G. A., Chirino, A. J., and Desjarlais, J. R. (2003) Rational design and engineering of therapeutic proteins. *Drug Discov. Today 8*, 212–221.

References

Appendix

www.nature.com/scientificreports

OPEN

Smoothing a rugged protein folding landscape by sequence-based redesign

Received: 22 June 2016
Accepted: 01 September 2016
Published: 26 September 2016

Benjamin T. Porebski[1,2,*], Shani Keleher[1,*], Jeffrey J. Hollins[3], Adrian A. Nickson[3], Emilia M. Marijanovic[1], Natalie A. Borg[1], Mauricio G. S. Costa[4], Mary A. Pearce[1], Weiwen Dai[1], Liguang Zhu[5], James A. Irving[6], David E. Hoke[1], Itamar Kass[1], James C. Whisstock[1,7], Stephen P. Bottomley[1], Geoffrey I. Webb[5], Sheena McGowan[1,8] & Ashley M. Buckle[1]

The rugged folding landscapes of functional proteins puts them at risk of misfolding and aggregation. Serine protease inhibitors, or serpins, are paradigms for this delicate balance between function and misfolding. Serpins exist in a metastable state that undergoes a major conformational change in order to inhibit proteases. However, conformational labiality of the native serpin fold renders them susceptible to misfolding, which underlies misfolding diseases such as α_1-antitrypsin deficiency. To investigate how serpins balance function and folding, we used consensus design to create *conserpin*, a synthetic serpin that folds reversibly, is functional, thermostable, and polymerization resistant. Characterization of its structure, folding and dynamics suggest that consensus design has remodeled the folding landscape to reconcile competing requirements for stability and function. This approach may offer general benefits for engineering functional proteins that have risky folding landscapes, including the removal of aggregation-prone intermediates, and modifying scaffolds for use as protein therapeutics.

The rugged energy landscapes of functional proteins reflect the delicate balance between efficient folding and function[1,2]. For proteins to fold, the interactions of the native state must outweigh the non-native interactions, which result in a funnel-shaped energy landscape[3–5]. However, it is not obvious how the myriad of non-covalent interactions that stabilise the native state can do so selectively over the vastly larger number of non-native conformations. Effective protein engineering has typically focused on stabilising low energy configurations as observed in X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy[6–9]. However, engineering robust proteins with funnel shaped energy landscapes may require not only stabilisation of the native state (positive design)[10–12], but also destabilisation of non-native states (negative design)[12–15]. This is especially true for engineering proteins with complex and rugged folding pathways, which often exhibit a delicate balance between function and misfolding[1,2].

Such a balance is exemplified by members of the serine protease inhibitor, or serpin superfamily[16–19]. Inhibitory members fold to a metastable native state that undergoes a major conformational change in order to inhibit target proteases[20]. The inhibitory mechanism of serpins is structurally well understood[20]. Briefly, a target protease initially interacts with and cleaves the serpin reactive center loop (RCL) that protrudes from the main body of the molecule. Following RCL cleavage, but prior to the final hydrolysis of the acyl enzyme intermediate, the RCL inserts into the central β-sheet to form an extra strand[20,21]. Since the protease is still covalently linked to the serpin, the process of RCL insertion results in the translocation of the protease to the opposite end of the molecule.

[1]Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia. [2]Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom. [3]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom. [4]Programa de Computação Científica, Fundação Oswaldo Cruz, 21949900 Rio de Janeiro, Brazil. [5]Faculty of Information Technology, Monash University, Clayton, Victoria 3800, Australia. [6]Wolfson Institute for Biomedical Research, University College London, Gower Street, London, WC1E 6BT, United Kingdom. [7]ARC Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia. [8]Biomedicine Discovery Institute, Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.M.G. (email: sheena.mcgowan@monash.edu) or A.M.B. (email: ashley.buckle@monash.edu)

In the final complex, the protease active site is distorted and trapped as the acyl enzyme intermediate[20,22]. This remarkable conformational change is termed the stressed [S] to relaxed [R] transition and is accompanied by a major increase in stability of the serpin protein.

As a consequence of folding to a metastable active state, serpins are prone to misfolding. Without being cleaved by a protease, the serpin RCL can self-insert, either partially (delta), or fully (latent)[16]; or polymerize by insertion of the RCL of one serpin into the body of another serpin[23–25]. Both such RCL insertion events result in a more stable protein species that is no longer functional as a protease inhibitor. Misfolding of the archetypal serpin, α1-antitrypsin (α1-AT), results in a deficiency of active protein, inducing emphysema through uncontrolled protease activity, and the retention of α1-AT polymers in the liver that induce cell death[18,26]. Serpin misfolding and serpinopathies are a direct result of the 'risky' energy landscape required to fold the protein to a metastable state[1,27]. To investigate how the folding energy landscape of serpins balances the competing requirements for function and stability we used consensus design to build a synthetic serpin. Consensus design is based on the hypothesis that at a given position in a multiple sequence alignment (MSA) of homologous proteins, the respective consensus amino acid contributes more than average to the stability of the protein than non-consensus amino acids[28–31]. The efficacy of consensus design has been demonstrated to increase the stabilities of a wide range of proteins, usually by stabilising the native state[29,31–38]. However, its potential for altering folding landscapes has not been thoroughly explored[29,31–38].

Hypothesizing that a serpin reflecting a highly conserved sequence may offer insight into the delicate balance between folding and function, we designed *conserpin* (*con*sensus *serpin*). Characterization of its function, structure and folding reveal a serpin that is inhibitory, folds reversibly, is thermostable and resistant to polymerisation. Our results suggest that consensus design has smoothed the folding landscape, reducing the lifetime of aggregation-prone intermediates. This work provides insights into the serpin function-stability balance and emphasises the wider potential for consensus design to remodel the risky folding landscapes of functional proteins.

Results

Conserpin is an inhibitory serpin. To design *conserpin* we used the consensus approach and a previously reported MSA of 219 serpin sequences[39]. Conserpin (396 aa) shares the highest similarity with α1-AT (137 residue differences; 62% sequence identity). There is an overall loss of 10 residues located at the N-terminus of the D-helix and C-terminus of the protein. The RCL contains 7 residue differences compared to α1-AT, notably an arginine at P1 compared to the methionine of α1-AT, and the deletion of a residue at P2.

Purified conserpin inhibits trypsin with a stoichiometry of inhibition (SI) of 1.8 and a k_{ass}^{app} of $7.5 \times 10^6 M^{-1} s^{-1}$ and hence a rate of association (k_{ass}) of $1.4 \times 10^7 M^{-1} s^{-1}$ (Fig. S1A–C). Higher order complex formation of conserpin with trypsin was observed on SDS PAGE; however, it was atypical compared to α1-AT (Fig. S1D). This unusual behaviour and the increased SI of conserpin may be a consequence of shortening the RCL on the 'prime' side of the recognition sequence for trypsin, or due to other biophysical differences. The crystal structure of conserpin (Table S1), confirms that it adopts the archetypal native serpin fold (Fig. 1A,B). Taken together, we propose that inhibition by conserpin occurs via the classical serpin mechanism.

Conserpin folds reversibly, is thermostable and resistant to polymerization. The majority of serpins unfold through an aggregation-prone intermediate ensemble and do not completely refold after chemical and/or thermal denaturation[40–47]. This is exemplified by α1-AT, which shows a very small amount of refolded monomer via chemical denaturation, rapid dilution and gel filtration (Fig. 1C). In contrast, conserpin refolds to a monomeric state (Fig. 1D). Equilibrium chemical unfolding and refolding curves overlay well, revealing a midpoint of denaturation, $[D]_{50}$, of 2.75 ± 0.10 M, an equilibrium *m*-value, $m_{D\text{-}N}$, of 8.45 ± 0.65 kcal mol^{-1} M^{-1}, and hence a stability, $\Delta G_{D\text{-}N}$, of -23.2 ± 2.0 kcal mol^{-1} (Fig. 1E). The correlation of unfolding and refolding curves, the single unfolding transition, and the steep *m*-value all suggest minimal formation of an intermediate ensemble. Refolded conserpin retained inhibitory activity, resulting in no significant change in SI (increased from 1.8 to 2.3, Fig. S1E), confirming that conserpin refolds to the native state after chemical denaturation.

Variable temperature circular dichroism (CD) thermal melt analysis at 222 nm reveals a highly thermostable protein with no defined unfolding transition up to a temperature of 110 °C (Fig. 1H). Far-UV spectral scans before and after the thermal melt showed no change in signal, indicating no detectable heat-induced structural changes (Fig. 1F). This contrasts with α1-AT, which upon heating undergoes a three-state transition with an initial midpoint temperature (T_m) of 61.8 °C and an incomplete transition that starts at 90 °C (consistent with other reports[48,49]; Fig. 1G). Upon cooling of α1-AT, we observed a white precipitate in the cuvette, consistent with irreversible aggregation. Refolding transverse urea gradient (TUG) gels further demonstrate that conserpin is more resistant to polymerization than α1-AT, which mostly formed polymers on refolding, with no formation of native protein (Fig. S2).

