HOW THE SERPIN $\alpha_1$-PROTEINASE INHIBITOR FOLDS

Klavs Dolmer and Peter G.W. Gettins

Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, USA.

Address correspondence to: Peter G.W. Gettins, Department of Biochemistry & Molecular Genetics, M/C 669, University of Illinois at Chicago, 900 S. Ashland, Chicago, Illinois 60607, USA. Phone; + 312 996 5534, e-mail: pgettins@uic.edu.

Running title: Serpin folding pathway

**BACKGROUND:** Functional serpins uniquely adopt a metastable conformation by an unknown folding pathway.

**RESULTS:** Ability of constituent $\alpha_1$-proteinase inhibitor ($\alpha_1$PI) peptides to associate reveals the order of folding.

**CONCLUSION:** Metastability results from inability of sheet A strand 4 to efficiently insert before completion of C-terminal sheet B.

**SIGNIFICANCE:** Pathway helps explains the nature of the polymerogenic intermediate of the Z variant of $\alpha_1$PI.

**SUMMARY**

Serpins are remarkable and unique proteins in being able to spontaneously fold into a metastable conformation without aid of a chaperone or prodomain. This metastable conformation is essential for inhibition of proteinases, so that massive serpin conformational change, driven by the favorable energetics of relaxation of the metastable conformation to the more stable one, can kinetically trap the proteinase-serpin acylenzyme intermediate. Failure to direct folding to the metastable conformation would lead to inactive, latent serpin. How serpins fold into such a metastable state is unknown. Using the ability of component peptides from the serpin $\alpha_1$PI to associate, we have now elucidated the pathway by which this serpin efficiently folds into its metastable state. In addition we have established the likely structure of the polymerogenic intermediate of the Z variant of $\alpha_1$PI.

Serpins, of which $\alpha_1$-proteinase inhibitor ($\alpha_1$PI, also frequently called $\alpha_1$-antitrypsin) is a very abundant and well studied example, are large proteinase inhibitors that possess a common core fold composed of three $\beta$-sheets and 8 or 9 $\alpha$-helices (Fig 1A). They can adopt two major conformations, native and latent. In the native state, the reactive center loop (RCL)$^1$, which contains the primary proteinase recognition site, is exposed and poised to interact with its target. Using a nomenclature in which strands of a given $\beta$-sheet X are represented as s#X, the exposed RCL is linked between s5A on the N-terminal end and s1C on the C-terminal end (Fig 1A). In the more stable latent state the RCL is inserted into $\beta$-sheet A as strand 4 (s4A), and s1C is absent from...
β-sheet C and exposed (1) (Figs 1A and B). In a unique departure from the protein-folding paradigm, the native, functional state of serpins is metastable, while the latent state is the most stable, though inactive, state (2). Although quantitative data on the difference in ΔG between these states is lacking, it is clearly manifested by a large difference in denaturation temperature between the two states. For the serpin PAI-1 the difference in Tm for unfolding of the two conformations is about 17 deg C (3), while for α1PI it is of similar, though less well-defined, magnitude (4). Surprisingly neither chaperone, nor pro-domain is required for efficient folding to the metastable state. Consequently, serpins without disulfides can be denatured from either native or latent states and refolded to the metastable conformation.

While the question of how proteins fold is of wide general interest, what makes the question of serpin folding in particular a fascinating one is that such metastability is an absolute requirement for the serpin inhibition mechanism. Thus serpins act as suicide substrate inhibitors of serine and cysteine proteinases by using a massive conformational change to translocate the proteinase >70 Å (5) and kinetically trap the acyl-enzyme intermediate by distortion of the proteinase active site (6,7). In this conformational change the exposed RCL (now cleaved by the attacking, but still attached, proteinase) inserts into β-sheet A in the same way as in formation of the latent conformation. The energy required for these changes derives from the metastability of the native conformation, which is thus crucial for function (8,9). As yet, however, the pathway that directs folding to the metastable rather than latent conformation (i.e. what determines that the RCL is exposed rather than immediately incorporated into β-sheet A as its central 4th strand), has not been elucidated.

A separate, but important aspect of serpin folding is that there is a major negative consequence of metastability in that it can lead to pathology when mutations affect the folding pathway. Such is the case for the Z variant of α1PI, where intracellular polymerization leads to liver disease, and emphysema results from reduction in circulating levels of functional α1PI (10), which is the principal inhibitor of neutrophil elastase. Here we have used folding of peptide fragments to identify for the first time the pathway by which the serpin α1PI, folds into the metastable state. By implication we consider that this is likely to be true for all serpins. In addition we have established the likely structure of the polymerogenic folding intermediate of the Z variant, which has implications for how polymerization is initiated.

