Supplementary Information

Dynamic local unfolding in the serpin alpha-1 antitrypsin provides a mechanism for loop insertion and polymerization

Beena Krishnan[1] and Lila M. Gierasch[1,2]*

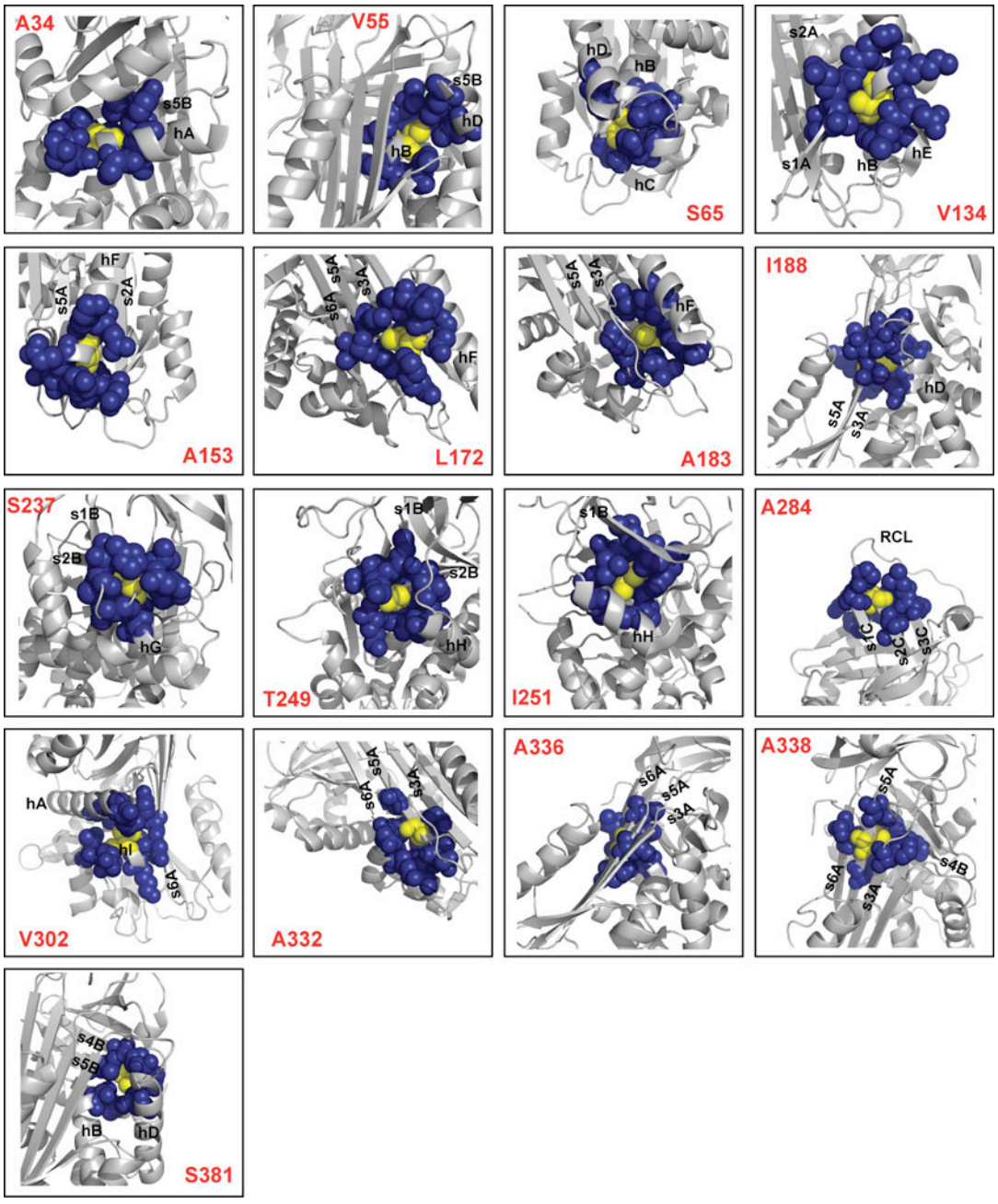
[1]Department of Biochemistry & Molecular Biology and [2]Department of Chemistry