To test if conserpin undergoes a transition to the more stable latent state upon heating, we assessed its inhibitory activity and structure after heating at 80 °C for 20 minutes. Heating caused a complete loss in inhibitory activity (Fig. S1F), suggesting formation of the latent state, which was then confirmed by native PAGE (Fig. S3A) and crystal structure determination (Table S1 & Fig. S3B).

Conserpin avoids polymerization by minimizing formation of folding intermediates. Although equilibrium unfolding/refolding data using intrinsic fluorescence indicated minimal formation of intermediates in the folding pathway of conserpin (Fig. 1E), this method is dependent on the difference in solvation of tryptophan residues during unfolding/refolding. In order to more thoroughly interrogate folding intermediates, we repeated the equilibrium unfolding experiments in the presence of bis-ANS (Fig. 1I). In native conditions, both folded conserpin and α1-AT show similar levels of fluorescence, however, by ~ 1 M GuHCl, a high intensity

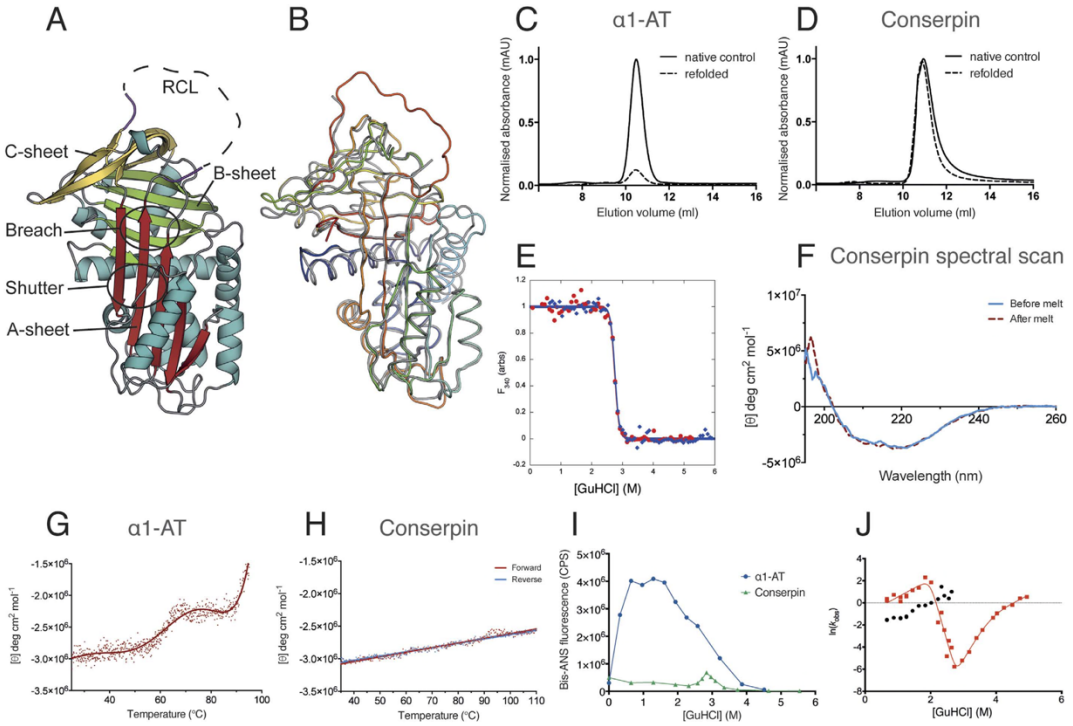


Figure 1. Conserpin conforms to the serpin fold and has superior biophysical properties compared with α1-AT. (A) Cartoon representation of the 2.4 Å X-ray crystal structure of native conserpin, identifying the breach and shutter regions, the A, B and C sheets (colored in red, green and yellow respectively), and the RCL stumps (magenta). **(B)** Structural alignment of conserpin (grey) with α1-AT (PDB: 3NE4; spectrum, blue to red). Root mean square deviation (RMSD) = 0.91 Å across 296 backbone Cα atoms. Chemical refolding of **(C)** α1-AT and **(D)** conserpin shows that conserpin can refold to a monomer. Chromatograms from a Superdex 75 10/300 size exclusion column are shown. Final protein concentrations loaded onto column were 2 μM. Samples were unfolded in 5 M GuHCl and then diluted out to 0.5 M GuHCl (dotted line). Control samples of native protein are shown as the solid black line. **(E)** Intrinsic fluorescence equilibrium unfolding (red dots) and refolding (blue diamonds) curves of conserpin coincide, demonstrating reversible folding. **(F)** CD spectral scans of conserpin before (solid blue line) and after (dashed red line) heating to 110 °C. Variable temperature thermal melts of **(G)** α1-AT and **(H)** conserpin as measured by CD at 222 nm. **(I)** Conserpin shows a significant reduction of intermediate formation during bis-ANS fluorescent equilibrium unfolding of α1-AT (blue circles) and conserpin (green triangles). **(J)** Kinetic unfolding and refolding experiments. The plot shows the [GuHCl]-dependence of the natural logarithm of the rate constants for unfolding and refolding of conserpin (chevron plot). Two discernable refolding rates are observed (red squares, fast rate; black circles, slower folding rate). The positive slope in each refolding arm suggests the presence of intermediate species that have to partially unfold to reach the native state.

fluorescent peak indicated the presence of a folding intermediate(s) for α1-AT (Fig. 1I), consistent with previous reports[47]. In contrast, the unfolding profile of conserpin in bis-ANS shows a small, sharp peak at approximately 3 M GuHCl (Fig. 1I). This is consistent with our $[D]_{50}$ measurement by intrinsic fluorescence (Fig. 1E) and confirms our hypothesis that conserpin has reduced intermediate ensemble formation.

To observe the kinetics of the folding intermediate, we used rapid mixing techniques during unfolding and refolding (Fig. 1J). As expected, the unfolding traces fitted well to a single exponential. When the protein was refolded from an equilibrated denatured solution (single-jump), the resulting traces could not be fitted to fewer than three exponentials (Fig. S1G) and showed inconsistencies between repeats (likely due to aggregate from previous runs). However, unfolding native conserpin followed by refolding (double-jump) resulted in more consistent refolding traces that fitted to a double exponential (SI methods; Fig. S1H). Both single and double jump refolding identified two rates that were independent of the delay time (Figs S1I,J and 1J). Plausible explanations for the presence of two refolding rates are: two denatured states folding on different timescales (e.g. folding limited by proline isomerisation); a fast rate of refolding to an intermediate, followed by a slow rate of refolding from that intermediate; or two fluorophores reporting on independent folding events (e.g. two independently nucleating subdomains). In our data, it is most likely that we are detecting folding from two similarly structured ground states. If we were observing a fast rate, followed by a slow rate, we should expect the fast rate to become kinetically

invisible when the two rates cross (~2 M GuHCl), which it does not. Similarly, if there are two independent folding events, then the relative amplitudes of each rate should be consistent, which they are not. Most interestingly, the refolding *m-values* are positive at low concentrations of denaturant (<2 M), suggesting that the two populated ground states are more structured than the subsequent folding transition state(s). Therefore, the starting states cannot be denatured states, and must be structured intermediates (I_1 and I_2) that fold on different time-scales (Figs 1J and S1I,J). The fast folding rate (red squares) matches up with the unfolding rate at the expected $[D]_{50\%,}$ (2.75 M), verifying that this rate shows folding over the major transition state. The "rollover" in this rate demonstrates that the first intermediate I_1 is in rapid pre-equilibrium with the denatured state (D) and there is a switch in ground state from I_1 to D when the two species are of equal stability (2 M GuHCl, red squares in Fig. 1J). The second intermediate I_2 (Fig. 1J, black circles) shows an almost identical folding *m*-value and, assuming this also folds over the major transition state, is likely to be very similar in structure to I_1. However, I_2 is more stable than I_1 and persists until the denaturant midpoint (2.75 M). As such, we propose that I_1 is likely to be the previously observed polymerogenic folding branch point[19,40,47,50–53]. As this species is highly aggregation prone in other serpins, it is possible that the second intermediate (I_2) is a multimer of the first intermediate.

Global structural features of native conserpin are not typical for a thermostable protein. Our data imply that the folding reversibility and low polymerization propensity of conserpin is due to alteration of the folding landscape, resulting in minimal formation of a folding intermediate. Comparison of native and latent state conserpin structures with available native, latent and cleaved structures of α1-AT[54–56], plasminogen activator inhibitor 1 (PAI-1)[57,58], α1-antichymotrypsin (ACH)[59], neuroserpin[60,61], antithrombin[62] and the thermostable serpins, thermopin[63] and tengpin[64] reveal that despite having the highest thermostability, native conserpin has the fewest H-bonds and salt bridges (Table S2). Further, native state conserpin has the largest accessible surface area and largest solvent inaccessible cavity volume of all assessed serpins. These characteristics are unusual for thermostable proteins, which typically feature more interactions and optimized packing compared to their mesophilic counterparts[37,65–73]. Comparison of the electrostatic surface potential of conserpin with that of α1-AT reveals minor differences on the surface-exposed face of the A-sheet, whilst the opposite face of the molecule is substantially more positively charged (Fig. 2A), consistent with the reported aggregation resistance of proteins featuring increased electrostatic surface potential[74,75]. In contrast to the majority of mesophilic proteins and their thermophilic homologues, a correlation between overall number of H-bonds/salt bridges and thermostability is not apparent for serpins, which must balance the relative stabilities of native and RCL-inserted states to enable unique conformational plasticity underpinning inhibitory function[63,64] (Table S2). This reasoning suggests that more subtle, context-dependent structural and dynamical features play a more dominant role in conserpin, which we explore next.

