

# SCIENTIFIC REPORTS



OPEN

## Probing the folding pathway of a consensus serpin using single tryptophan mutants

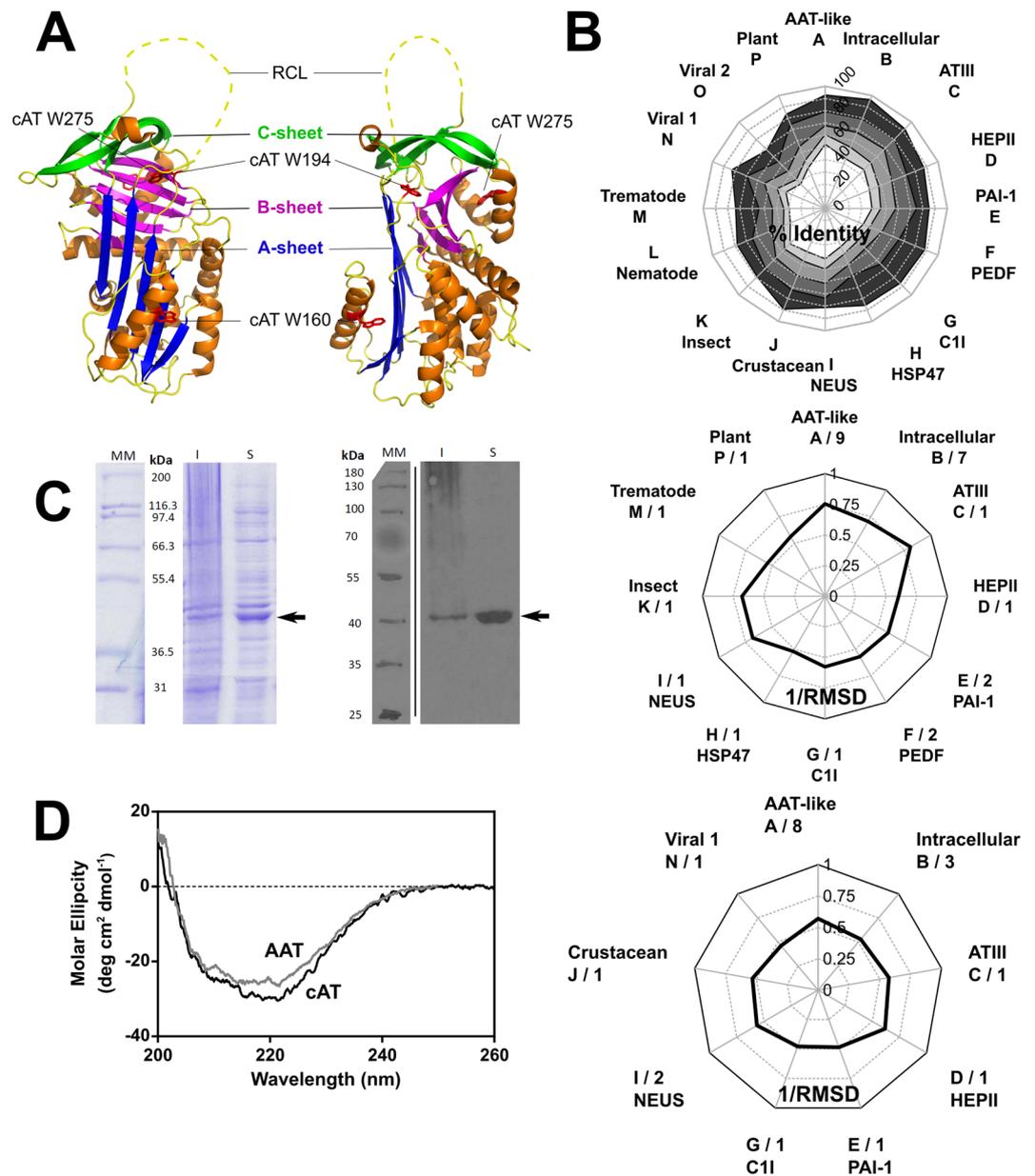
Li Yang<sup>1</sup>, James A. Irving<sup>2</sup>, Weiwen Dai<sup>1</sup>, Marie-Isabel Aguilar<sup>1</sup> & Stephen P. Bottomley<sup>1</sup>

Conserpin is an engineered protein that represents the consensus of a sequence alignment of eukaryotic serpins: protease inhibitors typified by a metastable native state and a structurally well-conserved scaffold. Previously, this protein has been found to adopt a native inhibitory conformation, possess an atypical reversible folding pathway and exhibit pronounced resistance to inactivation. Here we have designed a version of conserpin, cAT, with the inhibitory specificity of  $\alpha_1$ -antitrypsin, and generated single-tryptophan variants to probe its folding pathway in more detail. cAT exhibited similar thermal stability to the parental protein, an inactivation associated with oligomerisation rather a transition to the latent conformation, and a native state with pronounced kinetic stability. The tryptophan variants reveal the unfolding intermediate ensemble to consist of an intact helix H, a distorted helix F and 'breach' region structurally similar to that of a mesophilic serpin intermediate. A combination of intrinsic fluorescence, circular dichroism, and analytical gel filtration provide insight into a highly cooperative folding pathway with concerted changes in secondary and tertiary structure, which minimises the accumulation of two directly-observed aggregation-prone intermediate species. This functional conserpin variant represents a basis for further studies of the relationship between structure and stability in the serpin superfamily.

The serpin superfamily, with members in every phylogenetic kingdom, underwent a marked expansion in diversity following the plant-animal split<sup>1–3</sup>. The majority of serpins inhibit serine proteases, although some have been found to target classes of cysteine protease<sup>4–6</sup> or perform alternative roles such as hormone transport, as chaperones and as tumour suppressors<sup>7,8</sup>. In contrast to most proteins<sup>9</sup>, the serpin native conformation is a kinetically-trapped, thermodynamically metastable folding intermediate. This characteristic is intimately associated with the mechanism of inhibition<sup>10</sup>, the essence of which is a transition of a central 5-stranded  $\beta$ -sheet (denoted ' $\beta$ -sheet A') to a thermodynamically-preferred 6-stranded conformation. The serpin fold (Fig. 1A) is highly conserved despite this functional diversity, with structural deviations largely restricted to extensions at the N- and C- termini and in a defined inter-helical region<sup>1</sup>. The unadorned core structure is epitomised by one of the most studied serpins,  $\alpha_1$ -antitrypsin ( $\alpha_1$ -AT).

Like most mesophilic serpins,  $\alpha_1$ -antitrypsin follows a three-state folding mechanism that proceeds from an unfolded state *via* an intermediate to a fully folded form<sup>11</sup>. The intermediate is of great interest, as it has been proposed to be the aggregating species central to the pathology of a number of hereditary conditions involving serpin deficiency<sup>12,13</sup>. Mutations, or environmental conditions, that increase the intermediate population by perturbing the energy landscape increase the formation of non-functional serpin polymers<sup>12,14–16</sup>. Conversely, minimal population of this intermediate has been observed in serpins produced by hyperthermophiles<sup>17,18</sup>, and is proposed as a mechanism whereby folding is possible, and activity is maintained, under destabilising conditions. The deficiency in functional  $\alpha_1$ -AT resulting from mutations such as Z, M<sub>Malton</sub> or Siiyama is associated with neonatal hepatitis, liver disease, hepatocellular carcinoma and emphysema. Other serpins important to human physiology, such as antithrombin III,  $\alpha_1$ -antichymotrypsin, C1-inhibitor and plasminogen activator inhibitor-1, show a similar vulnerability to mutation<sup>19</sup>.

<sup>1</sup>Department of Biochemistry and Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia. <sup>2</sup>UCL Respiratory and the Institute of Structural and Molecular Biology, University College London, London, United Kingdom. Li Yang and James A. Irving contributed equally to this work. Marie-Isabel Aguilar and Stephen P. Bottomley jointly supervised this work. Correspondence and requests for materials should be addressed to L.Y. (email: [yang.le@wehi.edu.au](mailto:yang.le@wehi.edu.au)) or J.A.I. (email: [j.irding@ucl.ac.uk](mailto:j.irding@ucl.ac.uk))



**Figure 1.** Properties of a conserpin variant. **(A)** The serpin fold is comprised of three  $\beta$ -sheets and eight or nine  $\alpha$ -helices, illustrated using a cartoon representation of native conserpin (PDB 5CDX)<sup>34</sup>. Connecting  $\beta$ -strand 5 A and  $\beta$ -strand 1 C is the solvent exposed reactive centre loop (RCL) which dictates the target protease specificity of inhibitory serpins (unresolved in the structure and therefore denoted by a dashed line). Upon proteolytic cleavage, the RCL becomes a 6<sup>th</sup> central strand of the normally 5-stranded  $\beta$ -sheet A. The positions of the tryptophan residues mutated in this study are indicated. **(B) Upper panel:** Radial plot of the average sequence identity between the consensus sequence and members of individual serpin phylogenetic clades<sup>1</sup>, calculated using all alignment positions (white central area), or the 75%, 50%, 25% and 10% lowest variability sites (shaded from light grey to dark grey). **Middle panel:** Reciprocal RMSD of native conserpin (PDB 5CDX) compared with the common core of representative structures from different phylogenetic clades. **Lower panel:** A comparison between latent conserpin (PDB 5CDZ) and representative cleaved/latent structures. **(C) Left:** SDS-PAGE (10% w/v, see full gels in Supplementary Fig. S1) section of cAT after 20 hours of auto-induction, visualised using Coomassie Blue stain. **Right:** Western blot of the corresponding SDS-PAGE, developed using an anti-His<sub>6</sub> antibody. Lane MM, molecular weight markers; I, insoluble fraction; S, soluble lysate fraction. cAT is indicated by the arrow. **(D)** Far-UV spectrum of cAT (black) compared with  $\alpha_1$ -AT (grey) at a concentration of 0.2 mg mL<sup>-1</sup>, using a path length of 0.1 cm.

The formation of serpin polymers can be reduced under experimental conditions by stabilisation of the native state of the protein. This has been achieved by the rational introduction of point mutations to fill in surface cavities of  $\alpha_1$ -AT<sup>20</sup>, disulphide constraints of structural elements involved in conformational change<sup>13,21–23</sup>, the

addition of osmolytes<sup>23–25</sup> and random mutagenesis<sup>26</sup>. However, serpin function is intimately associated with a finely-tuned native state instability that is distributed throughout the molecule<sup>23,27</sup>, a context in which some, but not all, stabilising mutations compromise inhibitory activity<sup>28</sup>.