University of Massachusetts-Amherst, Amherst, MA 01003

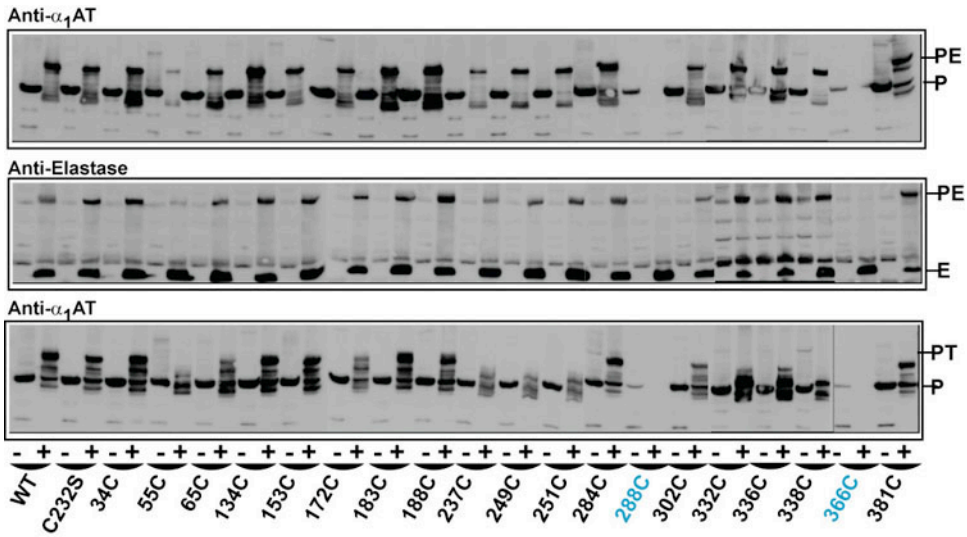
*** Correspondence should be addressed to L.M.G (gierasch@biochem.umass.edu)**