**EXPERIMENTAL PROCEDURES**

*Protein expression and purification:* Full length His6-tagged α1PI multi-8, cloned in pQE-30 (11,12), was expressed as soluble protein at 25° C and purified on Ni2+ Sepharose (GE), using the manufacturer’s protocol. The eluted α1PI was dialyzed against 20 mM Tris-HCl, pH 8.0, 50 mM NaCl overnight, and further purified by ion exchange chromatography on a Q-Sepharose HP column with a gradient of 50-1000 mM NaCl in 20 mM Tris-HCl, pH 8.0. The A350C and Z (E342K) variants were generated using the Quickchange protocol (Stratagene) on the α1PI multi-8 background. The A350C variant was expressed and purified as for wt α1PI. The Z variant was found exclusively in inclusion bodies and needed to be solubilized in 6M urea prior to purification. Further purification was from the denatured state using Ni2+-Sepharose, using the manufacturer’s protocol, followed by ion-
exchange chromatography on Q-Sepharose in 6 M urea.

DNA encoding the fragments 1-323, 324-394, 1-344 and Z 1-344 were generated from the α1PI multi-8 plasmid by PCR with Expand (Roche), using primers designed to introduce cysteines at positions 323, 324 or 344 (though the latter cysteine was always subsequently blocked with iodoacetamide (IAA)). The 1-323, 1-344 and Z 1-344 fragments, including an N-terminal His6-tag, were cloned in pQE-30, and the 324-394 fragment in pQE-60 (both Qiagen). All fragments were expressed in SG13009 cells (Qiagen) in 2YT medium, and isolated from inclusion bodies. All buffers for purification of the fragments were degassed and β-mercaptoethanol was added (0.1 % vol.) freshly. The 1-323, 1-344 and Z 1-344 fragments were purified by Co2+-affinity chromatography (TALON, Clontech), equilibrated in and loaded using 50 mM NaP, pH 7.4, 300 mM NaCl, 6 M urea, and eluted with 50 mM NaOAc, pH 5.0, 300 mM NaCl, 6 M urea. The eluted fragment was further purified by ion exchange chromatography on Q-Sepharose HP (20 mM Tris-HCl, pH 8.0, 0-1000 mM NaCl, 6 M urea). The 324-394 fragment was dissolved in 20 mM Tris-HCl, pH 8.0, 6 M urea and passed through a Q-Sepharose FF cartridge (GE). The flowthrough, containing the fragment, was dialyzed overnight against 20 mM MES, pH 6.0, 6 M urea. Final purification was on a SP-Sepharose HP column (GE) eluted with a 0-500 mM NaCl gradient in 20 mM MES, pH 6.0, 6 M urea. For NMR experiments, uniformly 15N labeled α1PI, 1-323 and 324-394 were expressed in minimal medium, containing 1g/l 15NH4Cl.

C26 (residues 369-394) was synthesized (Genscript) with an extra C-terminal Cys residue for labeling. C36 (residues 359-394 with an extra C-terminal Cys residue and an N-terminal His tag) was cloned by PCR from α1PI multi-8 and the fragment inserted in pQE30 (Qiagen), modified to contain a TEV cleavage site. C36 was expressed in SG13009 cells (Qiagen) in 2YT medium, grown to OD600=0.6-1.0 at 37° C, and cells were harvested 4-5 hours after induction with 1 mM IPTG. C36 was purified from inclusion bodies under denaturing conditions by Ni-NTA chromatography, followed by Q-HP ion exchange chromatography in 6 M Urea, 20 mM Tris-HCl, pH 8.0, eluting with a 0-1 M NaCl gradient.

For fluorescence studies, including native PAGE and quenching experiments, single-cysteine-containing species (C26, C36 and A350C variant of α1PI) were labeled with 5-IAF using a 10 fold molar excess of fluorophore over peptide, after incubation of the protein with DTT at a 1:1 ratio. Fluorescent fractions of labeled C26 and C36 were purified by reverse phase HPLC and checked by MS. This gave peptides with 1:1 fluorescein to peptide stoichiometries. To obtain the fluorescently-labeled fragment 1-354, with label at Cys350, the IAF-labeled A350C α1PI was cleaved at position 354 in the RCL with papain (1:100, enzyme:α1PI) molar ratio for 1 hour at 37° C. Papain activity was inhibited with IAA, and cleaved α1PI was denatured by adding guanidine HCl to a final concentration of 8 M and incubating the protein at 42° C for an hour. After dialysis against 20 mM Tris-HCl, pH 8.0, 6 M urea, the peptides were separated by Q-Sepharose HP chromatography as described for the 1-323 fragment. This gave a fragment, 1-354, with 0.84:1 fluorescein:protein label at Cys350. For quenching studies on the species obtained by associating C36 with Cys350-labeled 1-354, the folded two-chain species was purified from the folding mixture prior to use to ensure that a homogeneous species was used in the experiment.
Folding procedure: Small scale protein refolding assays were performed by placing a small droplet of denatured α1PI fragment(s) in 6 M guanidine HCl at the bottom of a micro centrifuge tube, and quickly diluting the protein with PBS. Final α1PI concentration was 210 nM (1-323, 1-344 and full length fragments) or 630 nM (C26 and C36). For peptide binding studies, the mixture was left on ice overnight. Refolding experiments involving full length α1PI and α1PI-Z were performed at 37º C. Large scale refolding was performed by slowly dripping denatured α1PI (1-323, 1-323 + 324-394, 1-354, or 1-354 + 359-394) into cold PBS with rapid stirring. After incubation overnight, the protein was concentrated using a Ni2+ NTA cartridge (Qiagen). Two-chain molecules were further purified on a Q-Sepharose HP column, eluted with a 50-1000 mM NaCl gradient in 20 mM Tris-HCl, pH 8.0.