A further strategy that has been employed to increase the thermodynamic stability of a protein is the consensus approach to rational protein design. This technique identifies residues in a native protein that are putatively the most compatible with a given fold based on conservation between members of a protein family. Mutations have systematically been identified that are compatible with function and enhance stability for a range of proteins which include DNA binding proteins<sup>29</sup>, antibodies<sup>30,31</sup>, leucine rich repeat proteins<sup>32</sup>, and enzymes<sup>33</sup>. By extending this approach to the whole protein rather than selected residues, conserpin, an entirely artificial serpin, has recently been engineered<sup>34</sup> with a sequence that reflects the most frequently observed residue at each site in an alignment of eukaryotic serpins<sup>1</sup>. This protein was found to adopt the canonical native, metastable conformation and showed a remarkable stability to inactivation by heat, an atypical fully reversible unfolding pathway, and an absence of polymer formation. Structural comparison with other eukaryotic serpins revealed features proposed to contribute to its stability, including stabilising interactions around  $\alpha$ -helix D and the  $\beta$ -sheet B/C barrel, an extended electrostatic network in the ‘breach’ region, and tight packing in the hydrophobic core and at the  $\beta$ -sheet A/helix F interface. While conserpin showed a two-state unfolding profile atypical of serpins by intrinsic tryptophan fluorescence, the presence of folding intermediates was inferred from kinetic refolding experiments and 4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid (bis-ANS) binding<sup>34</sup>.

Here we report the design, production and folding properties of a conserpin variant that incorporates the specificity-conferring reactive centre loop (RCL) residues of the disease-associated serpin,  $\alpha_1$ -AT. In common with its progenitor the resulting synthetic protein, cAT (consensus  $\alpha_1$ -AT), exhibited high thermal stability and reversible folding. We exploited the three tryptophans present in conserpin by generating single-tryptophan variants to probe unfolding in the proximity of helix F, the ‘breach’ at the top of  $\beta$ -sheet A, and helix H. These data, with the additional use of circular dichroism as a probe of change in secondary structure, extend the previous study by demonstrating recruitment of the whole molecule during folding. In combination with size-exclusion chromatography, we provide direct evidence of the presence of folding intermediate states inferred previously from kinetic data. We find the altered RCL favours inactivation by oligomerisation over transition to the latent form as observed for the parental protein, and utilise this property to assess the kinetic stability of the native state. Collectively, the data reflect a molecule with a highly efficient, concerted folding mechanism that minimises the accumulation of an aggregation-prone intermediate species as a result of a pronounced kinetic stability.

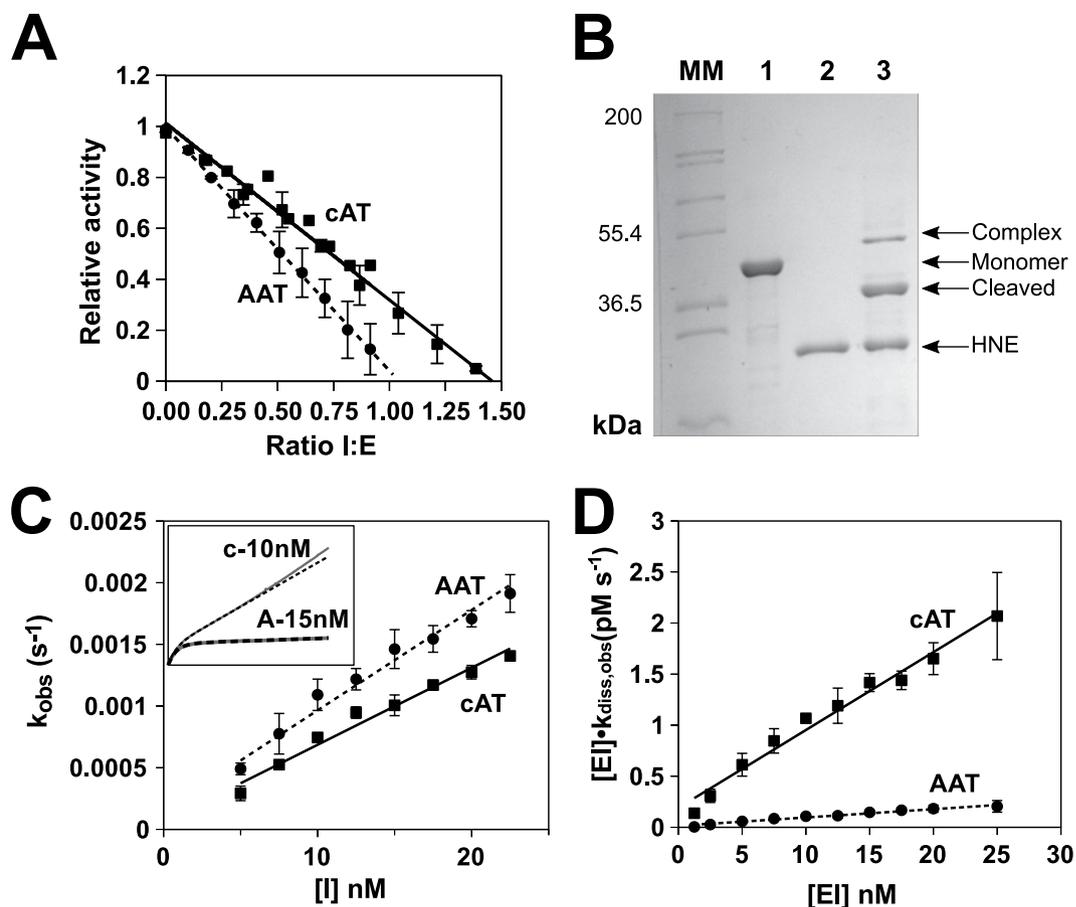
## Results

**Engineering a conserpin variant with  $\alpha_1$ -AT RCL residues.** The design of conserpin has been described in detail previously<sup>34</sup>. The consensus sequence it embodies, when placed in a phylogenetic context<sup>1</sup>, has a slightly higher similarity to  $\alpha_1$ -AT-like, intracellular, and antithrombin clades A–C over other subfamilies (Fig. 1B, upper panel). This is also reflected by a comparison between the conserpin structure<sup>34</sup> and other serpins in the native conformation (Fig. 1B, middle panel) but not between loop-inserted conserpin and cleaved structures (Fig. 1B, lower panel). This highlights a dichotomy between the serpin metastable native conformation and the thermodynamically stabilised loop-inserted form.

The RCL sequence of conserpin was previously found to support inhibitory activity against trypsin, with a stoichiometry of inhibition (SI) of 1.8<sup>34</sup>, equivalent to a ~45% non-productive turnover rate. The conserpin sequence represents a hybrid of the subsite preferences of a broad spectrum of target proteases. However, as the interaction between two defined partners is generally determined by multiple specific subsite interactions within the binding cleft<sup>35</sup>, this consensus RCL may not fully capture the potential inhibitory activity of the conserpin scaffold against a single target. Furthermore, the conserpin RCL is 1–4 residues shorter on the P' side than many inhibitory serpins. To introduce a sequence with a known cognate proteolytic profile, the P<sub>7</sub> to P<sub>2</sub>' region of the RCL was replaced by the corresponding region of  $\alpha_1$ -AT, to yield a functional variant, cAT. The purified protein (Fig. 1C) was found to exhibit a mixed  $\alpha$ -helix/ $\beta$ -sheet far UV-circular dichroism profile consistent with a folded serpin (Fig. 1D).

**cAT is a functional serpin with  $\alpha_1$ -AT-like specificity.** To determine whether the RCL substitution conferred cAT with an  $\alpha_1$ -AT-like activity, the SI was determined against chymotrypsin. SI is a measure of the efficiency of the inhibitory mechanism, governed by efficient loop incorporation and effective trapping of the tethered protease<sup>36,37</sup>. With a molar ratio of serpin-to-protease of  $1.46 \pm 0.03:1$  (SEM, n = 4) to achieve full inhibition (Fig. 2A), this interaction was more efficient than the interaction of conserpin with trypsin, with around a third of cAT molecules following the substrate pathway. cAT was also found to form an SDS-stable inhibitory complex – a hallmark of many serpin-enzyme interactions<sup>38</sup> – with human neutrophil elastase (Fig. 2B). This complex did not show evidence of secondary degradation by-products, in contrast to the interaction between conserpin and trypsin<sup>34</sup>. The second order association rate constant of cAT with chymotrypsin ( $k'_{\text{ass}}$ ), corrected for non-productive turnover, was  $7.7 \pm 0.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  at 25 °C (SE of the regression, n = 4), slightly faster than the rate of interaction with  $\alpha_1$ -AT under the same experimental conditions,  $6.9 \pm 0.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (Fig. 2C). Thus cAT can present its  $\alpha_1$ -AT RCL in an effective orientation for engagement with a protease and is capable of a sufficiently rapid conformational change to act as an inhibitor.

The inhibitory progress curves did exhibit a difference in behaviour between the two proteins, however, with a noticeably higher residual protease activity for cAT that increased towards the end of the experiment (Fig. 2C, inset). This is characteristic of complex dissociation. To assess this, 5  $\mu\text{M}$  inhibitory complex was diluted to 0–25 nM, and the regain of protease activity monitored by chromogenic substrate turnover. The apparent rate of complex dissociation observed for the cAT-chymotrypsin complex,  $6.9 \pm 1.0 \times 10^{-5} \text{ s}^{-1}$ , was around 10-fold greater than that of the  $\alpha_1$ -AT-chymotrypsin complex, at  $7.5 \pm 1.2 \times 10^{-6} \text{ s}^{-1}$  (SEM, n = 4) (Fig. 2D). Coupled