Supplementary Figures 1-6

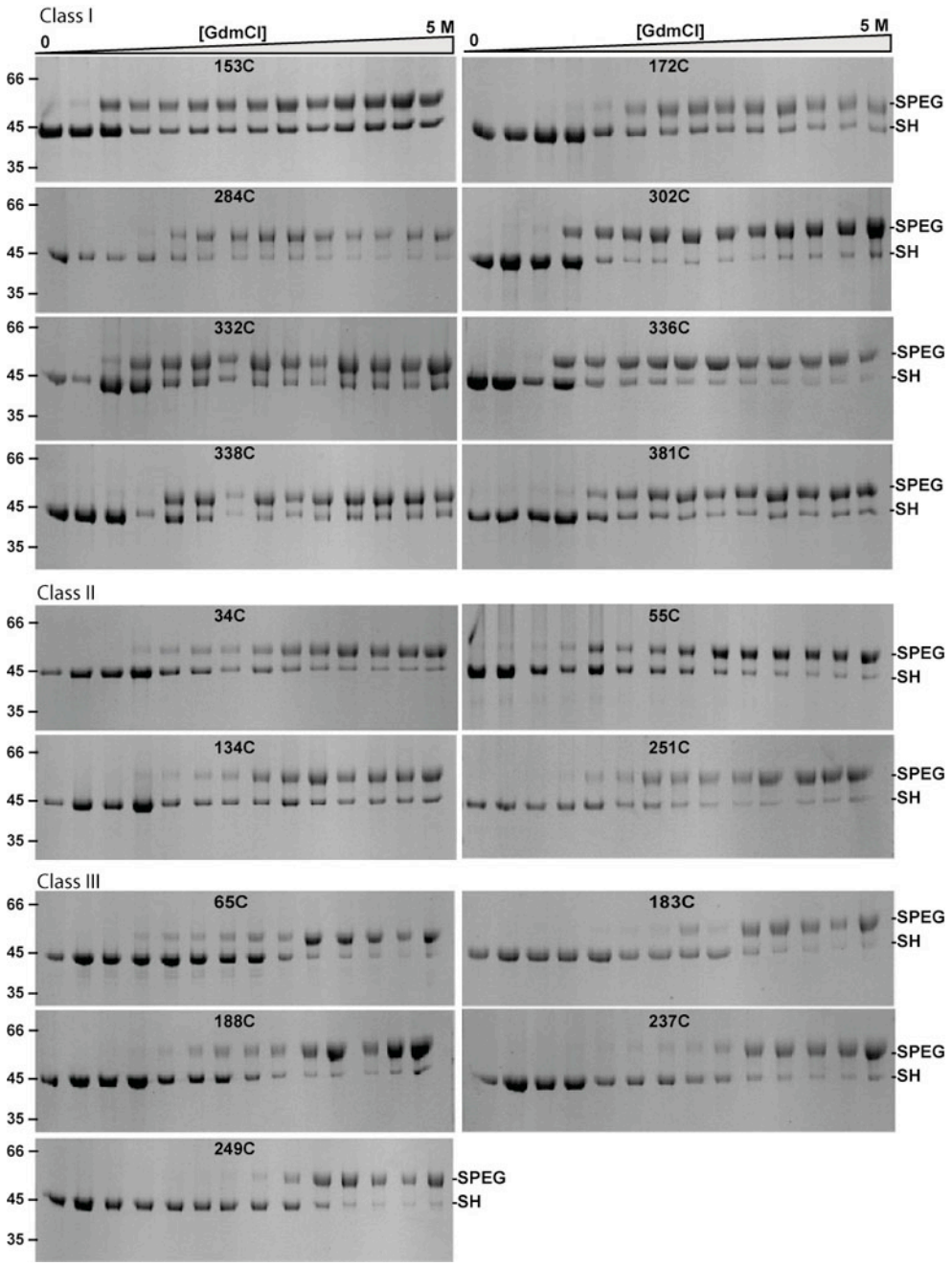
Supplementary Tables 1-3



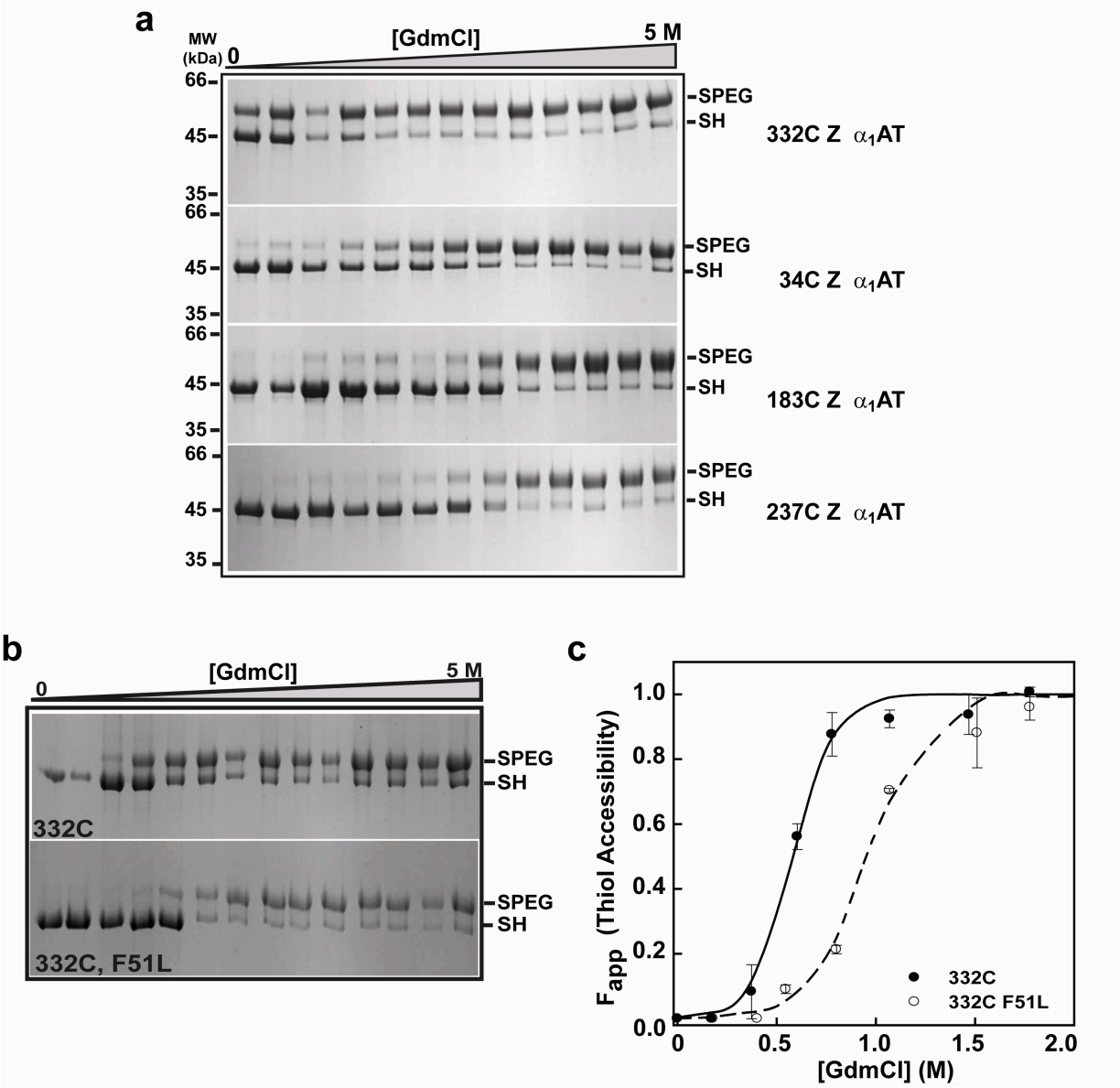
Supplementary Figure 1. *Residue packing around the positions substituted by cysteine.* Neighboring residues within 1 Å of the introduced cysteine residue position (in yellow spheres) are shown (in blue spheres) on a cartoon diagram of native alpha-1 antitrypsin (α_1AT) (PDB ID 1QLP).