CD and NMR spectroscopy: CD measurements were performed on a Jasco J-710 in 50 mM Na phosphate, pH 7.4 at 23º C. The proteins were diluted to a final concentration of 2 µM, and spectra recorded between 300-190 nm with 5 samplings in 2 mm quartz cells. Spectra are shown normalized to equal concentration. NMR [1H, 15N]-HSQC spectra were recorded on a 900 MHz Bruker US2, equipped with a cryoprobe. Spectra were recorded in 20 mM sodium phosphate, pH 7.4, 50 mM NaCl, supplemented with 10% (vol.) D2O. Sample concentrations were between 100 µM (for 15N wt α1PI) and 20 µM (for 15N 1-323).

Thermal denaturation: To measure the thermal stability of the product obtained from associating C36 with 1-354, the tryptophan fluorescence was followed at 335 nm, with excitation at 280 nm. Intrinsic fluorescence was recorded as each sample (1 µM) was heated in 5º C increments from 30-80º C.

Fluorescence spectroscopy: The environment of the reactive center loop was assessed by measuring the quenchability of a fluorescein label at the P9 (A350C) position. Fluorescein fluorescence at 515 nm was recorded at varying KI concentrations (0-200 mM), with excitation at 494 nm.

Kinetic assays: 50 µl α1PI samples were withdrawn from the refolding mixture at the times indicated, and mixed with one molar equivalent HNE. After incubation at RT for 2 min, the samples were mixed with 1 ml 200 µM AAPVpNA (Sigma) in 100 mM Tris-HCl, pH 8.0, 1 mM EDTA and 0.1% PEG 8000, and assayed by changes in absorbance at 405 nM for 180 sec in a Shimadzu UV-2101PC spectrophotometer at 25º C. Fractional activity was plotted relative to HNE activity in the absence of inhibitor.

Gel scanning: Gels were scanned using a BioRad Geldoc XR+ scanner, with automatic exposure setting. Gel bands were quantified using the supplied Image Lab software, using automatic background subtraction.

Size exclusion chromatography of fluorescein-labeled species: 1-344 or Z 1-344 (273 pmol in 4 or 5 µl respectively of 6M guanidine hydrochloride) was placed in a microfuge tube and renatured by rapid dilution with 1 mL PBS at room temp. After 10 min, 750 µL of the solution was transferred to a tube containing 3µL 66µM IAF-labeled C36. After 5 min incubation 200µL was analyzed on a Superdex 75 column (10 x 300 mm) run at 0.35 mL/min in 20 mM sodium phosphate buffer, pH 7.4, containing 250 mM NaCl and 0.1% PEG 3350. Elution was monitored by fluorescein
fluorescence at 515nm, with excitation at 494nm. The fraction of total fluorescence eluting with the peak corresponding to the position of native α,PI (run separately) is considered to represent C36 non-covalent complex with 1-344 or Z 1-344.

RESULTS AND DISCUSSION

The C-terminal region directs metastable folding

Because residues that change secondary structure between the native and latent conformational states of serpins all lie in the extreme C-terminal portion (see Fig 1), it seemed to us likely that the elements that direct folding to the metastable state would reside there. We therefore examined the ability of a polypeptide representing part of the α,PI C-terminal region to associate with the much longer chain that lies N-terminal to it to form a functional proteinase inhibitor.

The choice of break point was after residue 323. Residues 323 and 324 are both aspartates in an exposed loop at the bottom of β-sheet A, immediately preceding s5A (Fig 1A) and so are unlikely to be critical for folding. Both residues were changed to cysteine and separate polypeptides 1-323 and 324-394 expressed and purified. Refolding of the 1-323 fragment by the same protocol as used for refolding of single-chain α,PI, followed by addition of the 323-394 fragment, gave two species on SDS-PAGE run under non-reducing conditions. However, only a single, two-chain species was observed when reducing agent was added (data not shown), suggesting that the two species differed only by a disulfide between C323 and C324 in one form but not the other. Each of these species was then made in pure form for subsequent characterization. The disulfide-containing form could be obtained as the sole product by allowing oxidation of the two cysteines to proceed for a further day (Fig 2A). The two-chain, non-disulfide-containing form was obtained as a single species by blocking the cysteines with IAA prior to refolding the two chains (Fig 2A).