Favorable interactions and reduced dynamics surrounding the D-helix. Given the conformational plasticity required for serpin function, we next performed molecular dynamics (MD) simulations for 0.5 μs at 300 K in triplicate for both conserpin and α1-AT. Both systems reach equilibrium by 150 ns (Fig. S4A). Although the increased mobility of the RCL and the C-terminus of hA of conserpin leads to a higher overall RMSD, inspection of root mean square fluctuations (RMSFs) shows conserpin to exhibit an overall reduction in dynamics in the majority of regions, specifically the extended N-terminus of hA, hC/hD loop, hD, hE, hF, hG, hH (Fig. S4B,C). This is further supported by a large reduction in conformational sampling as shown by principle component analysis (Fig. S4D). The most notable reduction in dynamics is in the D-helix (hD; RMSD of 0.58 vs. 1.65 Å; Figs 2B and S4B,C). The D-helix of α1-AT has been implicated in stability; notably two mutations ($T114F_{\alpha1\text{-}AT}$ and $G117F_{\alpha1\text{-}AT}$) stabilize the D-helix and rescue the polymerogenic Z-variant[76,77]. The D-helix of conserpin is shortened by the deletion of five residues, four at the N-terminus (L84, E86, I87 and P88 in α1-AT) and one at the C-terminus (Q109 in α1-AT; Figs 2B & S5A). The deletion of $L84_{\alpha1\text{-}AT}$ and $I87_{\alpha1\text{-}AT}$ reduces overall hydrophobicity without affecting the packing of hD against the core of conserpin (Fig. S5A). Residue numbering will adhere to the following convention unless explicitly stated: $Q105_{\alpha1\text{-}AT}$ or $R79_{conserpin}$ or $Q105R_{79}$, where Q105 from α1-AT has been mutated to an R, which is residue number 79 in conserpin.

The rigidity of hD in conserpin is probably due to a salt bridge between $Q105R_{79}$ of hD and $E376_{346}$ and interactions of the N-terminus with hD. The salt bridge between the B-sheet and hD is present throughout the MD simulation and possibly stabilizes the top of the D-helix (Fig. 2B). In contrast, there are no similar salt bridges in the α1-AT crystal structure or during MD (Fig. 2B). Rather, hD in α1-AT undergoes conformational rearrangement and loss of secondary structure in one of the replicates (Fig. S5B). This is consistent with other reports, which indicates that minor changes to hD may accelerate or reduce polymer formation[76,77]. The N-terminus in conserpin is extended by the addition of a purification tag. Four residues of the extension were resolved in the crystal structure and a single H-bond is observed between the backbone of residue $A\text{-}1_{conserpin}$ and the N-terminus of hD ($D65_{conserpin}$; Fig. 2C). This H-bond is persistent throughout MD and extends to form a small β-sheet (Fig. 2D). Therefore, the extended N-terminus may impart stability to hD and may reflect similar interactions seen in the naturally extended N-termini of thermophilic serpins[63,64]. Taken together, our observations suggest that optimized interactions in and around hD increase the stability of the native state.

The electrostatic network of the serpin breach region is extended in conserpin. The breach region, consisting of a highly conserved electrostatic network between residues $E342_{\alpha1\text{-}AT}$, $K290_{\alpha1\text{-}AT}$ and $D341_{\alpha1\text{-}AT}$ at the top of the A-sheet is important for controlling the conformational change that drives protease inhibition[39,49,79]. This network is significantly extended in conserpin, compared to α1-AT (Fig. 3A). Specifically, the mutations of $T339E_{310}$ and $S292K_{264}$ contribute to a salt bridge network spanning s3A, s5A and s6A with $K191_{163}$. $T294E_{266}$ also forms a new salt bridge with $K335_{306}$ between s6A and s5A, whilst $D341N_{312}$ mediates an unfavorably charged cluster of $E310_{conserpin}$, $E313_{conserpin}$ and $E314_{conserpin}$ that is not present in α1-AT (Fig. 3A).

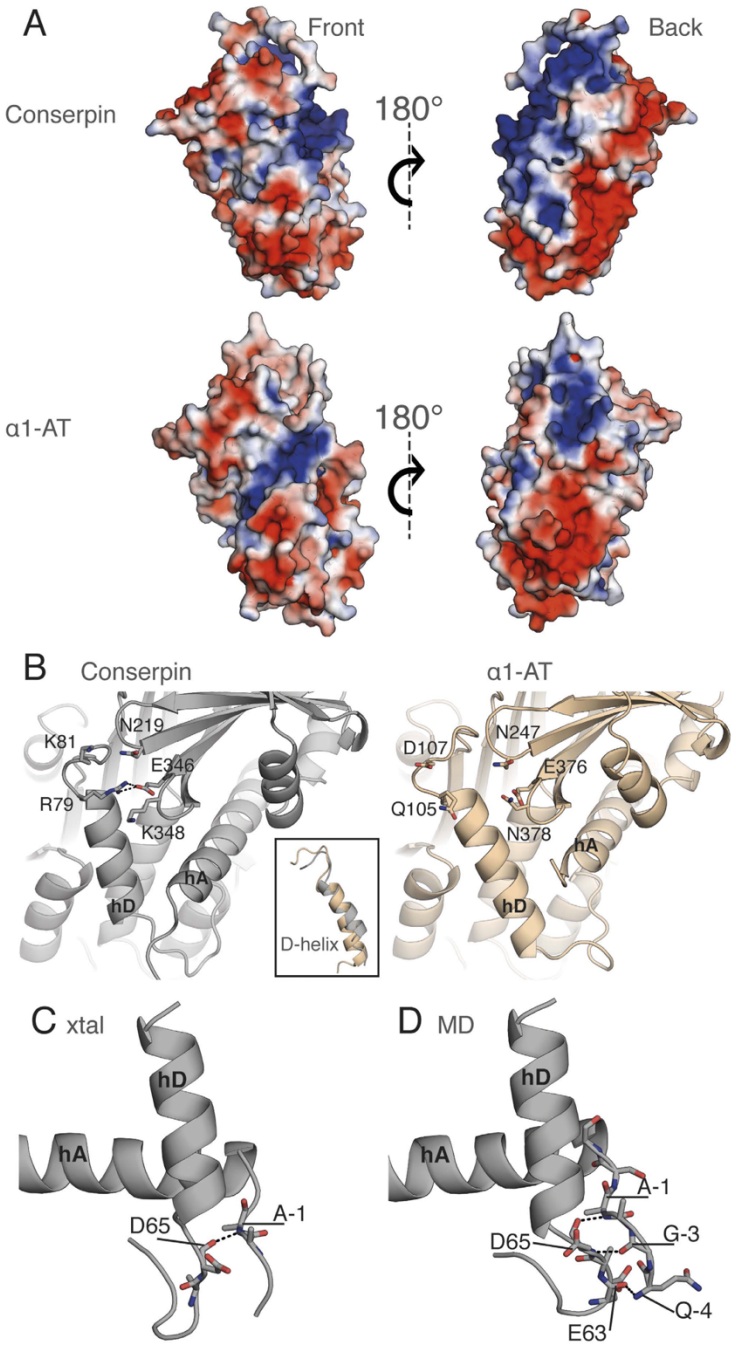


Figure 2. Structural analysis reveals alterations of the electrostatic surface and stabilization of the D-helix in conserpin. (A) The electrostatic potential surface of conserpin and α1-AT models (blue = +ve, red = −ve), in the same orientation as Fig. 1A (front) and a 180° rotation reveals an overall increase in positive charge on the back face of conserpin. **(B)** The introduced salt bridge in hD of conserpin with residues $Q105R_{79}$ and $E376_{346}$. There is no comparable interaction present in α1-AT. Inset shows the shortened D-helix in conserpin. **(C)** H-bonding between A-1 of the extended N-terminus and D65 of hD, as seen in the conserpin crystal structure. **(D)** Persistent hydrogen bonding between Q-4, G-3 and A-1 of the extended N-terminus and E63 and D65 of hD in conserpin as seen in MD simulation.