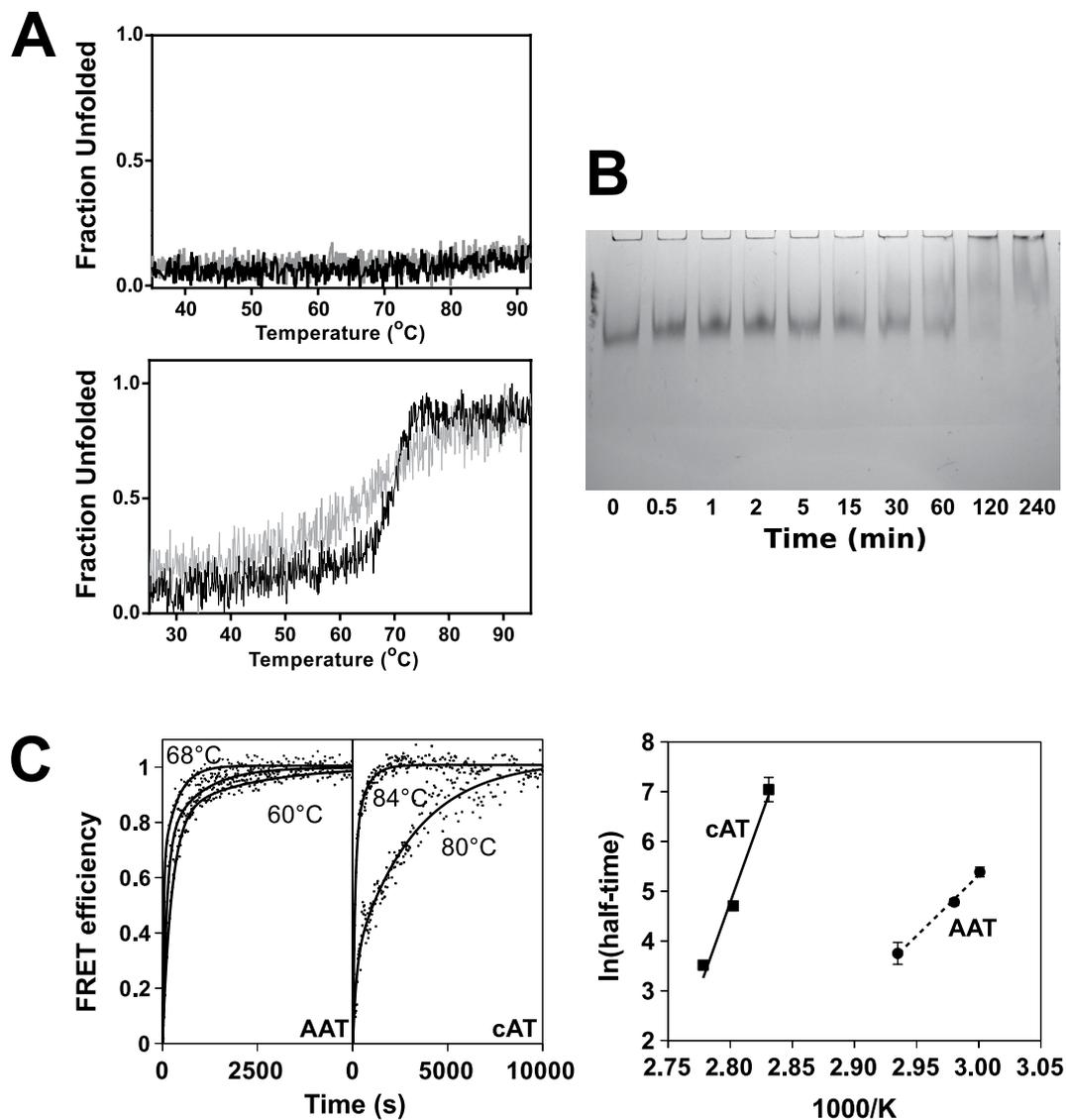


**Figure 2.** Serpin-enzyme complex formation by cAT. **(A)** Residual protease activity following the incubation at 25 °C of cAT with chymotrypsin at various molar ratios; the intercept of the linear regression with the abscissa is equivalent to the stoichiometry of inhibition for cAT (solid line) in comparison with  $\alpha_1$ -AT (dashed line). Error bars represent  $\pm$  SEM for four independent experiments. **(B)** SDS-PAGE (10% w/v) of cAT incubated with human neutrophil elastase (HNE) at 37 °C for 2 hours. MM, molecular marker; lane 1, cAT control; lane 2, HNE; lane 4, cAT and HNE incubated at equimolar concentration. Monomer, cleaved and serpin-enzyme complex bands are labelled. **(C)** The inhibition of chymotrypsin by several concentrations of cAT and  $\alpha_1$ -AT was followed over time at 25 °C, from which the apparent second-order rate constant ( $k_{obs}$ ) was calculated. The slope of the relationship between inhibitor concentration and  $k_{obs}$  provides the uncorrected second-order rate constant,  $k_{app}$ . Error bars represent  $\pm$  SEM ( $n = 4$ ). The inset graph shows representative progress curves of cAT (c) and  $\alpha_1$ -AT (A) with approximately equal  $k_{obs}$  values. **(D)** The dissociation of the complex between chymotrypsin and cAT or  $\alpha_1$ -AT was followed as the rate of regain of protease activity at 25 °C following dilution from 5  $\mu$ M to the concentrations shown. The slope of the resulting regression provides the apparent rate constant,  $k_{diss,app}$ . Error bars represent  $\pm$  SEM ( $n = 4$ ).

with a stoichiometry above parity, this is suggestive of a stability-function trade-off that yields an active but partially perturbed inhibitory mechanism.

**cAT forms higher-order species when heated.** In common with the parental protein<sup>34</sup>, cAT was found to be resistant to heat-induced denaturation, with minimal detectable change in secondary structure over the course of a thermal denaturation assay from 25 °C to 95 °C at 1 °C min<sup>-1</sup> (Fig. 3A, upper), requiring the addition of denaturant to unfold. In the presence of 2 M guanidine hydrochloride (GdnHCl) (Fig. 3A, lower), the unfolding temperature midpoint of denaturation ( $T_m$ ) of cAT was 70.0  $\pm$  0.1 °C (SEM,  $n = 3$ ), which is within 2.5 °C of the value for conserpin<sup>34</sup>. Following a decrease in temperature from 95 °C to 25 °C, the material showed no visible sign of precipitation, and retained ~50% activity (an SI value of 2.8  $\pm$  0.2, SEM,  $n = 3$ ) with respect to the pre-denaturation control. This partial loss of activity was not accompanied by a loss of circular dichroism (CD) signal, indicating that it was not associated with gross structural changes in, or significant precipitation of, the sample (Fig. 3A, lower). The thermal refolding curve revealed a similar midpoint transition of 69.5 °C albeit with a less co-operative behaviour than seen during unfolding, indicating that, in the presence of denaturant, the forward and reverse pathways of cAT are different.

Several serpins are known to form ordered long-chain polymers when heated due to population of an oligomerisation-prone intermediate state<sup>23,39,40</sup>. The latent conformation is an alternate monomeric end-state,



**Figure 3.** Heat-induced oligomerisation of cAT. **(A)** *Upper panel:* Thermal unfolding of  $0.2 \text{ mg mL}^{-1}$  cAT, heated from  $25^\circ\text{C}$  to  $95^\circ\text{C}$  at a rate of  $1^\circ\text{C min}^{-1}$ , monitored by the change in CD signal at  $222 \text{ nm}$  (black). The sample was then returned to the starting temperature at the same rate (grey). *Lower panel:* A representative thermal melt performed in the presence of  $2 \text{ M}$  of GdnHCl. **(B)** cAT (at  $10 \mu\text{M}$ ) was heated at  $75^\circ\text{C}$ ,  $3 \mu\text{L}$  aliquots removed at various timepoints, and the oligomerisation state resolved by 6% non-denaturing PAGE. No monomer was visible at the conclusion of the experiment. **(C)** *Left panel:* cAT or  $\alpha_1$ -AT labelled with Atto 488 and Atto 594 fluorescent dyes at a concentration of  $0.1 \text{ mg mL}^{-1}$  in PBS was heated at temperatures between  $80$ – $87^\circ\text{C}$  and  $60$ – $68^\circ\text{C}$ , respectively, and the increase in FRET between the fluorophores determined over time. Points reflect representative data normalised between 0 and 1, with the curves of best fit as solid lines. *Right panel:* The natural logarithm half-times of the change in FRET were plotted against the reciprocal inverse absolute temperature; the slopes of the regressions are proportional to the apparent activation energy of polymerisation under the conditions of the experiment. Error bars reflect  $\pm \text{SEM}$  ( $n = 4$ ).

in which the RCL is incorporated into  $\beta$ -sheet A without cleavage; this form can neither inhibit proteases nor polymerise<sup>41</sup>, and it also proposed to be accessed via an intermediate<sup>42</sup>. After heating at  $76^\circ\text{C}$  for 5 hours, conserpin had been observed to fully convert to this monomeric latent conformation<sup>34</sup>. In contrast, incubation of cAT at  $75^\circ\text{C}$  led to a gradual decrease in the monomer band intensity with a concurrent appearance of higher molecular mass species by non-denaturing PAGE, including a component unable to migrate into the gel (Fig. 3B). After four hours a complete loss of monomer was seen, with no residual monomer band that could indicate the presence of the latent species<sup>41</sup>. Thus the substitution of RCL residues in cAT confers behaviour more consistent with polymerisation-prone eukaryotic serpins, and by analogy the ability to adopt a polymerisation intermediate state.

To assess this possibility and follow polymerisation in real-time, mixtures of cAT and  $\alpha_1$ -AT (at  $0.1 \text{ mg mL}^{-1}$ ) labelled with Atto-488 and Atto-594 dyes were heated, and reaction progress monitored by a change in Förster resonance energy transfer (FRET). The resulting curves reported an increased proximity between donor and

acceptor fluorophores to within  $\sim 80 \text{ \AA}$ , consistent with intermolecular association between differentially labelled proteins (Fig. 3C, *left*). The temperature dependence of the rate of polymerisation was then used to calculate the apparent activation energy of this oligomerisation (Fig. 3C, *right*). Under these conditions, the energetic barrier to polymerisation was found to be around 3-fold higher for cAT than  $\alpha_1$ -AT, at  $139 \pm 12 \text{ kcal mol}^{-1}$  and  $48 \pm 6 \text{ kcal mol}^{-1}$ , respectively ( $\pm \text{SE}$  of regression,  $n = 4$ ). Thus, the cAT native state exhibits significant kinetic stability against heat-induced inactivation.