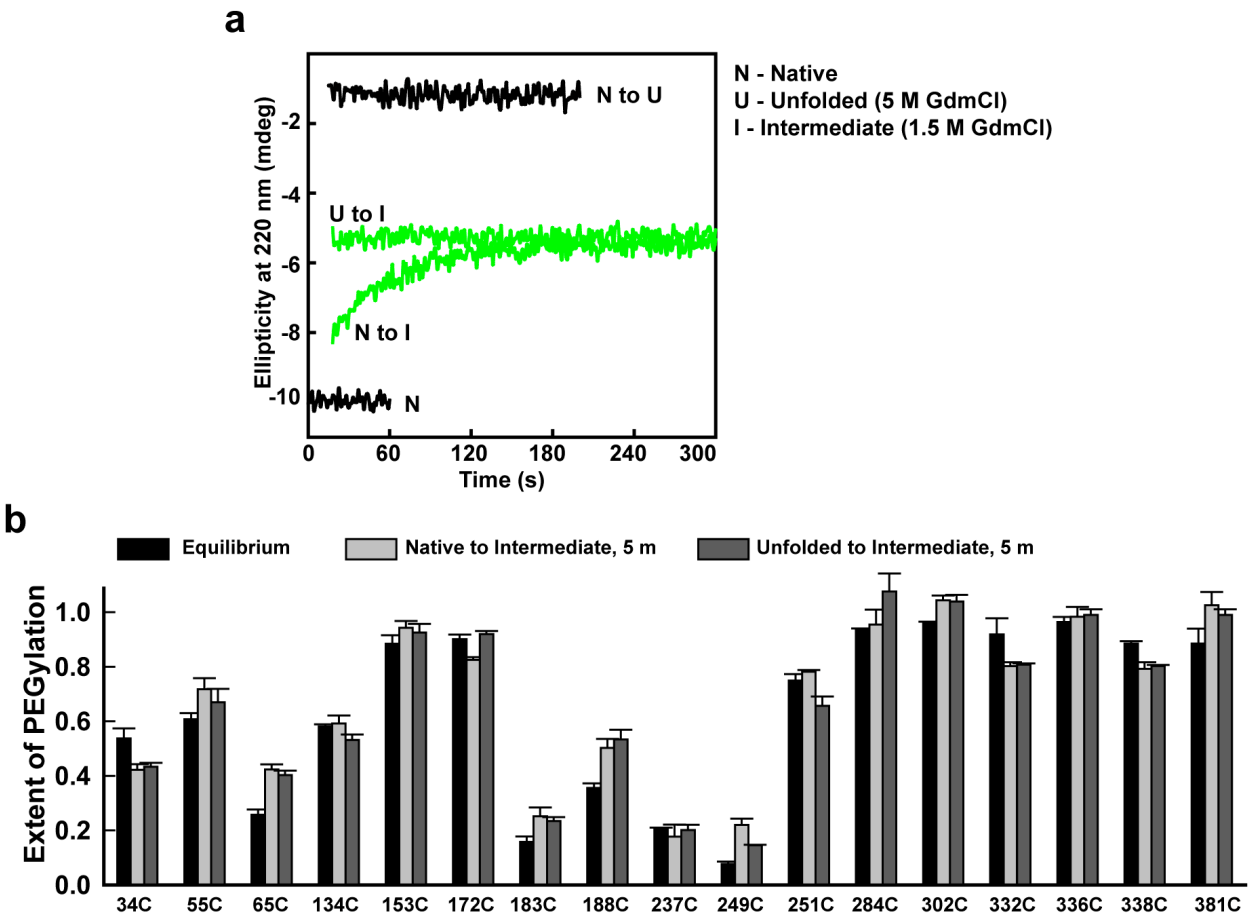
Supplementary Figure 2. *Activity screen for the cysteine variants present in the cell lysate.* Ten μL of the soluble cell lysate obtained from a small scale protein expression (in 20 ml LB), was diluted 3-fold into 50 mM HEPES, pH 8.0 containing 0.1 M sodium chloride and incubated at 37 °C for 5 m before adding 3.5 μg of porcine pancreatic elastase (Sigma) or 4.0 μg of bovine pancreatic trypsin (Sigma). The reaction was allowed to proceed at 37 °C for 15 m with shaking at 350 rpm using the Eppendorf Thermomixer, then quenched using the SDS-PAGE gel loading buffer and immediately boiled for 5 m. The samples were analyzed on a 10% Tricine SDS-PAGE and the covalent complex (if any) was detected by Western blot analysis. A Syngene G:Box gel documentation system was used for the chemiluminescence imaging of the Western blot. The labels P, PE, E, PT, and T indicate the band position of active inhibitor protein, inhibitor-elastase complex, elastase, inhibitor-trypsin complex, and trypsin, respectively. The signs '-' and '+' below the lanes indicate whether the reactions were run in the absence or presence of protease. Cysteine substitutions at positions 288 (on strand 2C) and 366 (on strand 1C), labeled in blue, resulted in inactive protein, as indicated by the absence of covalent protease-protein complex in the cell lysate.