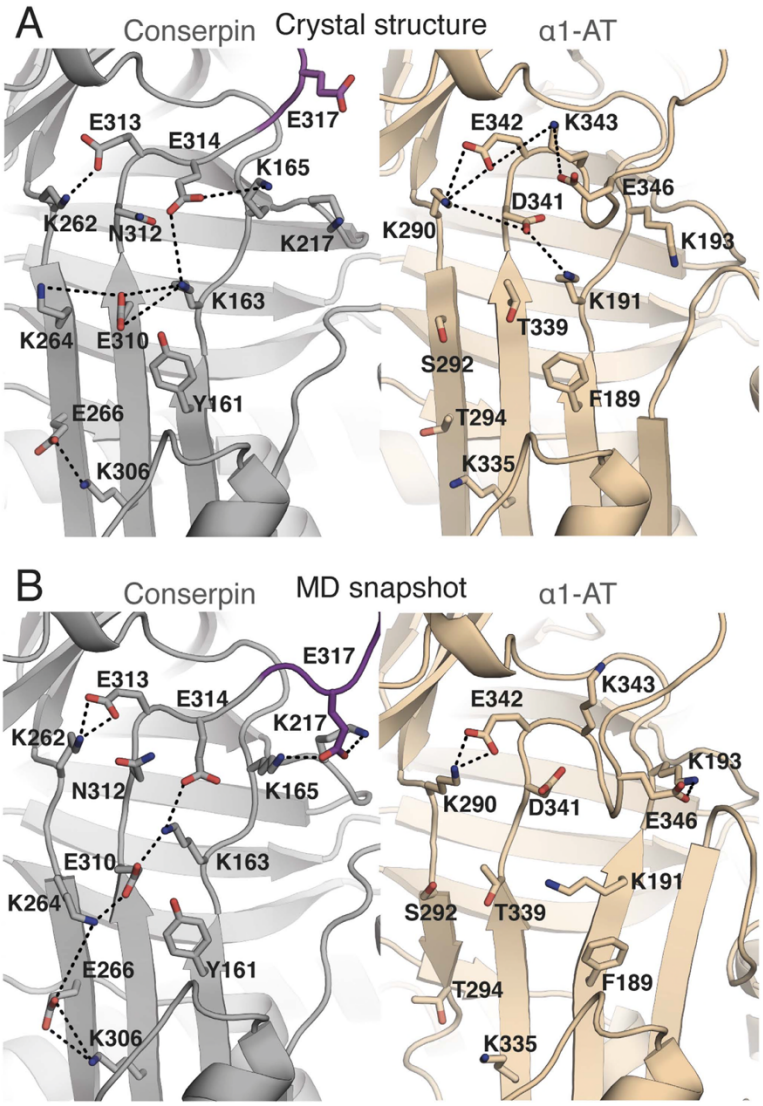


Figure 3. The electrostatic network of the breach region is extended in conserpin. (A) A-sheet salt bridge interactions (dashed lines) in the crystal structures of conserpin (carbon atoms in grey) and α1-AT (carbon atoms in wheat; PDB: 3NE4). **(B)** A simulation snapshot taken at 500 ns, showing A-sheet salt bridge interactions as described above. The modeled RCL of conserpin is colored magenta.

These observations are interesting in the context of serpin polymerization, which involves insertion of the RCL and/or s5A from one molecule into the flexible A-sheet of another[19,23,24,40,51,78]. In particular, the disease-causing Z-variant, $E342K_{\alpha1\text{-}AT}$ induces repulsion with $K290_{\alpha1\text{-}AT}$, which either retards the formation of the A-sheet during folding, increasing the lifetime of the polymerogenic intermediate ensemble, or destabilizes the structure and increases the dynamics of the native state, allowing for s5A and s6A to separate, and reduce the energy barrier for polymerization[19,40,80–82].

It is difficult to ascertain the effect of the extended salt bridge network on A-sheet dynamics during folding of conserpin as MD simulations only describe the dynamics of the native state. Nevertheless, simulations of α1-AT reveal its A-sheet salt bridge network to be weaker over time (Fig. 3B), allowing for the transient separation of strands s5A and s3A (Fig. S6). In contrast, the extensive network in conserpin is present throughout the majority of the simulation, with additional interactions being formed, notably an alternate conformation of $K264_{conserpin}$ mediating interactions between $E310_{conserpin}$, and $E266_{conserpin}$ (Fig. 3B). Furthermore, $E317_{conserpin}$ in the RCL of conserpin is able to adopt a stable conformation, mediating the salt-bridge between $K165_{conserpin}$ and $K217_{conserpin}$, with $K165_{conserpin}$ forming transient interactions to $E314_{conserpin}$. Equilibrium and kinetic folding studies of α1-AT

provide compelling evidence for the late folding of s5A during transition through the polymerogenic intermediate state[19,40]. Taken together, our observations suggest an improved energetically stable conserpin native state with possible increases to the folding cooperativity in this region, which may also be augmented by the hydrophobic core behind the A-sheet.

Biophysical and structural analysis of Z-conserpin. Intrigued by the stabilizing electrostatic interactions in the breach region of conserpin, we assessed the effect of introducing the disease-causing Z-mutation, $E342K_{313}$, into conserpin. Mutation of $E342_{\alpha1\text{-AT}}$ to a lysine results in an increased propensity of α1-AT to polymerize in the endoplasmic reticulum of hepatocytes, leading to a lack of secretion into the circulation[26]. Studying the effects of the Z-variant in α1-AT is difficult due to expression as insoluble aggregate[76,83]. The most likely mechanism of Z-variant polymerization involves perturbation of the folding energy landscape, thus increasing the lifetime of the polymerogenic intermediate ensemble[24,40,81,83]. Although there is evidence to suggest that the Z-mutation also results in structural and dynamic changes to the native state[80,82,84], a recent crystal structure of Z α1-AT shows minimal perturbation in comparison to wild-type[83]. In order to investigate the effects of a highly destabilizing mutation on conserpin, we introduced $E342K_{313}$ into conserpin to produce Z-conserpin.

Z-conserpin expressed well as a soluble monomer in *E. coli*, which has not been possible with Z α1-AT[76,83]. Z-conserpin showed a highly similar inhibitory profile to conserpin, with an SI of 2.3 and a k_{ass} of $2.1 \times 10^7\,M^{-1}\,s^{-1}$ (Fig. S7A,B). Z-conserpin exhibits reversible, two-state folding upon chemical denaturation (Fig. S7C,D). The equilibrium unfolding and refolding curves overlay almost perfectly, revealing a midpoint of denaturation, $[D]_{50}$ to be 2.51 ± 0.01 M, an equilibrium *m*-value, $m_{D\text{-}N}$, of 5.18 kcal mol^{-1} M^{-1}, and a stability $\Delta G_{D\text{-}N}$, of −12.8 kcal mol^{-1} (a loss of −10.04 kcal mol^{-1}; Fig. S7D). As with conserpin, equilibrium data did not reveal the presence of an intermediate species. We therefore repeated equilibrium unfolding using bis-ANS fluorescence, detecting a fluorescent peak at ~2.5 M GuHCl, that is slightly broader and more intense than observed in conserpin, indicating an increase in the intermediate ensemble population, but still smaller than in α1-AT (Fig. S7E). Variable-temperature far-UV CD melting curves in 2 M GuHCl gave a T_m of 60.7 °C (conserpin $T_m = 72.5$ °C; (Fig. S7F)). Native PAGE shows conserpin to remain monomeric except when heated to 90 °C for 10 minutes, whilst Z-conserpin has a complete loss of monomer at 80 °C and forms a slightly higher molecular weight species when heated to 70 °C for 10 minutes (Fig. S7G). The crystal structure of native Z-conserpin (Table S1) reveals almost no structural differences upon mutation (backbone RMSD = 0.23 Å); the sole differences surrounding $E342K_{313}$ are small side-chain shifts of $K342_{313}$ and $K290_{262}$, most likely as a result of electrostatic repulsion (Fig. S7H). A caveat is one local residue difference, $K343E_{314}$, in conserpin that may partially negate the effects of $E342K_{313}$, due to its salt bridge with $K165_{conserpin}$. As such, future studies of the double mutant $E342K_{313}$/$E314K_{conserpin}$ would be insightful. Regardless, the structure of Z-conserpin reveals essentially no structural changes to the native state which disagrees with reports of structural perturbations within the native state[80,82,84]; therefore favoring the mechanism of Z-variant polymerization via a folding intermediate[83]. However, the intermediate versus native state polymerization mechanisms may be reconciled if the intermediate ensemble is native-like in structure, consistent with our kinetic (un)folding data for conserpin. Considering the evidence in support of this for a wide range of proteins[85], our data is therefore consistent with the Z-mutation altering the folding energy landscape, possibly by lowering the kinetic barrier of the unfolding transition to the polymerogenic intermediate ensemble[81,83].

Importance of A-sheet/F-helix hydrophobic core packing. The hydrophobic core buried by the A-sheet is important for serpin stability[48,86,87]. Amongst 19 mutations designed to probe the stability of α1-AT, seven mutations in the hydrophobic core were found to be stabilizing[86]. Four of these mutations are found in conserpin ($T59S_{37}$, $T68A_{46}$, $A70G_{48}$ and $M374I_{244}$). In the remaining three mutations, the local environment adapts to improve packing and local interactions (Fig. S8).