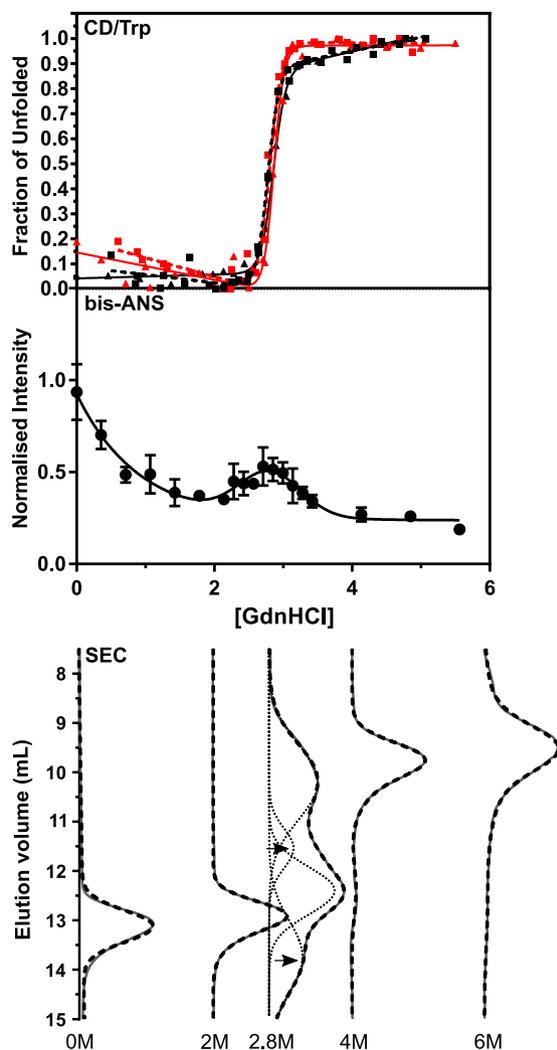
**cAT unfolds via a poorly populated intermediate.** In the seminal conserpin study<sup>34</sup>, intrinsic tryptophan fluorescence reported an apparently two-state equilibrium unfolding profile. This was inconsistent with data obtained using bis-ANS and rapid folding kinetics, which suggested the presence of a folding intermediate, and hence a three-state mechanism. This discrepancy could be explained either by a poorly populated intermediate, or well-populated intermediate with fluorescence properties very similar to either the native or unfolded state<sup>34</sup>. To investigate this further, GdnHCl-mediated equilibrium unfolding of cAT was undertaken whilst monitoring changes in far-UV CD signal, to use the global structure of the protein as an additional reporter of structural change. The resulting curves (Fig. 4, *upper panel*) showed an apparent two-state, strongly cooperative and near-contemporaneous transition in both fluorescence and CD, with the mid-point of denaturation ( $D_{50\%}$ ) centred at 2.9 M. As noted for conserpin, this contrasts with the three-state unfolding profile exhibited by most serpins<sup>12,40,43</sup>. The equilibrium unfolding and refolding data were superimposable – consistent with a fully reversible folding pathway – and protein refolded from 6 M GdnHCl to 0.2 M GdnHCl had an inhibitory activity close to the starting value (SI of  $1.7 \pm 0.1$ ; SEM,  $n = 3$ ). Bis-ANS, an environment-sensitive probe whose quantum yield increases upon binding to hydrophobic regions of equilibrium intermediates<sup>12,34,43</sup>, was found to exhibit a small peak in mid-range concentrations of denaturant (Fig. 4, *middle panel*), centred at  $2.8 \pm 0.1 \text{ M}$  GdnHCl (SE of global fit,  $n = 3$ ). This corresponded well with fluorescence and far-UV CD  $D_{50\%}$  values. Notably, the bis-ANS peak appeared over a noticeably more narrow concentration range than seen for  $\alpha_1$ -AT or  $\alpha_1$ -antichymotrypsin, with a width of  $\sim 1.7 \text{ M}$  compared with 3 M for the latter two proteins<sup>12,43</sup>. These data are consistent with a minimally-populated folding intermediate and the simultaneous loss of secondary and tertiary structure at increasing concentrations of denaturant.

**Analytical gel filtration reveals species with expanded and compact conformations.** Size exclusion chromatography (SEC) can be used to detect distinct intermediate ensembles with different hydrodynamic volumes to that of the native and unfolded conformations<sup>18</sup>. Accordingly, cAT was incubated in five different concentrations of GdnHCl chosen on the basis of the spectroscopic data, and the SEC elution profile of each sample determined (Fig. 4, *lower panel*). In the absence of denaturant, the native protein was found to elute at around 13 mL, while in 6 M GdnHCl, the expanded unfolded species eluted at 9.5 mL. At 2.8 M, which corresponds with the spectroscopic  $D_{50\%}$  value and the peak in bis-ANS fluorescence, both of these species were evident, but with additional peaks at 11.5 mL and 13.7 mL. These peaks are indicative of a population of expanded molecules, as expected for the molten globule-like unfolding intermediate of  $\alpha_1$ -AT<sup>44</sup>, and additionally a compact ensemble with respect to the native state. Notably, re-folding rate data indirectly suggested that conserpin populates two distinct intermediates<sup>34</sup>, an inference which appears to be consistent with the species directly observed here. Interestingly, it has been proposed that the heat-induced intermediate of  $\alpha_1$ -AT shows compaction with respect to the native state<sup>45,46</sup>. At either 2 M or 4 M GdnHCl, these additional species were absent. In combination with the spectroscopic data, these profiles are consistent with the presence of a poorly populated intermediate ensemble, comprised of two species with distinct properties, which exist over a narrow denaturant concentration range.

**Single-tryptophan variants as probes of local folding.** Tryptophan variants can be used as site-specific probes of the local folding behaviour of a protein, as described previously for  $\alpha_1$ -AT<sup>47,48</sup> and plasminogen activator inhibitor-1<sup>49</sup>. The cAT sequence contains tryptophans at positions 160, 194 and 275 (all designations made using  $\alpha_1$ -AT numbering), corresponding with helix F, the loop connecting strand 3 of  $\alpha$ -helix A and strand 4 of  $\beta$ -sheet C (situated in the ‘breach’ region at the top of  $\beta$  sheet A), and helix H respectively (Figs 1A and 5A). In order to use these residues as reporters of local structural change, single tryptophan variants were generated by systematically mutating two at a time to phenylalanine, to form the variants cAT<sub>W160</sub>, cAT<sub>W194</sub>, and cAT<sub>W275</sub>. The CD profiles of the resulting proteins were consistent with that of the wild-type cAT (Fig. 5B), and did not result in a loss of inhibitory activity (Table 1). Thus, the double tryptophan-to-phenylalanine substitutions did not substantially alter global structure or function.

In the presence of 6 M GdnHCl, the intrinsic fluorescence emission spectra of the denatured proteins were fully additive (Fig. 5C, *upper panel*). Spectra of cAT and variants were also recorded in the absence of denaturant (Fig. 5D). The sum of the emission spectra of all single tryptophan variants yielded a maximal intensity about 120% of that of cAT, with comparable peak emission wavelengths at 331–332 nm. A previous analysis of plasminogen activator inhibitor-1, with three of the four tryptophan residues at identical positions to those in cAT, revealed significant quenching of Trp275 by Trp194<sup>49</sup>. In the context of similar inhibitory activity, far-UV CD profiles and the peak fluorescence emission wavelength across the three cAT variants, this 20% increase in intensity was therefore more likely the consequence of resonance energy transfer in cAT rather than the result of marked structural perturbation.

Intrinsic fluorescent scans were also performed in the presence of 8 M urea. Two of the variants had identical emission profiles to that seen when unfolded in 6 M GdnHCl. The exception, cAT<sub>W275</sub> (Fig. 5C, *lower panel*), displayed an intensity at the emission maximum slightly higher than the other variants coupled with a  $-5 \text{ nm}$  blue shift. This indicates that full solvation was not achieved at position Trp275 and some structure persists around helix H. The vicinity of helix H is thus more resistant to unfolding, as seen previously in  $\alpha_1$ -antichymotrypsin<sup>50</sup>; the folding nucleus around which the rest of the serpin scaffold condenses therefore appears to be conserved in cAT.



**Figure 4.** cAT populates an intermediate state over a narrow range of denaturant. *Upper panel:* Equilibrium GdnHCl-mediated unfolding (solid lines) and refolding (dotted lines) for cAT followed via the change in CD signal at 222 nm (black lines) and intrinsic fluorescence at 330 nm (red lines). Each dataset is the result of at least three independent experiments. A two-state unfolding curve was satisfactorily fit to the data; midpoints are listed in Table 1. *Middle panel:* cAT was incubated in varying concentrations of GdnHCl at 25 °C and bis-ANS added at 5 times the concentration of the protein. The fluorescence intensity was measured at 480 nm, with an excitation wavelength of 390 nm, and slit widths of 5 nm. Error bars reflect SD from three independent experiments, whose profiles were normalised according to their total integrated fluorescence intensity. The curve reflects the sum of an empirically-determined single exponential decay and Gaussian function. *Lower panel:* GdnHCl unfolding was monitored by size exclusion chromatography using a Superose 12 10/300 column, at the denaturant concentrations shown. The absorbance at 280 nm is shown in grey. The deconvoluted components of the 2.8 M sample are shown as dotted lines, the sums of the fitted components shown as dashed lines, and the experimental data as solid lines. Arrows indicate the expanded (top) and compact (bottom) intermediates.

**Tryptophan residues contribute to the stability of cAT.** Thermal unfolding experiments were used to assess whether the double tryptophan substitutions resulted in a change in stability. In the absence of denaturant, the variants exhibited an observable unfolding profile, with  $T_m$  values of approximately 80 °C (Table 1 and Fig. 5E). This is appreciably lower than wild-type cAT but considerably higher than any single eukaryotic serpin studied. Similarly, in the presence of 2 M GdnHCl, the variants had a  $T_m$  of 6–8 °C lower than parental cAT (see Fig. 5F and Table 1). These data highlight the importance of the hydrophobic packing mediated by each of these residues, within the key helix F/ $\beta$ -sheet A, breach and sheet B/C barrel regions (Fig. 5A).

**cAT folding and unfolding is cooperative.** Equilibrium unfolding experiments were then performed on the mutants, with measurements by both CD and intrinsic fluorescence (Fig. 6A). The calculated  $D_{50\%}$  values (Table 1) reflected a marginally lower thermodynamic stability than the wild type cAT protein (2.5–2.6 M as compared with 2.9 M) with a  $\Delta\Delta G_{D-N}$  of greater than 4.4 kcal mol<sup>-1</sup> between cAT and the single tryptophan

Variant	SI (pre-unfold)	SI (refold)	T <sub>m</sub> (°C)	+GdnHCl T <sub>m</sub> (°C)	D <sub>50%</sub> (M)		ΔG <sub>D-N</sub> (kcal mol <sup>-1</sup> )
cAT	1.5 ± 0.1	1.7 ± 0.1	>95	69.9 ± 0.1	Fl CD	2.9 (2.8) 2.9 (2.8)	23 ± 3 19 ± 3
cAT <sub>W160</sub>	1.5 ± 0.2	1.4 ± 0.2	82.2 ± 0.3	64.0 ± 0.1	Fl CD	2.5 (2.6) 2.6 (2.6)	13 ± 3 13 ± 2
cAT <sub>W194</sub>	1.2 ± 0.1	1.1 ± 0.1	83.1 ± 0.3	64.1 ± 0.1	Fl CD	2.5 (2.6) 2.5 (2.5)	18 ± 3 13 ± 2
cAT <sub>W275</sub>	1.5 ± 0.1	1.6 ± 0.1	77.6 ± 0.1	61.9 ± 0.1	Fl CD	2.7 (2.7) 2.6 (2.6)	15 ± 2 11 ± 1

**Table 1.** Comparison of cAT and single tryptophan mutants. The SI was determined against chymotrypsin prior to unfolding in 6 M GdnHCl and after refolding by dilution back to a final GdnHCl concentration of 0.2 M. SI errors reflect ± SE from the fit to the data (n = 2 or 3). Thermal unfolding was monitored by the change in CD signal, in the absence and presence of 2 M GdnHCl; T<sub>m</sub> errors are SEM from three independent experiments. Refold D<sub>50%</sub>, shown in parentheses, was derived by unfolding protein in 6 M GdnHCl and then diluting it back into 0.2 M GdnHCl; midpoint and ΔG<sub>D-N</sub> values were calculated from equilibrium unfolding intrinsic fluorescence and CD profiles, as indicated.

variants. These data point to the importance of the local interactions made by the tryptophan residues. However, reversibility was not affected: similar curves were seen when the refolding of each variant was monitored in the same manner. The close correspondence in behaviour reported by tryptophan residues on different structural elements, by global (CD) and by local (intrinsic fluorescence) measures of unfolding, highlight the highly cooperative nature of cAT folding.