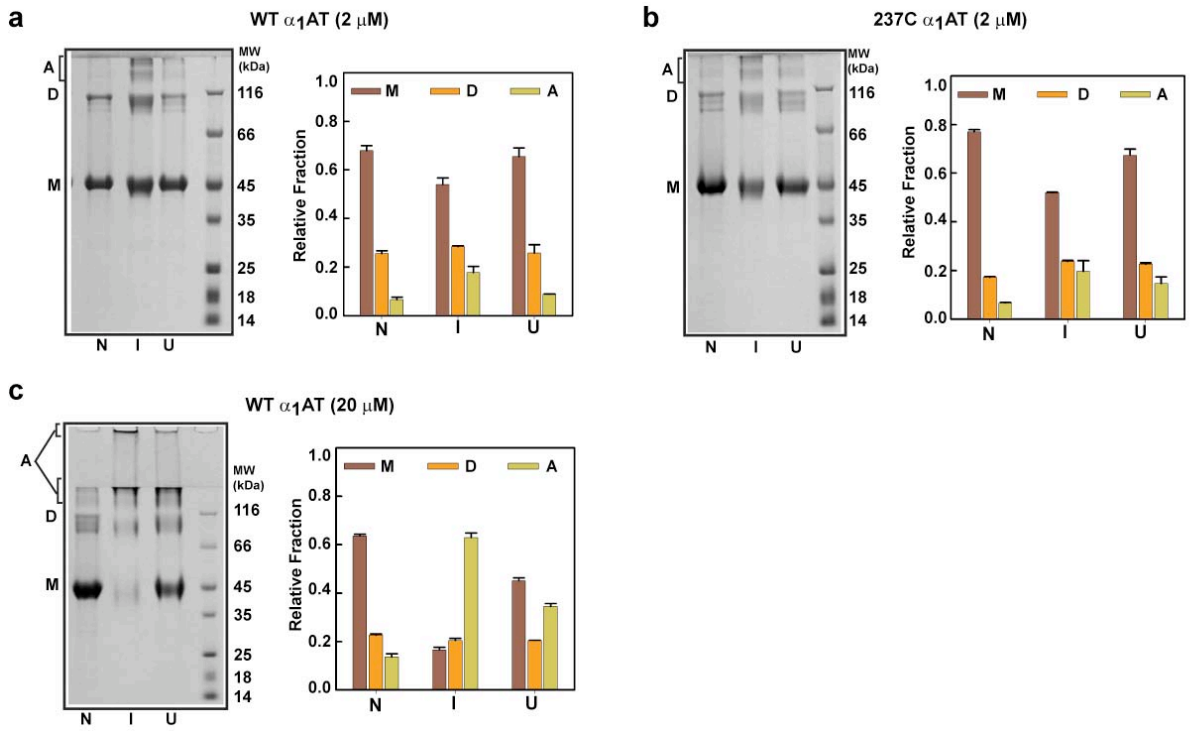
Supplementary Figure 3. *PEGylation of single cysteine α_1AT mutants.* Representative Coomassie-stained SDS-PAGE results for PEGylation of the individual cysteine variants as a function of GdmCl. Bands corresponding to the protein with free thiol and the PEG-modified thiol are indicated by '–SH' and '–SPEG' respectively. Positions of molecular weight markers (in kDa) are indicated on the left side of the gel.