To show that both oxidized single-chain and IAA-treated two-chain species were correctly-folded, we carried out inhibition studies with human neutrophil elastase (HNE) and porcine pancreatic elastase (PPE). Both α,PI species formed covalent 1:1 SDS-stable proteinase complexes, which are a hallmark for a functional serpin inhibiting by the suicide substrate inhibition mechanism (shown for the HNE reaction in Fig 2) and gave second order rate constants and stoichiometries of inhibition comparable to wild-type α,PI (Table 1). Since the serpin inhibition mechanism has an absolute requirement for the correctly-folded metastable conformation, these findings indicate that the two-chain serpin had the normal metastable fold. Significantly, however, the existence of a two-chain folded species suggested that folding of the smaller fragment onto the larger one did not require a covalent link between them. More importantly, this suggested that there must be a preferred folding pathway that results in association of the smaller C-terminal chain with the pre-folded larger N-terminal chain.

Confirmation of the correct fold of the two-chain species was provided by several additional approaches. CD spectroscopy gave a far UV spectrum indistinguishable from that of wild-type α,PI (Fig 3A), whereas the spectrum of 1-323 had much lower ellipticity, with secondary structure estimated as ~22% α-helix and 27% β-sheet compared to values of 31% and 32% respectively for wild-type α,PI (Fig 3). 2D NMR of two-chain α,PI, with 15N label in either the 323 residue N-terminal chain or the 71 residue C-terminal chain, gave an HSQC spectrum that was well dispersed and had resonances at
positions representing sub-sets of the spectrum of wild-type single chain uniformly-$^{15}$N labeled $\alpha_1$PI, so that superpositioning of spectra of the two two-chain $\alpha_1$PIs, with label alternately in the N- or C-terminal region gave a composite spectrum (Fig 4B) the same as for uniformly-labeled single chain $\alpha_1$PI (Fig 4A). The spectrum of 1-323 alone was more like that of a molten globule, with poor dispersion in the 'H dimension (Fig 4C), though one that, from the CD spectrum must contain much of the final $\beta$-sheet content. These findings are consistent with our hypothesis that the element(s) directing the metastable fold must lie C-terminal to residue 323, since a break at that position did not alter the ability of the two parts to fold together correctly.

The folding pathway

(i) $s5A$ inserts first

The above results suggest that 1-323 adopts much of its critical secondary structure first and only subsequently associates with the C-terminus, which comprises, in order, the secondary structure elements $s5A$, $s4A$, $s1C$, $s4B$ and $s5B$. We then examined which of these elements associates next. To accomplish this we attempted to fold a peptide comprising the residues that form strands $s1C$, $s4B$ and $s5B$ (C36, labeled with fluorescein at the C-terminus) onto pre-folded 1-323. If $s1C/s4B/s5B$ can bind efficiently, without the presence of a completed $\beta$-sheet A, the elimination of $s4A$ and $s5A$ from the C-terminal peptide, i.e. by using C36 (359-394) rather than the 324-394 used above, should not prevent it from associating with 1-323. We found, however, that these chains were minimally capable of associating (Fig 5) under identical conditions to those under which the longer C-terminal peptide (324-394) readily associated with 1-323 to give functional serpin (see above). However, when a longer fragment that contained $s5A$ (1-344) was used with C36 there was strong association of the two polypeptides, detected by fluorescence on native PAGE (Fig 5). These findings clearly indicate that $s5A$ must associate with the remainder of the serpin before $s1C$, $s4B$ and $s5B$ can efficiently associate. This finding is contrary to the recent suggestion that was based on a structure of dimeric antithrombin (13). In that structure the dimer was formed by a swap of strands $s4A$ and $s5A$ between monomers. It was suggested there that formation of the dimer occurred from an intermediate on the normal folding pathway, which implied that $s5A$ must be the last element of the metastable conformation to form since, by microscopic reversibility, it would be the element most easily removed. Our present results, and more recent results from the same group (14) are not consistent with this.

(ii) $s1C$ inserts next followed by $s4B$ and $s5B$

We next sought to determine whether there was further sequential association within the remaining C-terminal elements. We used the same approach of examining the ability of fluorescein-labeled C-terminal peptides to associate with 1-344, but used the smaller 26 residue fragment (C26), comprising only $s4B$ and $s5B$, but lacking $s1C$. Somewhat surprisingly, we found only poor association under the same conditions that C36 associated well (Fig 5). An estimate of the relative amounts of C-terminal peptide associated with 1-344, gave a preference for C36, containing $s1C$, over C26 of about 8-10:1, as judged from the relative intensities of the fluorescent bands of associated complex. This might result from the location of $s1C$ within $\alpha_1$PI. $s1C$ is the outermost strand of $\beta$-sheet C and ends close to where the long $s4B/s5B$ $\beta$-hairpin must insert into the hydrophobic center of $\beta$-sheet B, to complete the sheet. Association of $s1C$ prior
to insertion of the s4B/s5B hairpin, might optimally position the hairpin for insertion. However, it should be noted that the discrimination we observe between C26 and C36 is not absolute. This supports the role of s1C as promoting association, without being an absolute requirement. This may, however, explain the many examples of lowered levels of secreted serpin associated with mutations in s1C (15). As a final experiment, which reinforced the above conclusion of the need for s5A insertion prior to C-terminal association, we examined the ability of C26 to associate with 1-323. As expected, there was negligible binding (Fig 5).