Packing between hF and the A-sheet also stabilizes the native serpin state, with hF acting as a physical barrier for RCL insertion into the A-sheet during protease inhibition and polymerization[19,88–91]. Conserpin contains three mutations in this region (Fig. 4A); $Y187A_{159}$ and G115A[88], which allow s2A to more tightly pack against hF, and $Y160W_{132}$, which further improves the packing density (Fig. 4B). This is consistent with mutagenesis studies of α1-AT, where Y160A resulted in a 5 °C decrease in T_m and was attributed to the loss of a hydrogen bond and formation of a cavity[88]. In contrast, Y160W raised the T_m of α1-AT to 65 °C, and slowed the rate of polymerization[88]. MD reveals hF of conserpin to be slightly less flexible than that of α1-AT, with $W160_{132}$ remaining conformationally locked compared to Y160 of α1-AT, which frequently flips in and out of the hydrophobic pocket (Fig. 4C). Interactions within the "clasp" motif at the F-helix are structurally conserved in conserpin and maintained throughout simulation, consistent with its proposed role in regulating conformational change[92]. Taken together, these changes likely contribute to the stability of the native state.

Remodeling the B/C barrel, a folding nucleus. Formation of the B/C barrel is thought to occur early in the folding pathway of α1-AT, preceding formation of the A-sheet and acting as a "kinetic trap" that captures the RCL and prevents folding to other more stable states[19,22,40,55,93]. Conserpin contains several mutations in the B/C barrel that improve hydrophobic packing and form favorable interactions within the native state (Fig. 5A). Specifically, $F275W_{247}$ and $E279L_{251}$ allow tighter packing of hH. The introduction of a small salt-bridge network between $K274_{246}$, $C232D_{204}$ and $K234E_{206}$ in hH may further stabilise the hydrophobic core of the B/C barrel (Fig. 5A). Conserpin harbors two potentially destabilizing mutations, but surrounding mutations have compensatory effects: the known destabilizing mutation $F366A_{336}$[40,55], which in isolation would create a destabilizing cavity, is compensated by the mutation $V364F_{334}$ and the introduction of a coordinated salt-bridge network between $D256_{228}$, $E257_{229}$, $K368R_{338}$ and $N367D_{337}$ (Fig. 5B); the potentially destabilizing mutation $W238K_{210}$, which would likely weaken hydrophobic packing and introduce a large cavity, is offset by backbone polar contacts with

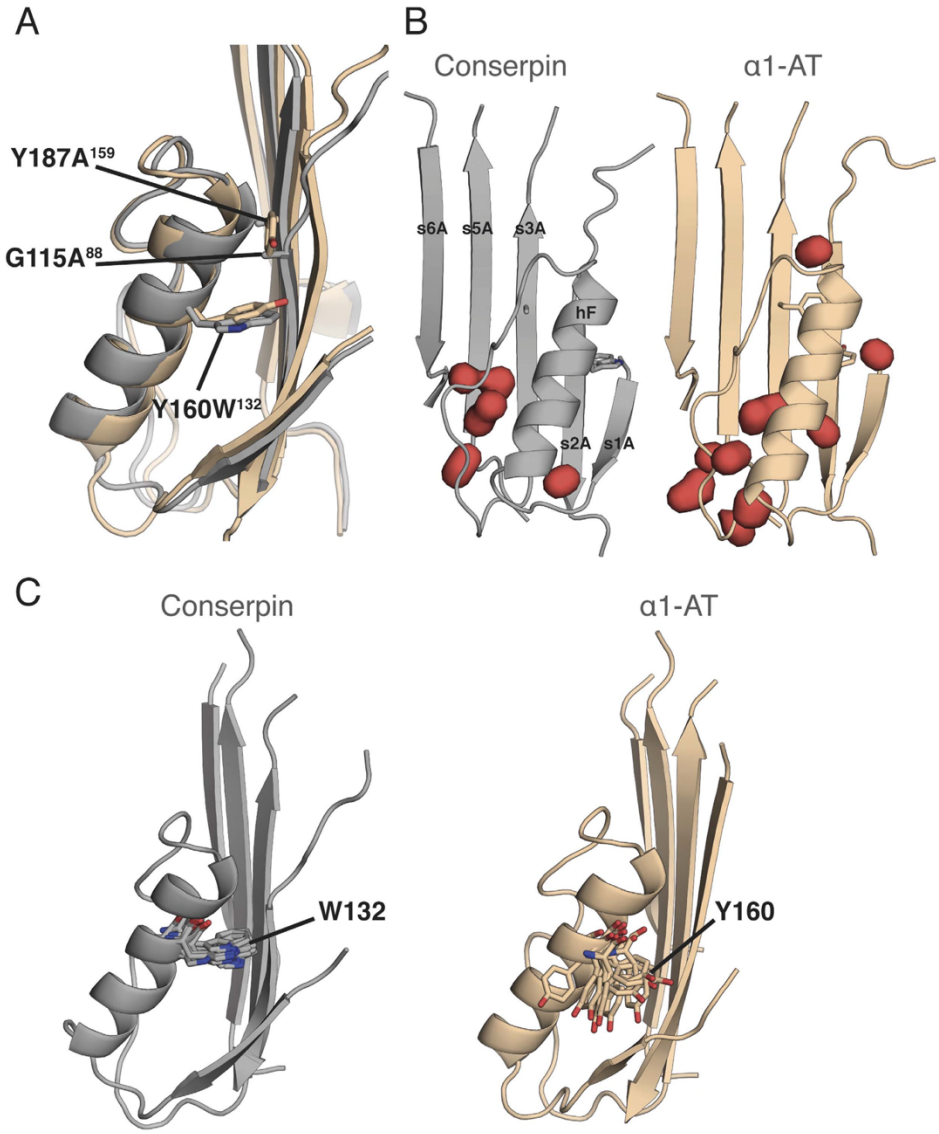


Figure 4. W160 stabilizes hF in conserpin. (A) A structural overlay of hF in conserpin (grey) and α1-AT (wheat), highlighting the positions of $Y160W_{132}$, $Y187A_{159}$ and $G115A_{88}$. **(B)** Solvent inaccessible cavities (red blobs) surrounding hF of conserpin and α1-AT. $Y160W_{132}$ reduces cavity volumes from 233.8 to 120.9 Å^3. **(C)** MD simulation frames (every 50 ns), highlighting the dynamic differences of W132 in conserpin and Y160 in α1-AT.

$E363_{333}$, and together with $I229Y_{201}$ and $A284V_{256}$ may function as a solvent barrier that shields the hydrophobic core (Fig. 5B). MD simulation also indicates a transient salt bridge between $W238K_{210}$ and $D256_{228}$. Conserpin also contains $L224K_{196}$ and $S285E_{257}$, which staples s2C and s3C together, further stabilizing the native state (Fig. 5B). Finally, $L241E_{213}$ and $N228Y_{200}$ are close to the B-sheet hydrophobic core and the region in which citrate was found to bind and stabilize α1-AT, thus potentially providing extra stability[93]. Taken together, these features may contribute to core nucleation rates during early protein folding, as well as native state resistance to unfolding, consistent with our unfolding and refolding data (Fig. 1).

Conserpin is less frustrated than α1-AT. We next investigated the distribution of energetic frustration within the structures of conserpin and α1-AT using the frustratometer webserver[94,95]. As proteins are thought to be minimally frustrated polymers with rugged energy landscapes, the degree of energetic frustration is related to the description of the proteins energy landscape[2,27,94]; that is, a high level of frustration implies flexibility and

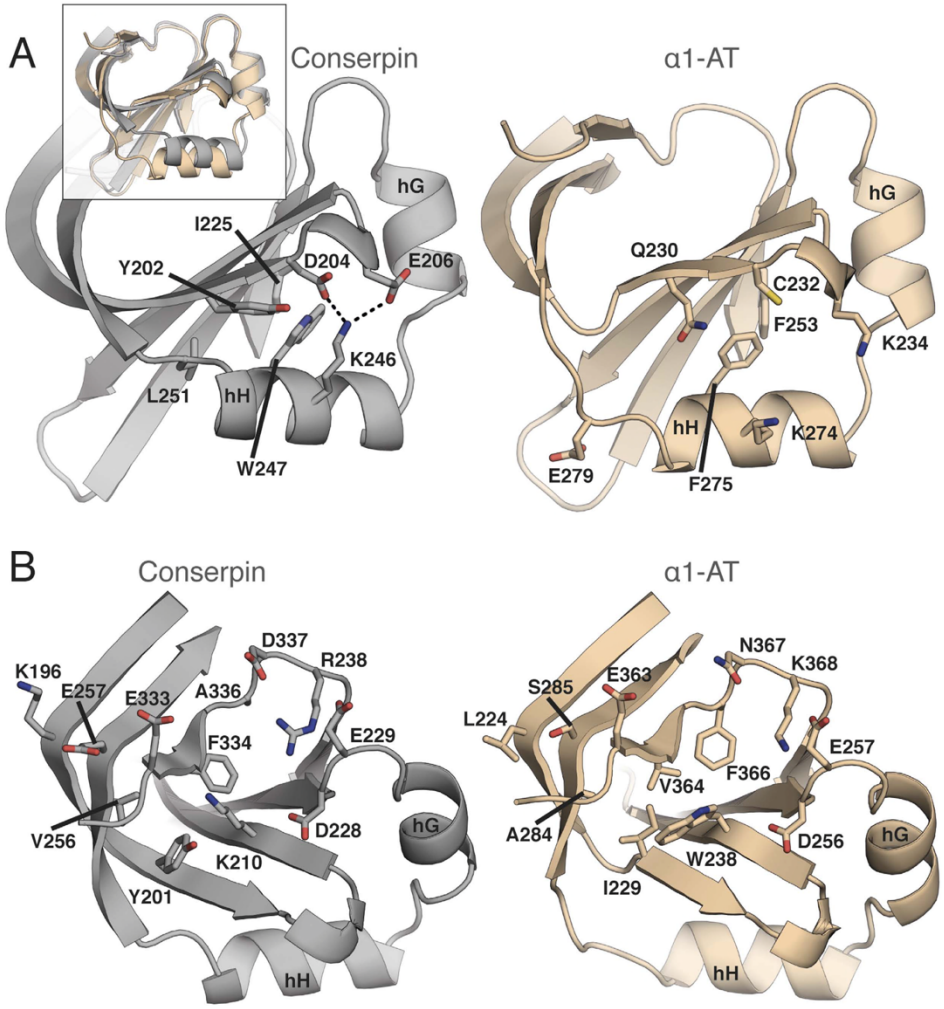


Figure 5. Structural analysis of the B/C barrel in conserpin (grey) and α1-AT (wheat). (A) Stabilizing hydrophobic mutations surrounding $F275W_{247}$. **(B)** Remodeling of the inner barrel surrounding $W238K_{210}$.