Supplementary Figure 4. *PEGylation of single cysteine(s) in the disease causing Z-variant and A-sheet stabilizing F51L α_1AT.* **(a)** SDS-PAGE analysis of PEGylation of protein equilibrated in varying [GdmCl]. The intermediate state of the single cysteine Z-protein was formed from the denatured state, *i.e.,* upon dilution from high (5M GdmCl) denaturant. SDS-PAGE of PEGylation of 332C variants as a function of GdmCl **(b)** and the fractional thiol accessibility obtained from band intensity analysis of data (c) clearly show the stabilizing effect of the F51L mutation on the A-sheet. The lines in **(c)** represent the fit to a two-state protein unfolding model.



Supplementary Figure 5. *Comparison of PEGylation of cysteine variants in the 1.5 M GdmCl-induced intermediate state under equilibrium and kinetically stable conditions.* **(a)** The second transition of unfolding α_1AT is reversible, and the formation of the intermediate in 1.5 M GdmCl from unfolding native protein or refolding 5.0 M GdmCl-denatured protein is complete within 5 m, as indicated by the data for WT α_1AT. **(b)** The pattern of extent of PEGylation of cysteines in various positions of α_1AT in the intermediate formed under equilibrium conditions is in good agreement with the intermediate generated under a kinetically controlled reaction. The intermediate formed in 1.5 M GdmCl in the short time of 5 m is aggregate-free, if any aggregation were to even occur at all under these reaction conditions. These data suggest that the solvent accessibility of the residues probed by PEGylation of the protein under equilibrium conditions reports on the structure of the monomeric intermediate species.



Supplementary Figure 6. *The α_1AT intermediate is predominantly monomeric at low protein concentrations.* The extent of oligomerization in 2 μM equilibrated samples of WT α_1AT **(a)** and a single cysteine variant (237C α_1AT) **(b)** at pH 7.0 in the native (N), intermediate (I, 1.5 M GdmCl) and unfolded (U, 4.0 or 5.0 M GdmCl) states was evaluated using glutaraldehyde cross-linking. The cross-linking reaction was carried out for a minute using a 500-fold molar excess of glutaradehyde (Aldrich, Grade I) at 25 °C, followed by quenching with 0.2 M Tris, pH 8.0 for 5 m. The reaction mixture (0.1 ml) was diluted to one ml with 50 mM Tris, pH 8.0 prior to TCA precipitation. TCA precipitated, cross-linked samples were analyzed on a 10% tricine SDS-PAGE (left panels), and the band intensities corresponding to the monomer (M), dimer (D), and higher molecular weight species (combined and referred to as A) were used to estimate the relative fraction of each of the species (bar graphs). As can be noted from the figure, the relative fraction of dimeric protein remains constant in different [GdmCl] conditions, and likely arises from an experimental artifact introduced during TCA precipitation, since the native protein that has not been subjected to the TCA precipitation step is >95% monomeric (data not shown). The requirement that no reductant be added during the cross-linking reaction because of interference with glutaraldehyde cross-linking, may lead to some non-specific disulfide-mediated dimer formation. A lower end estimate of the monomeric protein is about 80%. The presence of aggregates as a minor species cannot account for the observed trends of side chain accessibility for the various cysteine positions observed using the PEGylation assay. A control experiment **(c)** indicates that aggregation is strongly enhanced at 10-fold higher protein concentration (20 μM).