(iii) s4A inserts last
The above folding pathway of s5A insertion prior to association of the unit s1C/s4B/s5B leaves the RCL (s4A) exposed. If it were to insert into β-sheet A efficiently before the C-terminal unit could associate, the resulting serpin would be in the non-inhibitory latent conformation (Fig 1ii). Since the active metastable conformation is the normal end-point, this implies that insertion of s4A into sheet A requires the presence of s1C/s4B/s5B in the structure, so that, at the time that s4A might favorably associate with sheet A it is already constrained at both ends (by s5A and s1C) and is thus prevented from doing so. To test this, we examined the ability of 359-394 (C36) to associate with a longer N-terminal fragment that contained not only s5A, but most of s4A (1-354) (see Fig 1A). These two chains spontaneously associated to give a protein with a CD spectrum identical to that of cleaved α₁PI (i.e. a species with s4A inserted into sheet A) and distinct from that of native α₁PI (Fig 6). The species also migrated on native PAGE the same as for cleaved α₁PI and had greatly enhanced thermal stability compared to native α₁PI (data not shown), suggesting that it had adopted the cleaved conformation, in which s4A has inserted in sheet A to give a 6-stranded sheet. In contrast, the CD spectrum of 1-354 on its own was similar to that of 1-323 and to that of 1-323 + 359-394 (Figs 3 and 6), suggesting that insertion of the RCL (s4A) into β-sheet A has not occurred and that it must require the prior association of s1C, s4B and s5B. This is the critical step that ensures that folding is directed to the metastable state and not to the more stable latent state, since in the normal folding of the full-length α₁PI, association of s1C/s4B/s5B completes the native metastable structure with s4A (the RCL) exposed.

Separate evidence that insertion of s4A into β-sheet A can only occur after s1C/s4B/s5B (359-394) has associated, came from a fluorescein reporter attached to the RCL (s4A) at position 350 (P9). From x-ray structures of wild-type α₁PI it is known that this is a highly solvent-exposed position (16,17). Insertion of the RCL into β-sheet A, with concomitant long-distance movement of the P9 residue as it becomes part of β-sheet A results in a large change in environment for the P9 side chain, as has been found in studies on complex formation with proteinases (18). Accessibility of the fluorophore was probed by KI quenchability. In labeled native α₁PI, the fluorescein at P9 was readily quenched by KI, with KQ of 6.0 M⁻¹ (Fig 7). In labeled, folded, 1-354 the fluorescein was even more accessible, with KQ of 7.1 M⁻¹ (Fig 7), consistent with an RCL that was even more exposed than native α₁PI (as might be expected from the absence of the constraint that is imposed from attachment to s1C in full-length native α₁PI). However, after the C-terminal fragment C36 was added to labeled 1-354 (only residues P1-P4 were missing), and the resulting two-chain labeled species purified to obtain a homogeneous protein, the fluorophore was very much less
accessible to KI, with $K_Q$ of only 4.2 M$^{-1}$, suggesting a much more hindered environment, as expected if strand s4A had now inserted into $\beta$-sheet A (note that the P9 residue would be on the outer face of the sheet and so still be quenchable by KI). Consistent with this species having s4A inserted, the quenching constant for fluorescein at the P9 position of cleaved $\alpha_1$PI was nearly identical (4.0 M$^{-1}$). It should be noted that this experiment, in which a cleaved-like state (s4A inserted) that must proceed through a more native-like state first (s4A exposed) is generated, is analogous to one carried out by others on the RCL-cleaved form of the serpin ovalbumin (19). In that study it was shown that the two chains of RCL-cleaved ovalbumin, after denaturation and separation, could re-associate to form the same cleaved conformation, but that this occurred via a native-like species, i.e. that the last step was insertion of s4A.

(iv) The complete pathway for $\alpha_1$PI. Taken together, the above results suggest the following pathway for the folding of $\alpha_1$PI into the metastable state (Fig 8). First, the N-terminal portion up to residue 323 adopts a conformation with much secondary structure already present, as shown by the CD spectrum, though with conformational flexibility which is likely to result from incomplete $\beta$-sheets A and B, and which is manifested in an NMR spectrum that reflects conformational flux. The next segment to associate is the secondary structural element that occurs immediately following in the primary structure, namely s5A, rather than s1C/s4B/s5B (Fig 8ii). That s1C/s4B/s5B cannot efficiently associate with the large folded N-terminal region until s5A has bound (Fig 8 iv) may well be due to the extensive hydrophobic contact interface between $\beta$-sheets A and B. Thus, if $\beta$-sheet A has not been completed by incorporation of s5A, $\beta$-sheet B would not be able to make the many hydrophobic interactions with the under-side of sheet A that are presumably critical for stability. It is significant that many of the residues in this region are conserved amongst serpins, whether inhibitory or not (20), and so suggests that what is true for the folding of $\alpha_1$PI (M8 form) and ovalbumin is likely to be more generally true for serpins.