a more rugged energy landscape. We therefore used configurational frustration analysis which describes interactions with respect to structural decoys that may be encountered during the folding process[94,95]. Overall, conserpin is less frustrated than α1-AT in several regions, with the exception of the RCL, which is shown to be more dynamic during MD simulation (Fig. 6 and S4B,C). In combination with MD simulation, these results show an inverse correlation between the degree of frustration and degree of dynamics, which is mediated by electrostatic effects (Figs 6 and S4B,C). By modulating the electrostatic constant (k) from 4.15 to 16.6, it becomes apparent that long-range interactions are essential to the reduced frustration of conserpin, but are not as significant in α1-AT (Fig. 6). This is particularly noticeable for helix D and F, which have fewer highly frustrated contacts in conserpin (Fig. 6). The improved folding properties and increased conserpin stability may therefore be related to the higher number of charged residues that stabilise local contacts and introduce repulsion between patches enriched in like charges, which in turn must be correctly oriented in unfolded forms to avoid aggregation[96].

Discussion

The puzzle of how the folding polypeptide chain of serpins achieves a metastable native state has proven challenging to solve. Their unusual and complicated mechanism of protease inhibition challenges the characterization of their folding pathway[42]. It has been established that α1-AT, ACH and PAI-1 all unfold from their native states via an aggregation-prone intermediate ensemble[19,22,40,42,43,47,53,55,89,97]. More recent studies of α1-AT revealed relatively fast folding of the core B/C sheet β-barrel followed by much slower formation of the central A β-sheet[40]. These observations were consistent with models of how off-pathway α1-AT polymers form due to a folding "race" between the core barrel and the central β-sheet[19,24]. However, despite two decades of effort, the aggregation-prone

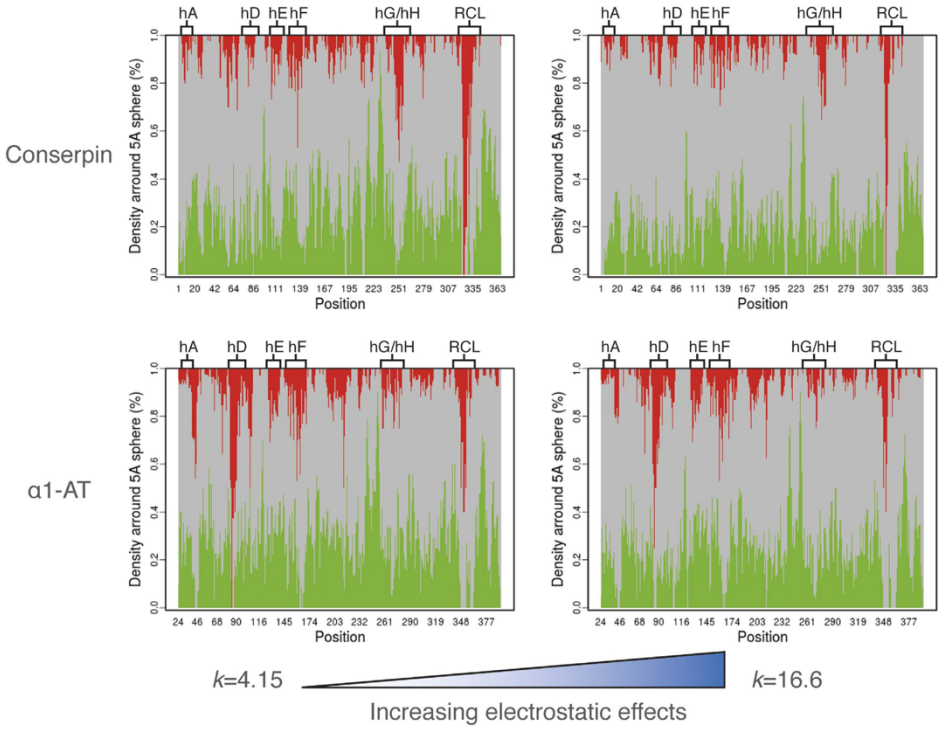


Figure 6. Configurational frustration analysis for conserpin and α1-AT. Minimal, neutral and highly frustrated contacts are represented in green, gray and red respectively. Calculations were performed with different electrostatic strengths by varying the electrostatic constant (*k*). According to ref 95, larger *k* values are related to stronger effects of the Debye–Hückel term.

nature and poor refolding properties of serpins have prevented a full, atomic level characterization of their folding landscape.

Our study reports for the first time, the successful engineering of a reversibly folding serpin that is highly resistant to polymerization and aggregation, even after the introduction of the polymerogenic, disease-causing Z-mutation. Structural analysis reveals the presence of many context dependant and stabilising interactions in regions that are known to be important for folding. These include stabilizing interactions around the D-helix, a salt bridge network in the A-sheet that may resist aberrant RCL insertion, optimization of A-sheet hydrophobic core packing, stabilising mutations in the F-helix that may raise the energy barrier for RCL insertion, and improved packing in the B/C barrel. Although some single mutations in these same regions have been reported to stabilise α1-AT, we found that many mutations within conserpin act together cooperatively. The relatively large accessible surface area and solvent inaccessible cavity volume compared to all other serpins are unusual for a thermostable protein, suggesting that stabilization is achieved by highly context-specific interactions. Although structure and dynamics suggest stabilisation of the native state in key regions, this represents a conundrum: how can function be maintained, which requires metastability, flexibility and conformational change, in combination with a high degree of stability?

Biophysical and structural analysis paints a complex picture. Although native conserpin features fewer overall number of polar contacts compared to other metastable serpins, new, specific interactions stabilise a rigidified native state that is less frustrated than α1-AT. The inhibitory activity of conserpin confirms that its native state retains metastability required for function. The slightly increased SI of conserpin is most likely due its sub-optimal RCL sequence hindering association with the target protease. However, functional impairment might also be caused by a slowing of the rate of insertion of its RCL into the central A β-sheet, which may also contribute to its aggregation resistance. The latent state structure reveals an overall increase in H-bonds and salt bridges. These observations underline the functional importance of maintaining the relative stabilities of the native versus RCL-inserted state (latent/cleaved) irrespective of the specific stabilising features, as observed previously for thermostable serpins[63,64]. However, the structural data do not completely explain the remarkable stability of conserpin. Indeed, the most interesting properties are seen during folding and on exposure to heat; conserpin folds in a concerted fashion, with a relatively minimised population of the aggregation prone intermediate ensemble, and avoids aggregation on heating, with preferential transition to the latent state. Taken together, these results reveal two distinct effects from chemical denaturation and thermal treatment that uniquely provides conserpin with two-state reversible folding, a high degree of thermostability and aggregation resistance. Furthermore, our

findings suggest that the robustness of conserpin folding is due to remodelling of its energy landscape, specifically the smoothing of rugged features that trap aggregation-prone intermediates.

Remodelling of the energy landscape is fascinating from an evolutionary and protein engineering perspective. Consensus design typically accumulates residues important to native state stability[28,29,31,32,34,98], but could equally alter the folding landscape[37,38,99]. As such, conserved features of the energy landscape would be solidified or even amplified, whilst non-conserved features would be minimised. For conserpin, this mechanism implies that aggregation and off-pathway folding events are not conserved across the serpin family, possibly because the functional and regulatory requirements of divergently evolved serpin clades sculpted rugged landscapes as an unfortunate consequence, as may be the case for other functional proteins[1,2]. Therefore, consensus design can potentially smooth the "risky", rugged folding landscapes of functional proteins. This may offer several benefits for protein engineering in general, including the removal of aggregation-prone intermediates and modifying protein scaffolds for use as protein therapeutics and diagnostic reagents. In the case of serpins, our structural and folding data for conserpin and Z-conserpin demonstrate the potential of this engineered scaffold as a model system for studying pathological disease mutations. Finally, the fragile nature of serpin folding has thus far hindered residue-level kinetic characterization of all species on the folding pathway, for example using phi-value analysis[100]; the robustness of conserpin may finally provide the basis for such characterization.