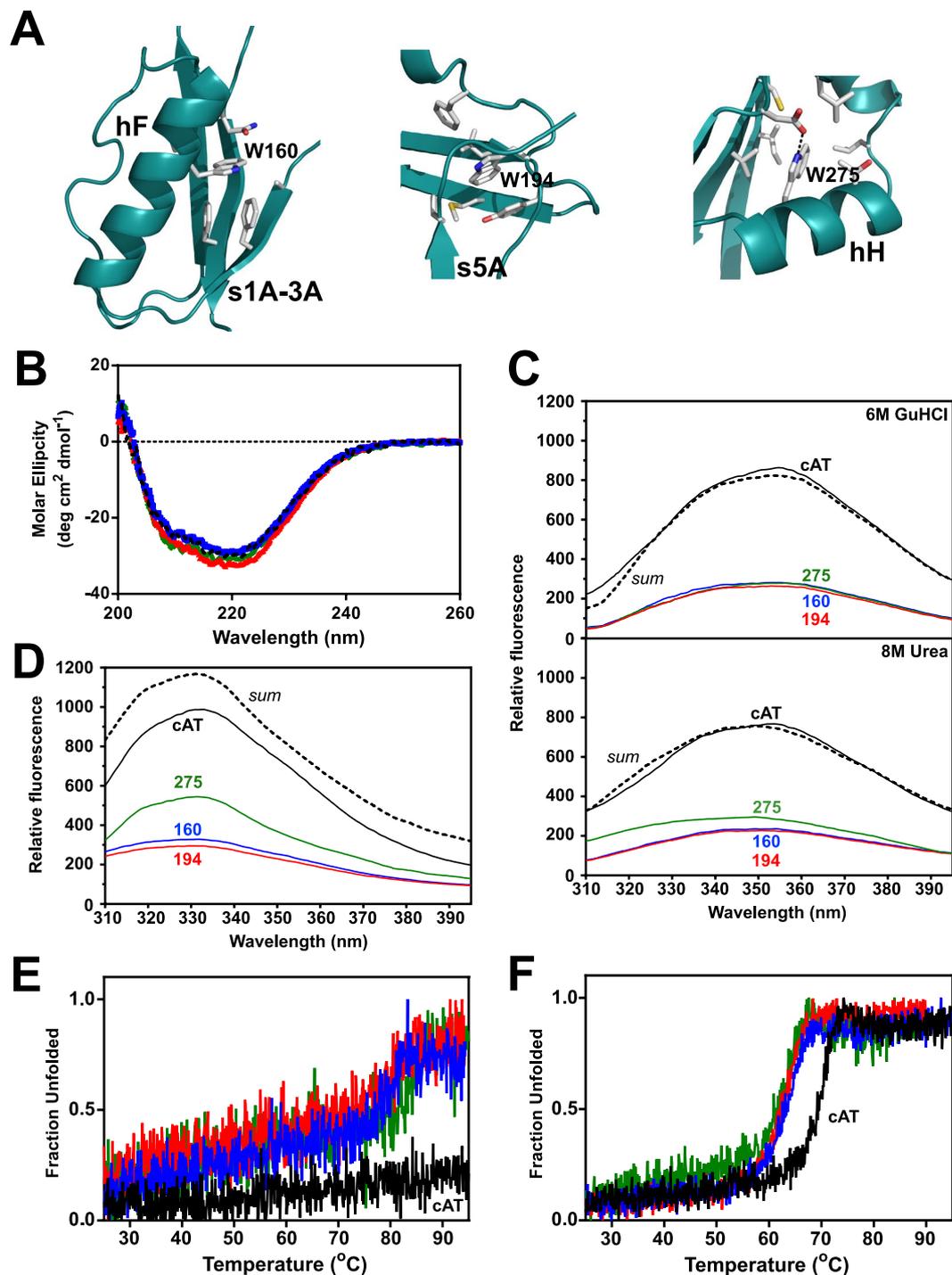
**Intrinsic fluorescence spectra reveal the presence of a folding intermediate.** Fluorescence scans of α<sub>1</sub>-AT at the midpoint of chemical denaturation exhibited a red shift from 330 nm to 343 nm, which has been ascribed to the presence of an intermediate ensemble<sup>47</sup>. Similar scans were conducted here at 3 M and 2.6 M GdnHCl for cAT and the tryptophan variants, respectively (Fig. 6B). Under these conditions, all double mutants showed a λ<sub>max</sub> of approximately 348 nm, signifying the partial exposure of each individual tryptophan to the solvent due to local conformational changes. This common behaviour is consistent with a molten globule intermediate<sup>51</sup>. These data, in combination with the results obtained using bis-ANS and SEC, support the conclusion that cAT folds through a three-state mechanism via a molten globule intermediate. Therefore, the reversible nature of cAT unfolding is a consequence of a poorly populated, rather than absent, folding intermediate, facilitated by a native state with pronounced thermodynamic and kinetic stability and a highly efficient and cooperative folding pathway.

## Discussion

Serpins are present in every taxonomic phylum, including prokaryotes that live at extremes of temperature. However, well-characterised serpins relevant to human physiology generally possess a finely-balanced stability easily perturbed by mutation, which on the face of it appears to be a necessary compromise to maintain the unique mechanism of action. It follows that prokaryotic hyperstable serpins should therefore be the product of specialised evolutionary adaptations to an intrinsically unstable scaffold. It is notable then that conserpin, a protein representative of eukaryotic serpins, exhibits both hyperstability and inhibitory activity. Indeed, this protein exhibits structural features that are distinct from the kinds of changes typically associated with adaptation to destabilising environments. Instead, it is proposed that the folding landscape has been smoothed, reducing opportunities for populating aggregation-prone intermediate states<sup>34</sup>.

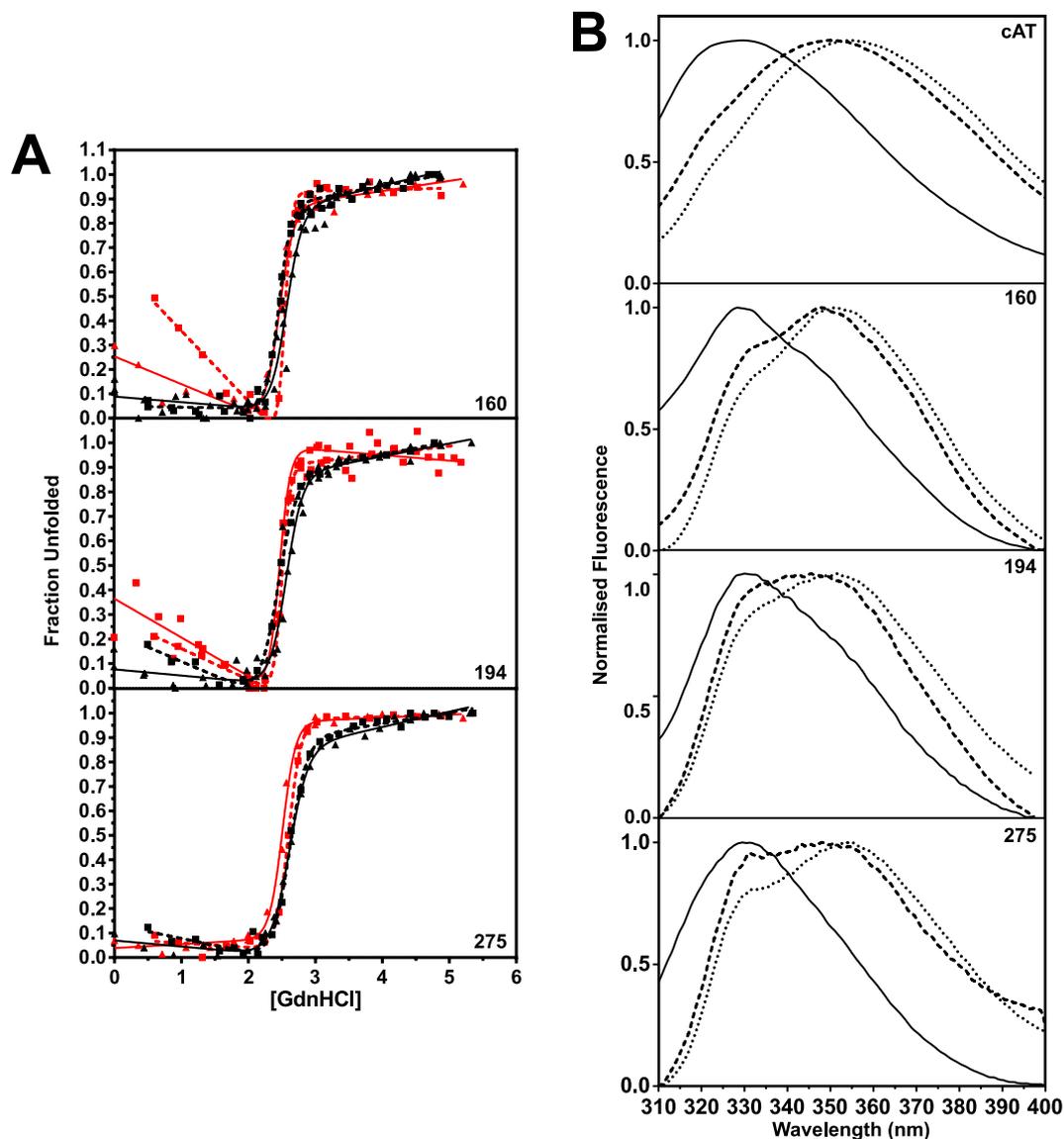
The whole-protein consensus approach to the design of conserpin was simple: each residue corresponds to the most frequently observed amino acid at each position in an alignment of eukaryotic serpins. The underlying premise is that, over the course of evolution, extensive residue-level sampling has occurred, and changes which are the least incompatible with function and stability will be over-represented in modern-day sequences<sup>52</sup>. This is an approach distinct from the derivation of an ancestral sequence, which reconstructs a protein based on phylogenetic relationships, and which by definition attempts to negate the influence of changes in sequence that have occurred subsequent to the chosen branch-point. Thus it is unsurprising that there is a higher representation of residues in conserpin from the most populous phylogenetic clades, α<sub>1</sub>-AT-like clade A and intracellular clade B (Fig. 1B).