At this point, with s5A inserted, the native 5-strand conformation of $\beta$-sheet A has been completed, while the polypeptide from residue 345 to the C-terminus is still not associated (Fig 8 ii). This includes the RCL (s4A) and strands 1C, s4B and s5B. Of these, it is the latter group that next associates. If s1C is absent, the hydrophobic hairpin of s4B/s5B binds much less well, suggesting either that prior association of s1C positions the hairpin correctly to fit into the opening in sheet B, or that a concerted process involving all elements occurs. The resulting structure is then that of the correctly folded native, metastable form of the serpin (Fig 8 iii). Thus, the key step in ensuring that the metastable conformation is adopted is that s4A cannot associate efficiently until s1C/s4B/s5B has done so. As with the requirement for s5A to associate before s1C/s4B/s5B can do so, this is likely to result from the intimate contacts between sheets B and A, which are thus likely to help to stabilize the expanded 6-strand $\beta$-sheet that forms in cleaved $\alpha_1$PI (Fig 8 v). In a timely publication from another group, support for this folding pathway comes from an x-ray structure of a trimeric form of $\alpha_1$PI in which the elements s1C/s4B/s5B from one monomer associate with and complete the structure of a second monomer (14) in such a way that none of these units remains exposed. This is evidence of a folding intermediate in which these elements are not yet associated with the folded N-terminal region, but in which s5A is associated.
Folding of Z variant slowed

The above folding pathway has implications for the folding and polymerogenicity of Z-variant α₁PI. This variant involves mutation of Glu 342 to Lys at the top of s5A (Fig 8iii), which results in a normal folding pathway intermediate having a lengthened half-life, with the consequence that, while the intermediate can still successfully proceed to functional metastable serpin, it can also lead to polymers from abnormal build-up of the concentration of intermediate (21). In homozygotes, such polymers cause liver disease when formed within the hepatocyte (22). Based on our proposed folding pathway, the Z mutation might be expected to make it harder for s5A to fully insert, since another positively charged residue (Lys290) is adjacent to Lys342 at the top of s6A.

To examine this we incubated fluorescein-labeled C36 for 5 min with peptides 1-344 or 1-344 containing the Z mutation and then quantitated the amount of C36 that associated with the 1-344 species by separating peptide from protein by size exclusion chromatography. The first-eluting fluorescein-containing peak, which was confirmed to also contain 1-344 by SDS-PAGE (not shown), was more than twice as intense for the wt 1-344 than for the Z-variant form (integrated intensities of 11.5±2.3% and 5.2±2.2% of total peptide respectively from four separate runs on each species) (Fig 9). Non-associated fluorescein-labeled C36 was well separated and eluted much later (Fig 9). In addition, a recent study using single tryptophan variants of α₁PI demonstrated a structural difference between wild-type and Z variant α₁PIs in the vicinity of the mutation at the top of the A sheet, suggestive of greater solvent accessibility in this region (23), perhaps resulting from a widened opening at the top of the sheet.

To further characterize this incompletely-folded intermediate, we followed the folding of Z α₁PI by both native gel and activity measurements. Wild type α₁PI folded rapidly and gave a single band on native PAGE within the first time point (5 min) (Fig 10A). Separate kinetic measurements of the attainment of activity as an elastase inhibitor, and thus of correct folding, were consistent with this (Fig 10B). In contrast, Z α₁PI was much slower to achieve maximum activity (Fig 10B), which correlated with the initial formation of a slower-moving band on native PAGE, which slowly converted to a faster-moving, presumably native species (Fig 10A). Fitting of the kinetic data to a simple exponential decay process, gave half-lives of 0.7 min for wild-type and 4.2 min for the Z variant. Taken together with our above finding of the poorer association of C36 with the 1-344 containing the Z mutation, this suggests that the Z-variant intermediate contains an incompletely-inserted s5A, since we have shown above that prior association of s5A is necessary before C36 can efficiently associate. This is exactly as predicted from our folding pathway, both in terms of the accumulation of a not yet active intermediate (s1C/s4B/s5B not associated) and the final attainment of similar folding efficiency as wild-type, rather than the accumulation of inactive latent conformation (s4A cannot associate until s1C/s4B/s5B has done so). In addition, it explains an otherwise odd finding in a recent study on the accessibility of α₁PI residues in an unfolding intermediate that may correspond to the present folding intermediate. There, residue 381, otherwise buried and at the end of s5B, became fully accessible in the unfolding intermediate (24).

It has been shown that formation of polymers of the Z variant occur from a retarded normal folding intermediate (21). Our above findings of a rapidly-formed, but
not yet active intermediate for the Z variant is consistent with it being the same polymerogenic intermediate. In light of our proposed folding pathway such an intermediate is expected to have the structure depicted in Fig 8ii, though with the very C-terminal residue(s) not yet inserted as a result of the Z mutation (and hence accounting for its longer half life). The implication with regard to the nature of Z-polymers formed in the hepatocyte during folding is that they are initiated by association of the exposed s1C/s4B/s5B of one monomer with the incomplete β-sheet B of another. Significantly this is very similar to what has recently been found in the x-ray structure of a heat-induced trimeric form of α1PI (14).