Materials and Methods

To design *conserpin* we used a previously reported MSA of 219 serpin sequences[39] to generate a consensus sequence as described previously[37]. After filtering to remove incomplete sequences and the application of redundancy reduction, we aligned 212 sequences and generated a new protein sequence by selecting the most frequently observed residue at each column of the MSA (the 'consensus method') (Dataset S1). Protein expression and purification, SI measurement and spectroscopic analysis was performed as described previously[89]. Protein Crystallography was performed at the MX1 and MX2 beamlines at the Australian Synchrotron[101]. All experimental and computational methods are described in detail in SI Methods.

References

1. Gershenson, A., Gierasch, L. M., Pastore, A. & Radford, S. E. Energy landscapes of functional proteins are inherently risky. *Nat. Chem. Biol.* **10,** 884–891 (2014).
2. Gianni, S. *et al.* Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. USA* **111,** 14141–14146 (2014).
3. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences* **89,** 8721–8725 (1992).
4. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proceedings of the National Academy of Sciences* **92,** 3626–3630 (1995).
5. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4,** 10–19 (1997).
6. Hilvert, D. Design of protein catalysts. *Annu. Rev. Biochem.* **82,** 447–470 (2013).
7. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38,** 16419–16423 (1999).
8. Schreiber, G., Buckle, A. M. & Fersht, A. R. Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* **2,** 945–951 (1994).
9. Russell, R. J. & Taylor, G. L. Engineering thermostability: lessons from thermophilic proteins. *Current Opinion in Biotechnology* **6,** 370–374 (1995).
10. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A Large Scale Test of Computational Protein Design: Folding and Stability of Nine Completely Redesigned Globular Proteins. *Journal of Molecular Biology* **332,** 449–460 (2003).
11. Kuhlman, B. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302,** 1364–1368 (2003).
12. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).
13. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **99,** 2754–2759 (2002).
14. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* **249,** 884–891 (1990).
15. Jin, W., Kambara, O., Sasakawa, H., Tamura, A. & Takada, S. De novo design of foldable proteins with smooth folding funnel: automated negative design and experimental verification. *Structure/Folding and Design* **11,** 581–590 (2003).
16. Gettins, P. G. W. Serpin Structure, Mechanism, and Function. *Chem. Rev.* **102,** 4751–4804 (2002).
17. Lomas, D. A. & Carrell, R. W. Serpinopathies and the conformational dementias. *Nat Rev Genet* **3,** 759–768 (2002).
18. Law, R. H. P. *et al.* An overview of the serpin superfamily. *Genome Biol.* **7,** 216 (2006).
19. Krishnan, B. & Gierasch, L. M. Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nature Structural & Molecular Biology* **18,** 222–226 (2011).
20. Huntington, J. A., Read, R. J. & Carrell, R. W. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407,** 923–926 (2000).
21. Stratikos, E. & Gettins, P. G. Major proteinase movement upon stable serpin-proteinase complex formation. *Proceedings of the National Academy of Sciences* **94,** 453–458 (1997).
22. Tew, D. J. & Bottomley, S. P. Probing the equilibrium denaturation of the serpin α1-antitrypsin with single tryptophan mutants; evidence for structure in the urea unfolded state. *Journal of Molecular Biology* **313,** 1161–1169 (2001).
23. Yamasaki, M., Li, W., Johnson, D. J. D. & Huntington, J. A. Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* **455,** 1255–1258 (2008).
24. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. & Huntington, J. A. Molecular basis of α1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep.* **12,** 1011–1017 (2011).
25. Ekeowa, U. I. *et al.* Defining the mechanism of polymerization in the serpinopathies. *Proc. Natl. Acad. Sci. USA* **107,** 17146–17151 (2010).
26. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* **357,** 605–607 (1992).
27. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **47,** 285–363 (2014).
28. Lehmann, M. & Wyss, M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology* **12,** 371–375 (2001).

29. Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *Journal of Molecular Biology* **240,** 188–192 (1994).
30. Steipe, B. Evolutionary approaches to protein engineering. *Curr. Top. Microbiol. Immunol.* **243,** 55–86 (1999).
31. Lehmann, M. *et al.* The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **15,** 403–411 (2002).
32. Wang, Q., Buckle, A. M., Foster, N. W., Johnson, C. M. & Fersht, A. R. Design of highly stable functional GroEL minichaperones. *Protein Sci.* **8,** 2186–21193 (1999).
33. Maxwell, K. L. & Davidson, A. R. Mutagenesis of a Buried Polar Interaction in an SH3 Domain: Sequence Conservation Provides the Best Prediction of Stability Effects †. *Biochemistry* **37,** 16172–16182 (1998).
34. Nikolova, P. V., Henckel, J., Lane, D. P. & Fersht, A. R. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl. Acad. Sci. USA* **95,** 14675–14680 (1998).
35. Dai, M. *et al.* The creation of a novel fluorescent protein by guided consensus engineering. *Protein Eng. Des. Sel.* **20,** 69–79 (2007).
36. Jacobs, S. A. *et al.* Design of novel FN3 domains with high stability by a consensus sequence approach. *Protein Eng. Des. Sel.* **25,** 107–117 (2012).
37. Porebski, B. T. *et al.* Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Eng. Des. Sel.* **28,** 67–78 (2015).
38. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29,** 245–251 (2016).
39. Irving, J. A., Pike, R. N., Lesk, A. M. & Whisstock, J. C. Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Research* **10,** 1845–1864 (2000).
40. Tsutsui, Y., Cruz, Dela, R. & Wintrode, P. L. Folding mechanism of the metastable serpin α1-antitrypsin. *Proc. Natl. Acad. Sci. USA* **109,** 4467–4472 (2012).
41. Kwon, K. S., Lee, S. & Yu, M. H. Refolding of alpha 1-antitrypsin expressed as inclusion bodies in Escherichia coli: characterization of aggregation. *Biochim. Biophys. Acta* **1247,** 179–184 (1995).
42. Whisstock, J. C. & Bottomley, S. P. Molecular gymnastics: serpin structure, folding and misfolding. *Current Opinion in Structural Biology* **16,** 761–768 (2006).
43. Wang, Z., Mottonen, J. & Goldsmith, E. J. Kinetically controlled folding of the serpin plasminogen activator inhibitor 1. *Biochemistry* **35,** 16443–16448 (1996).
44. Shirai, N., Tani, F., Higasa, T. & Yasumoto, K. Linear polymerization caused by the defective folding of a non-inhibitory serpin ovalbumin. *J. Biochem.* **121,** 787–797 (1997).
45. Takehara, S. *et al.* Refolding and polymerization pathways of neuroserpin. *Journal of Molecular Biology* **403,** 751–762 (2010).
46. Onda, M. & Hirose, M. Refolding mechanism of ovalbumin: investigation by using a starting urea-denatured disulfide isomer with mispaired CYS367-CYS382. *J. Biol. Chem.* **278,** 23600–23609 (2003).
47. Pearce, M. C., Rubin, H. & Bottomley, S. P. Conformational change and intermediates in the unfolding of alpha 1-antichymotrypsin. *J. Biol. Chem.* **275,** 28513–28518 (2000).
48. Kwon, K. S., Kim, J., Shin, H. S. & Yu, M. H. Single amino acid substitutions of alpha 1-antitrypsin that confer enhancement in thermal stability. *J. Biol. Chem.* **269,** 9627–9631 (1994).
49. Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P. & Lomas, D. A. A kinetic mechanism for the polymerization of alpha1-antitrypsin. *J. Biol. Chem.* **274,** 9548–9555 (1999).
50. James, E. L., Whisstock, J. C., Gore, M. G. & Bottomley, S. P. Probing the unfolding pathway of alpha1-antitrypsin. *J. Biol. Chem.* **274,** 9482–9488 (1999).
51. Knaupp, A. S. *et al.* The Roles of Helix I and Strand 5A in the Folding, Function and Misfolding of α1-Antitrypsin. *PLoS ONE* **8,** e54766 (2013).
52. Tran, S. T. & Shrake, A. The folding of alpha-1-proteinase inhibitor: kinetic vs equilibrium control. *Archives of Biochemistry and Biophysics* **385,** 322–331 (2001).
53. Kim, D. & Yu, M. H. Folding pathway of human alpha 1-antitrypsin: characterization of an intermediate that is active but prone to aggregation. *Biochem. Biophys. Res. Commun.* **226,** 378–384 (1996).
54. Patschull, A. O. M. *et al.* Therapeutic target-site variability in α1-antitrypsin characterized at high resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun*. **67,** 1492–1497 (2011).
55. Im, H. Interactions Causing the Kinetic Trap in Serpin Protein Folding. *J. Biol. Chem.* **277,** 46347–46354 (2002).
56. Dunstone, M. A. *et al.* Cleaved antitrypsin polymers at atomic resolution. *Protein Sci.* **9,** 417–420 (2000).
57. Zhou, A., Huntington, J. A., Pannu, N. S., Carrell, R. W. & Read, R. J. How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration. *Nat. Struct. Biol.* **10,** 541–544 (2003).
58. Stout, T. J., Graham, H., Buckley, D. I. & Matthews, D. J. Structures of active and latent PAI-1: a possible stabilizing role for chloride ions. *Biochemistry* **39,** 8460–8469 (2000).
59. Baumann, U. *et al.* Crystal structure of cleaved human alpha 1-antichymotrypsin at 2.7 A resolution and its comparison with other serpins. *Journal of Molecular Biology* **218,** 595–606 (1991).
60. Takehara, S. *et al.* The 2.1-A crystal structure of native neuroserpin reveals unique structural elements that contribute to conformational instability. *Journal of Molecular Biology* **388,** 11–20 (2009).
61. Ricagno, S., Caccia, S., Sorrentino, G., Antonini, G. & Bolognesi, M. Human neuroserpin: structure and time-dependent inhibition. *Journal of Molecular Biology* **388,** 109–121 (2009).
62. Skinner, R. *et al.* The 2.6 A structure of antithrombin indicates a conformational change at the heparin binding site. *Journal of Molecular Biology* **266,** 601–609 (1997).
63. Fulton, K. F. *et al.* The high resolution crystal structure of a native thermostable serpin reveals the complex mechanism underpinning the stressed to relaxed transition. *J. Biol. Chem.* **280,** 8435–8442 (2005).
64. Zhang, Q. *et al.* The N terminus of the serpin, tengpin, functions to trap the metastable native state. *EMBO Rep.* **8,** 658–663 (2007).
65. Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. Strength and co-operativity of contributions of surface salt bridges to protein stability. *Journal of Molecular Biology* **216,** 1031–1044 (1990).
66. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **29,** 9343–9352 (1990).
67. Fersht, A. R. & Serrano, L. Principles of protein stability derived from protein engineering experiments. *Current Opinion in Structural Biology* **3,** 75–83 (1993).
68. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput Biol* **4,** e1000002 (2008).
69. Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci. USA* **86,** 8237–8241 (1989).
70. Chothia, C. & Finkelstein, A. V. The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* **59,** 1007–1039 (1990).
71. DeDecker, B. S. *et al.* The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *Journal of Molecular Biology* **264,** 1072–1084 (1996).
72. Levitt, M., Gerstein, M., Huang, E., Subbiah, S. & Tsai, J. Protein folding: the endgame. *Annu. Rev. Biochem.* **66,** 549–579 (1997).