While the consensus requirements of the N-terminal RCL hinge region are well-known<sup>53</sup>, the specificity of a serpin is reliant on an effective substrate-like interaction between the C-terminal region of the RCL and a target protease. This in turn depends on protease subsite preferences and active site topology<sup>54</sup>, in which both RCL sequence and length play a role. Thus the observation that around 45% of conserpin molecules are cleaved non-productively may be a consequence of an RCL sequence built on consensus principles. To test this, and generate a tool protein with a known inhibitory profile, the specificity-determining residues of the α<sub>1</sub>-AT RCL were introduced onto the conserpin scaffold. Whilst exhibiting a comparable rate of association for α<sub>1</sub>-AT with chymotrypsin, and a higher degree of inhibition with respect to conserpin, this change did not convert the resulting protein into a stoichiometric inhibitor (Fig. 2A and C). Additionally, the inhibitory complex exhibited compromised stability with respect to α<sub>1</sub>-AT, with a gradual regain of protease activity at an accelerated rate (Fig. 2D). The basis for this is not immediately clear; the two have a comparable RCL length and the resting place of the inhibited protease is structurally similar in both. It has been shown that the protease catalytic triad can be distorted to various extents in some complexes<sup>55</sup> and that partial translocation of a protease results in a relatively unstable inhibitory



**Figure 5.** Properties of single-tryptophan mutants of cAT. (A) Residues and structural elements proximate to tryptophans 160, 194 and 275 ( $\alpha_1$ -AT numbering) present in the structure of native conserpin (5CDX) are shown. (B) Far-UV spectra of cAT<sub>W160</sub> (blue) cAT<sub>W194</sub> (red) and cAT<sub>W275</sub> (green) are shown with respect to cAT (black dashes), at a concentration of 0.2 mg mL<sup>-1</sup>, using a path length of 0.1 cm, at 25 °C. (C) Fluorescence emission spectra for cAT (black), cAT<sub>W160</sub> (blue) cAT<sub>W194</sub> (red) and cAT<sub>W275</sub> (green), in the presence of 6 M of GdnHCl (*upper panel*) and 8 M urea (*lower panel*). The dashed spectrum is the summation of all tryptophan mutants. Sample fluorescence, with excitation at 295 nm and 5 nm slit widths, was buffer-corrected. (D) As in panel C, but in the absence of denaturant. (E) Single-tryptophan variants cAT<sub>W160</sub> (blue) cAT<sub>W194</sub> (red) and cAT<sub>W275</sub> (green) at 0.2 mg mL<sup>-1</sup>, heated from 25 °C to 95 °C at a rate of 1 °C min<sup>-1</sup>, with unfolding monitored by the change in CD signal at 222 nm. (F) As in panel E, performed in the presence of 2 M GdnHCl.

state<sup>56</sup>. It is possible that the instability is due to an impedance of full translocation, or an otherwise incompletely inhibited catalytic triad. Thus, it appears that the less efficient inhibitory mechanism is a trade-off for the gain in stability.



**Figure 6.** Equilibrium unfolding of single-tryptophan variants of cAT. **(A)** Equilibrium GdnHCl-mediated unfolding (solid lines) and refolding (dotted lines) for cAT<sub>W160</sub>, cAT<sub>W194</sub> and cAT<sub>W275</sub> followed by the change in CD signal at 222 nm (black lines) and intrinsic fluorescence at 330 nm (red lines). Each dataset is the result of at least three independent experiments. A two-state unfolding curve was fitted to the data. The concentration at which half the protein is unfolded ( $D_{50\%}$ ) or refolded ( $D_{50\% \text{ refold}}$ ) is shown in Table 1. **(B)** Normalised fluorescence emission spectra of parental cAT, cAT<sub>W160</sub>, cAT<sub>W194</sub> and cAT<sub>W275</sub> in buffer without denaturant (—), and in the presence of 2.6 M or 3 M of GdnHCl (---) and 6 M GdnHCl (---). All scans were conducted at 25 °C, with an excitation wavelength of 295 nm and slit widths of 5 nm.

Like its progenitor protein, cAT exhibited a remarkable thermal stability at odds with the eukaryotic serpins from which the consensus sequence was derived (Fig. 3A); the only naturally-occurring serpin characterised to date with a higher stability is aeropin, produced by an archaeon that lives at temperatures close to 100 °C<sup>18</sup>. This was associated with a pronounced increase in the kinetic barrier to polymerisation (Fig. 3C), while thermodynamic stability with respect to the unfolded state was partially attributable to the three tryptophan residues in helix F, at the top of  $\beta$ -sheet A, and helix H (Fig. 5E). The lack of distinction between denaturant and thermal measures of stability (Figs 5E,F and 6A) reinforces a contribution to global stability and the high degree of cooperativity between the three locales during folding. It has been reported that introduction of a tryptophan at position 160 of  $\alpha_1$ -AT elevated the  $T_m$  from 59 to 65 °C<sup>48</sup>; correspondingly, this side-chain contributes to an enhanced packing with  $\beta$ -sheet A in the conserpin structure<sup>34</sup>. Trp194 is the most highly conserved of the three, situated in the breach region, and substitution with a phenylalanine residue has been found to decrease the kinetic stability of  $\alpha_1$ -antichymotrypsin<sup>57</sup>, although not the thermodynamic stability of  $\alpha_1$ -AT<sup>47</sup>. The loss of Trp275 does not compromise reversibility or lead to a disproportionate impact on thermodynamic stability with respect to the other tryptophan residues; thus, it does not appear to be required for the putative folding nucleus (Fig. 5C, lower panel).

The equilibrium unfolding and refolding profiles of cAT and its tryptophan variants were consistent with those observed for conserpin: they exhibited a two-state transition, showed full reversibility, and exhibited a high degree of cooperativity (Figs 4 and 6A). In contrast, all characterised eukaryotic serpins have been found to unfold via an intermediate species<sup>18</sup> often described as molten globule-like in structure<sup>51</sup>, and no other eukaryotic serpin has been shown to unfold reversibly and refold fully. Accordingly, intrinsic fluorescence spectra of individual tryptophan residues and analytical size exclusion chromatography confirmed the presence of a distinct expanded intermediate ensemble (Figs 4 and 6B). The resolution of two novel components over a narrow denaturant range is consistent with indirect evidence of intermediates inferred from rapid folding of conserpin<sup>34</sup> and with the observation of distinct transition midpoints in plasma  $\alpha_1$ -AT reported by different spectroscopic approaches<sup>58</sup>. The concurrent appearance of these intermediates at equilibrium under restrictive denaturant concentrations suggests that they are interconvertible, directly or indirectly. In the context of folding from a denatured state, this further supports a highly efficient collapse from expanded to compact ensembles.

Polymers and the inactive latent conformation form via intermediate states that can be induced under destabilising conditions<sup>23,39,40,42</sup>. In  $\alpha_1$ -AT, the degree of native state perturbation is associated with generation of two different polymer configurations, attained through two different intermediates - a 'denaturant-induced' expanded molten globule unfolding intermediate<sup>44</sup> and a compact 'heat-induced' polymerisation intermediate<sup>45,46</sup>. Whilst definitive equivalence cannot be determined from these data, it remains noteworthy that the hydrodynamic volumes of cAT intermediate species are consistent with those implicated in  $\alpha_1$ -AT polymerisation. Further investigation would be required to establish whether the intermediate states identified here represent branch points leading to different conformational outcomes. Nevertheless, with intermediates apparently representing an obligate component of the serpin folding pathway, it is most likely their low abundance that asserts the greatest influence on thermal stability and reversibility of unfolding of cAT. Indeed, our data combined with published observations of conserpin<sup>34</sup> and aeropin<sup>18</sup> support a general mechanism in which serpin stability against misfolding and polymerisation is achieved not by elimination of such intermediates, but by limiting the extent to which they are populated.

When considering mechanisms of inactivation, a noteworthy divergence in behaviour of cAT from conserpin is the appearance of higher-order species during prolonged heating at high temperature (Fig. 3B). Under these conditions the parental conserpin protein also undergoes an inactivating change, but in contrast does so in a unimolecular fashion, resulting in formation of the latent species<sup>34</sup>. The prevailing evidence is that polymerisation and latency arise through a common process: both have been observed to occur simultaneously with neuroserpin<sup>42</sup>, Z  $\alpha_1$ -AT and M  $\alpha_1$ -AT in the presence of citrate<sup>41,59</sup>, and thus have been proposed to represent alternate outcomes determined by a decision point along the pathway<sup>42</sup>. Structurally, both forms require a mobile strand 1 of  $\beta$ -sheet C, and at least partial insertion of the RCL as an additional central 6<sup>th</sup> strand in  $\beta$ -sheet A<sup>14,21,60,61</sup>. As cAT and conserpin undergo conformational change with a similar half-time of around 1-2 hours under equivalent conditions, the substitution of the  $\alpha_1$ -AT RCL residues has not made cAT substantially more prone to inactivation, but instead shifted the balance towards the formation of oligomers. A basis for this could be greater compatibility of the P<sub>7</sub>-P<sub>1</sub>' RCL residues with the inserted state, in line with an improved stoichiometry of inhibition relative to the parental conserpin protein (Fig. 2A). Whether this is the case or not, these data indicate that an attempt to further limit inactivating conformational change should focus in the first instance on residues outside of the RCL.

In  $\alpha_1$ -AT, the intermediate ensemble has been characterised as partially folded, possessing an intact  $\beta$ -sheet B and helices G and H, an expanded  $\beta$ -sheet A and a disrupted helix F<sup>47,48,51,62,63</sup>. This is compatible with the presence of residual structure in the vicinity of Trp275 (on helix H) in 8 M urea and the disordered breach and helix F regions (Trp160 and Trp194 respectively) observed here. By extension, this suggests that not only is the native structure and function of the consensus serpin protein conserved, but the folding pathway is as well.

The development of a synthetic serpin using a consensus strategy was found to yield a functional protein that is highly thermostable, resistant to polymerisation and able to fold reversibly. This study has provided direct evidence for two transiently populated intermediates, and the use of single-tryptophan variants has revealed the folding pathway to exhibit a high degree of cooperativity and conservation of the folding nucleus. Substitution of the RCL has improved inhibitory efficiency but suggests that further optimisation of the conserpin core may be necessary to maximise this. Even so, with a fully reversible folding pathway, pronounced stability, and adaptable specificity, conserpin and the derivatives characterised here provide a useful platform for the investigation of conformational change and stability in the serpin superfamily.