ACKNOWLEDGEMENT
We thank Steven Olson and Miljan Simonovic for encouragement and many helpful suggestions and comments.
LITERATURE CITED

1. Mottonen, J., Strand, A., Symersky, J., Sweet, R. M., Danley, D. E., Georghegan, K. F., Gerard, R. D., and Goldsmith, E. J. (1992) Nature (London) 355, 270-273
2. Creighton, T. E. (1992) Nature (London) 356, 194-195
3. Lawrence, D. A., Olson, S. T., Palaniappan, S., and Ginsburg, D. (1994) Biochemistry 33, 3643-3648
4. Lomas, D. A., Elliott, P. R., Chang, W.-S. W., Wardell, M. R., and Carrell, R. W. (1995) J. Biol. Chem. 270, 5282-5288
5. Stratikos, E., and Gettins, P. G. W. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4808-4813
6. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Nature (London) 407, 923-926
7. Dementiev, A., Dobo, J., and Gettins, P. G. W. (2006) J. Biol. Chem. 281, 3452-3457
8. Kaslik, G., Kardos, J., Szabó, L., Závodszky, P., Westler, W. M., Markley, J. L., and Gráf, L. (1997) Biochemistry 36, 5455-5474
9. Gettins, P. G. W. (2002) FEBS Lett. 523, 2-6
10. Mahadeva, R., and Lomas, D. A. (1998) Thorax 53, 501-505
11. Lee, K. N., Park, S. D., and Yu, M. H. (1996) Nat. Struct. Biol. 3, 497-500
12. Peterson, F. C., Gordon, N. C., and Gettins, P. G. W. (2000) Biochemistry 39, 11884-11892
13. Yamasaki, M., Li, W., Johnson, D. J. D., and Huntington, J. A. (2008) Nature 455, 1255-1259
14. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C., and Huntington, J. A. (2011) EMBO reports 12, 1011-1107
15. Stein, P. E., and Carrell, R. W. (1995) Nat. Struct. Biol. 2, 96-113
16. Elliott, P. R., Abrahams, J. P., and Lomas, D. A. (1998) J. Mol. Biol. 275, 419-425
17. Dementiev, A., Simonovic, M., Volz, K., and Gettins, P. G. W. (2003) J. Biol. Chem. 278, 37881-37887
18. Shore, J. D., Day, D. E., Francis-Chmura, A. M., Verhamme, I., Kvassman, J., Lawrence, D. A., and Ginsburg, D. (1995) J. Biol. Chem. 270, 5395-5398
19. Onda, M., Nakatani, K., Takahara, S., Nishiyama, M., Takahashi, N., and Hirose, M. (2008) J. Biol. Chem. 283, 17568-17578
20. Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000) Genome Res. 10, 1845-1864
21. Yu, M.-H., Lee, K. N., and Kim, J. (1995) Nat. Struct. Biol. 2, 363-367
22. Lomas, D. A., Evans, D. L., Finch, J. T., and Carrell, R. W. (1992) Nature (London) 357, 605-607
23. Knaupp, A. S., and Bottomley, S. P. (2011) J. Mol. Biol. 413, 888-898
24. Krishnan, B., and Giersch, L. M. (2011) Nature Structural and Molecular Biology 18, 222-226
25. Ogushi, F., Fells, G. A., Hubbard, R. C., Straus, S. D., and Crystal, R. G. (1987) J. Clin. Invest. 80, 1366-1374
Footnotes: 1. Abbreviations used: 
\( \alpha_1 \text{PI} \), \( \alpha_1 \)-proteinase inhibitor or \( \alpha_1 \)-antitrypsin; HNE, human neutrophil elastase; PPE, porcine pancreatic elastase; RCL, reactive center loop; SI, stoichiometry of inhibition (i.e. number of mols of serpin required to inactivate 1 mol of proteinase; Z \( \alpha_1 \text{PI} \), Z variant of \( \alpha_1 \text{PI} \) with mutation of 342 from Glu to Lys; C26, peptide comprising residues 369-394 of \( \alpha_1 \text{PI} \) with an additional C-terminal cysteine; C36, peptide comprising residues 359-394 of \( \alpha_1 \text{PI} \) with an additional C-terminal cysteine and N-terminal His tag; s4A, s1C, etc, strand 4 of \( \beta \)-sheet A or strand 1 of \( \beta \)-sheet C respectively.
Table 1  Inhibition parameters for $\alpha_1$PI species

| Species                  | SI (HNE) | SI (PPE) | $k_{app}$ PPE | $k_{app} \times SI$ |
|--------------------------|----------|----------|---------------|---------------------|
| wt $\alpha_1$PI          | 1.0      | 1.7      | $1.5 \times 10^5$ | $2.7 \times 10^5$   |
| 1-323:324-394 disulfide linked | 1.1      | 1.9      | $1.6 \times 10^5$ | $3.0 \times 10^5$   |
| 1-323:324-394 two-chain IAA | 1.5      | 2.4      | $1.3 \times 10^5$ | $3.0 \times 10^5$   |
FIGURE LEGENDS

Figure 1  Serpin conformations. Native $\alpha_1$PI (panel A) and latent $\alpha_1$PI (panel B), showing the key residues and secondary structural elements mentioned in the text. The color scheme for s5A (green), s4A(RCL) (yellow), the remainder of sheet A (red), and s1C/s4B/s5B (cyan) is maintained in the folding pathway shown in Fig 8.