73. Kellis, J. T., Nyberg, K., Sali, D. & Fersht, A. R. Contribution of hydrophobic interactions to protein stability. *Nature* **333,** 784–786 (1988).
74. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129,** 10110–10112 (2007).
75. Miklos, A. E. *et al.* Structure-based design of supercharged, highly thermoresistant antibodies. *Chemistry & Biology* **19,** 449–455 (2012).
76. Parfrey, H. *et al.* Targeting a Surface Cavity of 1-Antitrypsin to Prevent Conformational Disease. *J. Biol. Chem.* **278,** 33060–33066 (2003).
77. Gooptu, B. *et al.* Crystallographic and Cellular Characterisation of Two Mechanisms Stabilising the Native Fold of α1-Antitrypsin: Implications for Disease and Drug Design. *Journal of Molecular Biology* **387,** 857–868 (2009).
78. Sivasothy, P., Dafforn, T. R., Gettins, P. G. W. & Lomas, D. A. Pathogenic 1-Antitrypsin Polymers Are Formed by Reactive Loop–Sheet A Linkage. *J. Biol. Chem.* **275,** 33663–33668 (2000).
79. James, E. L. & Bottomley, S. P. The mechanism of alpha 1-antitrypsin polymerization probed by fluorescence spectroscopy. *Archives of Biochemistry and Biophysics* **356,** 296–300 (1998).
80. Kass, I., Knaupp, A. S., Bottomley, S. P. & Buckle, A. M. Conformational properties of the disease-causing Z variant of α 1-antitrypsin revealed by theory and experiment. *Biophys. J.* **102,** 2856–2865 (2012).
81. Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C. & Bottomley, S. P. Kinetic instability of the serpin Z alpha1-antitrypsin promotes aggregation. *Journal of Molecular Biology* **396,** 375–383 (2010).
82. Hughes, V. A., Meklemburg, R., Bottomley, S. P. & Wintrode, P. L. The Z mutation alters the global structural dynamics of α 1-antitrypsin. *PLoS ONE* **9,** e102617 (2014).
83. Huang, X. *et al.* Molecular Mechanism of Z α1-Antitrypsin Deficiency. *J. Biol. Chem.* **291,** 15674–15686 (2016).
84. Knaupp, A. S. & Bottomley, S. P. Structural change in β-sheet A of Z α(1)-antitrypsin is responsible for accelerated polymerization and disease. *Journal of Molecular Biology* **413,** 888–898 (2011).
85. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA* **110,** 17874–17879 (2013).
86. Lee, K. N., Park, S. D. & Yu, M. H. Probing the native strain in alpha1-antitrypsin. *Nat. Struct. Biol.* **3,** 497–500 (1996).
87. Kim, J., Lee, K. N., Yi, G. S. & Yu, M. H. A thermostable mutation located at the hydrophobic core of alpha 1-antitrypsin suppresses the folding defect of the Z-type variant. *J. Biol. Chem.* **270,** 8597–8601 (1995).
88. Cabrita, L. D., Whisstock, J. C. & Bottomley, S. P. Probing the Role of the F-Helix in Serpin Stability through a Single Tryptophan Substitution †. *Biochemistry* **41,** 4575–4581 (2002).
89. Cabrita, L. D., Dai, W. & Bottomley, S. P. Different Conformational Changes within the F-Helix Occur during Serpin Folding, Polymerization, and Proteinase Inhibition †. *Biochemistry* **43,** 9834–9839 (2004).
90. Gettins, P. G. W. The F-helix of serpins plays an essential, active role in the proteinase inhibition mechanism. *FEBS Letters* **523,** 2–6 (2002).
91. Gooptu, B. *et al.* Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. *Proc. Natl. Acad. Sci. USA* **97,** 67–72 (2000).
92. Nyon, M. P. *et al.* Structural dynamics associated with intermediate formation in an archetypal conformational disease. *Structure* **20,** 504–512 (2012).
93. Pearce, M. C. *et al.* Preventing serpin aggregation: the molecular mechanism of citrate action upon antitrypsin unfolding. *Protein Sci.* **17,** 2127–21133 (2008).
94. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA* **104,** 19819–19824 (2007).
95. Parra, R. G. *et al.* Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **44,** W356–W360 (2016).
96. Tsai, M.-Y. *et al.* Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Sci.* **25,** 255–269 (2016).
97. Pearce, M. C., Cabrita, L. D., Rubin, H., Gore, M. G. & Bottomley, S. P. Identification of residual structure within denatured antichymotrypsin: implications for serpin folding and misfolding. *Biochem. Biophys. Res. Commun.* **324,** 729–735 (2004).
98. Lehmann, M. *et al.* From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng.* **13,** 49–57 (2000).
99. Barrick, D., Ferreiro, D. U. & Komives, E. A. Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Current Opinion in Structural Biology* **18,** 27–34 (2008).
100. Fersht, A. R. Characterizing transition states in protein folding: an essential step in the puzzle. *Current Opinion in Structural Biology* **5,** 79–84 (1995).
101. Cowieson, N. P. *et al.* MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation* **22,** 187–190 (2015).

Acknowledgements

We thank Jane Clarke, Andrew Ellisdon, and Gordon Lloyd for helpful discussions and advice. BTP is a Medical Research Council Career Development Fellow. AAN and JJH are supported by the Wellcome Trust (grant number WT 095195). SM acknowledges fellowship support from the Australian Research Council (FT100100960). NAB is an Australian Research Council Future Fellow (110100223). GIW is an Australian Research Council Discovery Outstanding Researcher Award Fellow (DP140100087). AMB is a National Health and Medical Research Senior Research Fellow (1022688). JCW is an NHMRC Senior Principal Research fellow and also acknowledges the support of an ARC Federation Fellowship. We thank the Australian Synchrotron for beam-time and technical assistance. This work was supported by the Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) (www.massive.org.au). We acknowledge the Monash Protein Production Unit and Monash Macromolecular Crystallization Facility.

Author Contributions

B.T.P., J.I., S.M., S.P.B. and A.M.B. designed the study. J.A.I., J.C.W., L.Z., S.P.B. and G.I.W. performed the protein design. S.K., E.M.M., M.A.P., W.D. and B.T.P. performed the protein expression, purification and CD thermal melt experiments. B.T.P., N.A.B., S.M. and S.K. performed the crystallography. B.T.P. performed molecular dynamics simulations and analysis with contributions from IK. D.E.H. assisted in biophysical data analysis. M.G.S.C. performed the frustration analysis, structural analysis and contributed to discussion of data. S.K., J.J.H. and A.A.N. performed the folding kinetics and equilibrium measurement experiments. B.T.P. and S.K. generated figures. B.T.P., S.M., A.A.N. and A.M.B. wrote the manuscript.

Additional Information

Accession Numbers: The coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5CDX, 5CDZ and 5CE0.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Porebski, B. T. *et al.* Smoothing a rugged protein folding landscape by sequence-based redesign. *Sci. Rep.* **6**, 33958; doi: 10.1038/srep33958 (2016).