## Methods

**Materials and software.** All reagents were from Sigma Aldrich unless specified. The concentration of stock solutions of guanidine hydrochloride (GdnHCl) was determined using refractive index measurements<sup>64</sup>. Molecular weight markers were from Life Technologies and Fermentas. A primary mouse anti-histidine tag antibody (AbD Serotec) and secondary sheep anti-mouse antibody (Chemicon) were used for western blots. Bovine  $\alpha$ -chymotrypsin was stored in 1 mM HCl. Human neutrophil elastase (HNE) was from Calbiochem and prepared in 50 mM sodium acetate, 200 mM NaCl, pH 5. Non-linear regression and numerical calculation of reaction half-times was performed using Prism (GraphPad Inc.) and GNU Octave. Structural representations were generated using Pymol (Schrodinger Inc.).

**Design of the consensus serpin (cAT).** The engineering of the conserpin sequence has been described previously<sup>34</sup>. Residues P<sub>7</sub>-P<sub>1</sub>' (GVEIVPRS) were replaced with the P<sub>7</sub>-P<sub>2</sub>' region of the  $\alpha_1$ -AT RCL (FLEAIPMSI). The constructs used for the single tryptophan mutation experiments were codon-optimised for *E. coli* and commercially synthesised by DNA 2.0 (USA).

**Expression constructs.** Plasmids encoding cAT and the mutants cAT<sub>W160</sub> (W194F/W275F), cAT<sub>W194</sub> (W160F/W275F), and cAT<sub>W275</sub> (W160F/W194F) were generated using ligation-independent cloning with the pLIC-HIS vector using standard protocols; these constructs were transformed into BL21(DE3) *E. coli* and subjected to small scale expression<sup>65</sup>. Colonies were screened for expression using a crude activity assay in which 90  $\mu$ L of the lysate was mixed with 10  $\mu$ L of 10  $\mu$ M of chymotrypsin.

**Protein expression and purification.** Protein was expressed using Overnight Express™ Instant TB Medium (Merck-Millipore) as described previously<sup>65</sup>. Cells were harvested, lysed in 40 mL lysis buffer (10 mM imidazole, 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0) supplemented with 2 mM  $\beta$ -mercaptoethanol, 0.125 mM PMSF, 0.25 mg mL<sup>-1</sup> lysozyme and 1 mg of DNase I. Following centrifugation, the soluble fraction was filtered through a 0.22  $\mu$ m filter membrane (Millipore) and applied to a pre-equilibrated 5 ml HisTrap HP column (GE Healthcare), washed with 6 volumes of 20 mM imidazole, 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0 and eluted with 500 mM imidazole, 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0. Peak fractions were loaded onto a Superdex 200 16/60 column and eluted with 50 mM Tris, 90 mM NaCl, pH 8.0; protein was concentrated to ~300  $\mu$ M and stored at -80 °C.

**Characterisation of inhibitory properties.** The SI and association rate constant ( $k_{\text{ass}}$ ) of cAT and single tryptophan variants against bovine chymotrypsin was determined as described previously<sup>66</sup> using protease assay buffer (20 mM Tris, 100 mM NaCl, 0.1% (w/v) PEG 8000, 10 mM CaCl<sub>2</sub>, pH 8.0) and 200  $\mu$ M N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide substrate. The refolding SI was measured using protein unfolded in 6 M of GdnHCl, 90 mM NaCl and 50 mM Tris, pH 8.0 for 2 hours, and refolded into 50 mM Tris, 90 mM NaCl, pH 8.0, so the final concentration of the GdnHCl was 0.2 M, for an additional 2 hours. For complex dissociation experiments, 5  $\mu$ M serpin-chymotrypsin complex was diluted 200–2000-fold into assay buffer containing substrate, and the resulting progress curves describing the regain of activity were fit by a quadratic equation as described<sup>67</sup> with the turnover number calculated separately for each discrete experiment.

**Heat-induced oligomerisation.** The disappearance of monomer was analysed by incubating 10  $\mu$ L of 10  $\mu$ M of cAT protein samples at 75 °C for different times. 3  $\mu$ L of the protein sample was then added to 1  $\mu$ L of native-PAGE loading buffer and resolved by 6% native-PAGE as described<sup>68</sup>. For FRET experiments,  $\alpha_1$ -AT or cAT, in phosphate-buffered saline (PBS), were incubated with a 2–3-fold molar excess of either NHS-Atto 488 or NHS-Atto 594 dyes (Atto-Tec, Germany) in separate reactions for 4 hours at 25 °C. Following the quenching of the reaction by 10 mM hydroxylamine, the preparations were separated from unconjugated dye by anion exchange chromatography as described<sup>66</sup>. Polymerisation experiments, with protein at a concentration of 0.1 mg mL<sup>-1</sup> in PBS and a total volume of 20  $\mu$ L, were performed using a Mastercycler Realplex4 instrument (Eppendorf) and the resulting curves were well-described by a double exponential equation as noted previously<sup>66</sup>. The temperature dependence of these reaction half-times was assessed using an Arrhenius plot, with the apparent activation energy ( $E_{\text{act,app}}$ ) calculated by multiplying the slope of the regression by the gas constant.

**Circular dichroism scans and thermal denaturation.** Circular dichroism (CD) measurements were performed on a Jasco J-815 CD spectrometer (Jasco) at a protein concentration of 0.2 mg mL<sup>-1</sup> with 90 mM NaCl and 50 mM Tris, pH 8.0 using a quartz cell with a path-length of 0.1 cm. Far-UV scans were performed between 200–260 nm for samples in 90 mM NaCl and 50 mM Tris, pH 8.0. For thermal denaturation, a heating rate of 1 °C min<sup>-1</sup> from 25 °C to 95 °C was used, with the change in signal measured at 222 nm. Refolding was measured directly after the thermal melt by holding the temperature at 95 °C for 1 min before the temperature was decreased to 25 °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<sup>69</sup> for both forward and reverse thermal denaturation experiments.

**Size exclusion chromatography (SEC).** SEC was performed on a Superose 12 10/300 gel filtration column equilibrated with the respective buffer that the protein was incubated in: 0 M, 2 M, 2.8 M, 4 M or 6 M GdnHCl in 90 mM NaCl and 50 mM Tris, pH 8.0. The elution profiles, reflecting the absorbance at 280 nm, were fit to the sum of one or more Gaussian functions.

**Tryptophan fluorescence scans.** Intrinsic fluorescence was measured using a FluoroMax-4 spectrofluorometer (HORIBA Jobin Yvon) with 0.5  $\mu$ M of each protein in a 1 cm path-length quartz cell at 25 °C. The excitation wavelength ( $\lambda_{\text{ex}}$ ) was 295 nm and the emission wavelength ( $\lambda_{\text{em}}$ ) was 330 nm, with 5 nm slit widths.

**Equilibrium unfolding and refolding.** Equilibrium unfolding was performed as described<sup>18</sup> using both CD and intrinsic fluorescence. Refolding curves were obtained through first unfolding the protein in 6 M of GdnHCl, 90 mM NaCl and 50 mM Tris, pH 8.0 for 2 hours and then refolding back into buffer containing various concentration of GdnHCl with 50 mM Tris, 90 mM NaCl, pH 8.0 for an additional 2 hours. The fraction of unfolded was plotted against the final GdnHCl concentration.

**Sequence and structural comparisons.** MEGA<sup>670</sup> was used to calculate the average pairwise sequence identity between conserpin and aligned eukaryotic serpins separated into phylogenetic clades<sup>1</sup>. To produce sub-alignments in which the most variable sites were removed at different thresholds, positions were ranked by their Kabat variability score (calculated as the number of amino acids at a site  $\div$  the frequency of the most common amino acid). For structural comparisons, the native (PDB 5CDX) and latent (PDB 5CDZ) conformations of

conserpin<sup>34</sup> were aligned to native and loop-inserted serpin structures, respectively, using SUPERPOSE<sup>71</sup>. Two rounds of superpositions were performed: following the first, those residue positions in conserpin that failed to match in any pairwise comparison were excluded the subsequent round, following which root-mean square deviations (RMSD) were calculated.