Figure 2  SDS-PAGE of $\alpha_1$PI under non-reducing (panel A) and reducing (panel B) conditions, with (+) or without (-) HNE. WT, wild-type $\alpha_1$PI; Cov, $\alpha_1$PI with disulfide linkage between separate chains of 1-323 and 324-394; TC, two-chain $\alpha_1$PI formed from separate, non disulfide-linked, IAA-blocked, 1-323 and 324-394 chains. HNE$^+$ represents HNE covalently-linked to 324-358 present in covalent complex of HNE with $\alpha_1$PI. M, molecular mass markers.

Figure 3  Evidence for folding of two-chain $\alpha_1$PI. CD spectra of wt $\alpha_1$PI (black), two-chain $\alpha_1$PI formed from 1-323 + 324-394 (orange), disulfide-linked "two-chain" $\alpha_1$PI formed from 1-323 + 324-394 (plum) and fragment 1-323 (cyan).

Figure 4  NMR evidence for correct folding of two-chain $\alpha_1$PI. 2D $^1$H-$^{15}$N HSQC NMR spectra of uniformly labeled wild-type $\alpha_1$PI (panel A); overlay of spectra of 1-323 + 324-394 two-chain $\alpha_1$PI, with label in the 1-323 fragment (black) and of 1-323 + 324-394 two-chain $\alpha_1$PI, with label in the 324-394 fragment (red) (panel B); uniformly-labeled fragment 1-323 (panel C).

Figure 5  Association of fluorescein-labeled C36 with 1-344 detected by fluorescence. The gel is overexposed to show the minimal association of fluorescein-labeled C26 with 1-344 or of 1-323 with either C26 or C36.

Figure 6  CD evidence that s4A inserts only after s1C/s4B/s5B. Two-chain $\alpha_1$PI formed from 1-354 + 359-394 (red) has a spectrum that resembles that of cleaved $\alpha_1$PI (green) rather than of native $\alpha_1$PI (black). In contrast, 1-354 (blue) alone gives a spectrum similar to that of 1-323 + 359-394 (cyan) which neither associate nor contain s4A. The spectrum of 359-394 (purple) adds little to the overall spectrum.

Figure 7  Fluorescence evidence that s4A only associates after s1C/s4B/s5B. Stern-Volmer plot of KI quenching of fluorescein attached to position 350 in the RCL. Native $\alpha_1$PI (black), 1-354 (blue), cleaved $\alpha_1$PI (red), 1-354 after addition of 359-394 (green).

Figure 8.  Folding pathway of $\alpha_1$PI: (i) Molten globule-like conformation of 1-323 forms first; (ii) incorporation of s5A (327-342) into $\beta$-sheet A; (iii) full insertion of s5A (green) now allows incorporation of s4B and s5B (cyan) behind s5A and of s1C (cyan). This leaves the RCL (s4A, yellow) exposed, since the RCL can only insert into $\beta$-sheet A after s4B, s5B and s1C have associated. RCL can insert into $\beta$-sheet A either after extraction of s1C or RCL cleavage, to give either latent (iv) or
cleaved structures (v) respectively. The requirement that s4B and s5B must associate before s4A can associate with sheet A, ensures that state (ii) cannot proceed directly to the latent state (iv). Color scheme: 1-323 (gray), s5A (324-344) green, RCL (s5A) yellow and s1C/s4B/s5B (359-394) (cyan).

Figure 9 Ability of fluorescein-labeled C36 peptide to associate with wt 1-344 or Z-variant 1-344. FPLC Superdex 75 elution profiles, monitored by fluorescein fluorescence, of fluorescein-labeled C36 incubated for 5 min on its own (A), with wt 1-344 peptide (B) or with Z-mutation-containing 1-344 (C). The elution position of folded α1PI, run separately, is indicated by an arrow. The slightly earlier elution of the non-covalent complex compared with native α1PI is likely due to the extra His-tag and linker in the complex.

Figure 10 Kinetics of folding of wt and Z variant α1PI followed by native PAGE (panel A) and activity measurements against HNE (panel B). Sufficient α1PI was used in panel B to give complete inhibition of the HNE if refolding were 100% successful and the SI in each case were ~1. For both the folding efficiency was ~75%, since for Z α1PI, the SI is closer to 2 (25).
Figure 3

Figure 4
Figure 7

The graph shows a set of lines with different slopes, each labeled with the corresponding $K_Q$ value:

- $K_Q = 7.1$
- $K_Q = 6.0$
- $K_Q = 4.2$
- $K_Q = 4.0$

The graph plots $F_0/F$ on the y-axis against [KI] (M) on the x-axis.