## References

- 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 Res* **10**, 1845–1864 (2000).
- Irving, J. A. *et al.* Serpins in prokaryotes. *Mol Biol Evol* **19**, 1881–1890 (2002).
- Ragg, H., Lokot, T., Kamp, P. B., Atchley, W. R. & Dress, A. Vertebrate serpins: construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses. *Mol Biol Evol* **18**, 577–584 (2001).
- Ray, C. A. *et al.* Viral inhibition of inflammation: cowpox virus encodes an inhibitor of the interleukin-1 beta converting enzyme. *Cell* **69**, 597–604 (1992).
- Schick, C. *et al.* Cross-class inhibition of the cysteine proteases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a kinetic analysis. *Biochemistry* **37**, 5258–5266 (1998).
- Irving, J. A. *et al.* Evidence that serpin architecture intrinsically supports papain-like cysteine protease inhibition: engineering alpha(1)-antitrypsin to inhibit cathepsin proteases. *Biochemistry* **41**, 4998–5004 (2002).
- Dafforn, T. R., Della, M. & Miller, A. D. The molecular interactions of heat shock protein 47 (hsp 47) and their implications for collagen biosynthesis. *J Biol Chem* **276**, 49310–49319 (2001).
- Zou, Z. *et al.* Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. *Science* **263**, 526–529 (1994).
- Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230, <https://doi.org/10.1126/science.181.4096.223> (1973).
- Seo, E. J., Im, H., Maeng, J. S., Kim, K. E. & Yu, M. H. Distribution of the native strain in human alpha 1-antitrypsin and its association with protease inhibitor function. *J Biol Chem* **275**, 16904–16909, <https://doi.org/10.1074/jbc.M001006200> (2000).
- Powell, L. M. & Pain, R. H. Effects of glycosylation on the folding and stability of human, recombinant and cleaved alpha 1-antitrypsin. *J Mol Biol* **224**, 241–252 (1992).
- Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C. & Bottomley, S. P. Kinetic instability of the serpin Z alpha1-antitrypsin promotes aggregation. *J Mol Biol* **396**, 375–383, <https://doi.org/10.1016/j.jmb.2009.11.048> (2010).
- Yamasaki, M., Li, W., Johnson, D. J. & Huntington, J. A. Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* **455**, 1255–1258, <https://doi.org/10.1038/nature07394> (2008).
- 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, <https://doi.org/10.1038/357605a0> (1992).
- Mast, A. E., Enghild, J. J. & Salvesen, G. Conformation of the reactive site loop of alpha 1-proteinase inhibitor probed by limited proteolysis. *Biochemistry* **31**, 2720–2728 (1992).
- Knaupp, A. S. & Bottomley, S. P. Structural change in beta-sheet A of Z alpha(1)-antitrypsin is responsible for accelerated polymerization and disease. *J Mol Biol* **413**, 888–898, <https://doi.org/10.1016/j.jmb.2011.09.013> (2011).
- Zhang, Q. *et al.* The N terminus of the serpin, tengpin, functions to trap the metastable native state. *EMBO Rep* **8**, 658–663, <https://doi.org/10.1038/sj.embor.7400986> (2007).
- Cabrita, L. D., Irving, J. A., Pearce, M. C., Whisstock, J. C. & Bottomley, S. P. Aeropin from the extremophile *Pyrobaculum aerophilum* bypasses the serpin misfolding trap. *J Biol Chem* **282**, 26802–26809, <https://doi.org/10.1074/jbc.M705020200> (2007).
- Lomas, D. A. & Mahadeva, R. Alpha-1-Antitrypsin polymerization and the serpinopathies: pathobiology and prospects for therapy. *Journal of Clinical Investigation* **110**, 1585–1590, <https://doi.org/10.1172/Jci200216782> (2002).
- Gooptu, B. *et al.* Crystallographic and cellular characterisation of two mechanisms stabilising the native fold of alpha1-antitrypsin: implications for disease and drug design. *J Mol Biol* **387**, 857–868, <https://doi.org/10.1016/j.jmb.2009.01.069> (2009).
- Chang, W. S. *et al.* Importance of the release of strand 1C to the polymerization mechanism of inhibitory serpins. *Protein Sci* **6**, 89–98, <https://doi.org/10.1002/pro.5560060110> (1997).
- Baek, J. H. *et al.* Probing the local conformational change of alpha 1-antitrypsin. *Protein Sci* **16**, 1842–1850, <https://doi.org/10.1110/ps.072911607> (2007).
- Irving, J. A., Haq, I., Dickens, J. A., Faull, S. V. & Lomas, D. A. Altered native stability is the dominant basis for susceptibility of alpha1-antitrypsin mutants to polymerization. *Biochem J* **460**, 103–115, <https://doi.org/10.1042/BJ20131650> (2014).
- Devlin, G. L., Parfrey, H., Tew, D. J., Lomas, D. A. & Bottomley, S. P. 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**, 727–732, <https://doi.org/10.1165/ajrcmb.24.6.4407> (2001).
- Sharp, L. K. *et al.* Sugar and alcohol molecules provide a therapeutic strategy for the serpinopathies that cause dementia and cirrhosis. *FEBS J* **273**, 2540–2552, <https://doi.org/10.1111/j.1742-4658.2006.05262.x> (2006).
- Lee, K. N., Im, H., Kang, S. W. & Yu, M.-H. Characterization of a human alpha1-antitrypsin variant that is as stable as ovalbumin. *J Biol Chem* **273**, 2509–2516 (1998).
- Seo, E. J., Lee, C. & Yu, M. H. Concerted regulation of inhibitory activity of alpha 1-antitrypsin by the native strain distributed throughout the molecule. *J Biol Chem* **277**, 14216–14220, <https://doi.org/10.1074/jbc.M110272200> (2002).
- Im, H., Seo, E. J. & Yu, M.-H. Metastability in the inhibitory mechanism of human alpha1-antitrypsin. *J Biol Chem* **274**, 11072–11077 (1999).
- 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).
- Steipe, B., Schiller, B., Pluckthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol* **240**, 188–192, <https://doi.org/10.1006/jmbi.1994.1434> (1994).
- Knappik, A. *et al.* Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol* **296**, 57–86, <https://doi.org/10.1006/jmbi.1999.3444> (2000).
- Stumpp, M. T., Forrer, P., Binz, H. K. & Pluckthun, A. Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J Mol Biol* **332**, 471–487 (2003).
- Anbar, M., Gul, O., Lamed, R., Sezerman, U. O. & Bayer, E. A. Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. *Appl Environ Microbiol* **78**, 3458–3464, <https://doi.org/10.1128/AEM.07985-11> (2012).
- Porebski, B. T. *et al.* Smoothing a rugged protein folding landscape by sequence-based redesign. *Sci Rep* **6**, 33958, <https://doi.org/10.1038/srep33958> (2016).
- Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157–162 (1967).
- Zhou, A., Carrell, R. W. & Huntington, J. A. The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. *J Biol Chem* **276**, 27541–27547, <https://doi.org/10.1074/jbc.M102594200> (2001).
- Huntington, J. A., Read, R. J. & Carrell, R. W. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**, 923–926, <https://doi.org/10.1038/35038119> (2000).

38. Lawrence, D. A. *et al.* Serpin-Protease complexes are trapped as stable acyl-enzyme intermediates. *Journal of Biological Chemistry* **270**, 25309–25312 (1995).
39. 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).
40. James, E. L. & Bottomley, S. P. The mechanism of alpha1-antitrypsin polymerization probed by fluorescence spectroscopy. *Arch Biochem Biophys* **356**, 296–300 (1998).
41. Lomas, D. A., Elliott, P. R., Chang, W. S., Wardell, M. R. & Carrell, R. W. Preparation and characterization of latent alpha 1-antitrypsin. *J Biol Chem* **270**, 5282–5288 (1995).
42. Chiou, A. *et al.* Probing neuroserpin polymerization and interaction with amyloid-beta peptides using single molecule fluorescence. *Biophys J* **97**, 2306–2315, <https://doi.org/10.1016/j.bpj.2009.07.057> (2009).
43. 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, <https://doi.org/10.1074/jbc.M004310200> (2000).
44. Tsutsui, Y., Dela Cruz, R. & Wintrobe, P. L. Folding mechanism of the metastable serpin alpha1-antitrypsin. *Proc Natl Acad Sci USA* **109**, 4467–4472, <https://doi.org/10.1073/pnas.1109125109> (2012).
45. Tsutsui, Y., Kuri, B., Sengupta, T. & Wintrobe, P. L. The structural basis of serpin polymerization studied by hydrogen/deuterium exchange and mass spectrometry. *J Biol Chem* **283**, 30804–30811, <https://doi.org/10.1074/jbc.M804048200> (2008).
46. Ekeowa, U. I. *et al.* Defining the mechanism of polymerization in the serpinopathies. *Proc Natl Acad Sci USA* **107**, 17146–17151, <https://doi.org/10.1073/pnas.1004785107> (2010).
47. Tew, D. J. & Bottomley, S. P. 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, <https://doi.org/10.1006/jmbi.2001.5104> (2001).
48. 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, <https://doi.org/10.1021/bi0158932> (2002).
49. Verheyden, S., Sillen, A., Gils, A., Declerck, P. J. & Engelborghs, Y. Tryptophan properties in fluorescence and functional stability of plasminogen activator inhibitor 1. *Biophys J* **85**, 501–510, [https://doi.org/10.1016/S0006-3495\(03\)74495-6](https://doi.org/10.1016/S0006-3495(03)74495-6) (2003).
50. 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, <https://doi.org/10.1016/j.bbrc.2004.09.105> (2004).
51. Tsutsui, Y. & Wintrobe, P. L. 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, <https://doi.org/10.1016/j.jmb.2007.05.039> (2007).
52. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng Des Sel* **29**, 245–251, <https://doi.org/10.1093/protein/gzw015> (2016).
53. Hopkins, P. C., Carrell, R. W. & Stone, S. R. Effects of mutations in the hinge region of serpins. *Biochemistry* **32**, 7650–7657 (1993).
54. Ye, S. *et al.* The structure of a Michaelis serpin-protease complex. *Nat Struct Biol* **8**, 979–983, <https://doi.org/10.1038/nsb1101-979> (2001).
55. Plotnick, M. I. *et al.* Heterogeneity in serpin-protease complexes as demonstrated by differences in the mechanism of complex breakdown. *Biochemistry* **41**, 334–342 (2002).
56. Liu, L., Mushero, N., Hedstrom, L. & Gershenson, A. Short-lived protease serpin complexes: partial disruption of the rat trypsin active site. *Protein Sci* **16**, 2403–2411, <https://doi.org/10.1110/ps.07311207> (2007).
57. Pearce, M. C., Cabrita, L. D., Ellisdon, A. M. & Bottomley, S. P. The loss of tryptophan 194 in antichymotrypsin lowers the kinetic barrier to misfolding. *FEBS J* **274**, 3622–3632, <https://doi.org/10.1111/j.1742-4658.2007.05897.x> (2007).
58. Herve, M. & Ghelis, C. Conformational changes in intact and papain-modified alpha 1-proteinase inhibitor induced by guanidinium chloride. *Eur J Biochem* **191**, 653–658 (1990).
59. Tan, L. *et al.* Characterising the association of latency with alpha(1)-antitrypsin polymerisation using a novel monoclonal antibody. *Int J Biochem Cell Biol* **58**, 81–91, <https://doi.org/10.1016/j.biocel.2014.11.005> (2015).
60. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. & Huntington, J. A. Molecular basis of alpha1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep* **12**, 1011–1017, [embor.2011.171](https://doi.org/10.1038/embor.2011.171) (2011).
61. Mottonen, J. *et al.* Structural basis of latency in plasminogen activator inhibitor-1. *Nature* **355**, 270–273 (1992).
62. 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).
63. 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).
64. Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266–280 (1986).
65. Cabrita, L. D., Dai, W. & Bottomley, S. P. A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol* **6**, 12, <https://doi.org/10.1186/1472-6750-6-12> (2006).
66. Haq, I. *et al.* Reactive centre loop mutants of alpha-1-antitrypsin reveal position-specific effects on intermediate formation along the polymerization pathway. *Biosci Rep* **33**, <https://doi.org/10.1042/BSR20130038> (2013).
67. Calugaru, S. V., Swanson, R. & Olson, S. T. The pH dependence of serpin-proteinase complex dissociation reveals a mechanism of complex stabilization involving inactive and active conformational states of the proteinase which are perturbable by calcium. *J Biol Chem* **276**, 32446–32455, <https://doi.org/10.1074/jbc.M104731200> (2001).
68. Stone, S. R., Dennis, S. & Hofsteenge, J. Quantitative evaluation of the contribution of ionic interactions to the formation of the thrombin-hirudin complex. *Biochemistry* **28**, 6857–6863 (1989).
69. Knaupp, A. S. *et al.* The roles of helix I and strand 5A in the folding, function and misfolding of alpha1-antitrypsin. *PLoS One* **8**, e54766, <https://doi.org/10.1371/journal.pone.0054766> (2013).
70. Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163 (2004).
71. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256–2268, <https://doi.org/10.1107/S0907444904026460> (2004).

## Acknowledgements

We would like to thank Shani B. Keleher for consultations. This work was supported by a NHMRC National Health and Medical Research Council Program Grant and a NHMRC Principal Research Fellowship to SPB; LY was supported by a Monash Graduate Scholarship. This work was funded in part by a Project Grant from the Alpha-1 Foundation to JAI.

## Author Contributions

L.Y., S.P.B. and J.A.I. designed research; L.Y., J.A.I. and W.D. performed research; L.Y., J.A.I., S.P.B. and M.I.A. analysed data; L.Y., J.A.I. and M.I.A. wrote and W.D. proofread the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-19567-9>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018