Supplementary Table 1. Structural details of single cysteine α_1AT variants

Variant[a]	Location[b]	Neighboring Residues[c]	SI[d]
A34C	hA	30-32 (hA), 36-38 (hA), 52(hA-loop-s6B), 54 (s6B-loop-hB), 382(s5B), 385(s5B)	1.1 ± 0.02
V55C	hB	53(hB), 57-59(hB), 99(hD), 103(hD), 381-383(s5B)	1.2 ± 0.05
S65C	hB	61(hB), 63(hB), 67(hB-loop- hC), 68 (hB-loop-hC), 73(hC), 76(hC), 77(hC), 93(hD)	1.0 ± 0.03
V134C	hE	63(hB), 118(s2A), 130(hE), 132(hE), 136(hE), 138-142(hE-loop-s1A)	0.8 ± 0.04
A153C	hF	146(s1A-loop-hF), 147(s1A-loop-hF), 149-151(hF), 155-157(hF), 176(hF-loop-s3A)	1.4 ± 0.08
L172C	hF-s3A loop	158(hF), 161(hF), 169, 170, 174-(hF-loop-s3A) 185(s3A), 296(s6A), 333(s5A), 335(s5A)	1.0 ± 0.0
A183C	s3A	117-119(s2A), 147(s1A-loop-hF), 157(hF), 180-185(hF-loop-s3A) and s3A, 329(hI-loop-s5A), 331(s5A)	1.0 ± 0.01
I188C	s3A	53(hB), 56(hB), 112-114(s2A), 190(s3A), 334-337(s5A), 384(s5B)	1.1 ± 0.05
S237C	s2B	231(s1B), 232(s1B), 235(s1B-loop-s2B), 239(s2B), 253-256(s3B), 263(hG)	1.0 ± 0.03
T249C	s3B	241-244(s2B), 251(s3B), 276(hH), 373-375(s4B), 377(s4B-loop-s5B)	1.1 ± 0.02
I251C	s3B	239-241(S2B), 249(s3B), 253(s3B), 272(hH), 276(hH), 372(s4B), 373(s4B), 375(s4B)	1.5 ± 0.17
A284C	s2C	223-225(s3C), 227(s1B), 229(s1B), 282-286(s2C), 361(RCL), 362(RCL)	1.0 ± 0.06
V302C	hI	37(hA), 41(hA), 44(hA), 297-300(s6A and hI), 304-306(hI)	1.2 ± 0.08
A332C	s5A	61(hB), 184(s3A), 298-300(hI), 312(hI-loop-s5A), 329(hI-loop-s5A), 330(hI-loop-s5A), 334(s5A)	1.1 ± 0.04
A336C	s5A	51(s6B), 52(s6B), 188(s3A), 294(s6A), 295(s6A), 334-338(S5A)	0.9 ± 0.04
L338C	s5A	51(s6B), 190(s3A), 292(s6A), 293(s6A), 336(s5A), 340(s5A), 372(s4B), 374(s4B)	1.2 ± 0.07
S381C	s5B	30(hA), 55(hB), 99(hD), 375(s4B), 376(s4B), 379(s4B-loop-s5B), 383(s5B)	1.0 ± 0.05

a. Mutations were introduced into cysteine-free (C232S) α_1AT.

b. The location of the residue with respect to the main secondary structural elements of α_1AT: helices (h), strands (s) in sheets A, B, or C.

c. Residues that are within 1 Å of the variant position and the structural element to which the contacting residue belongs.

d. The mean value of stoichiometry of inhibition (SI) determined against trypsin as described [Liu, L., et al., *Biochemistry* **45**, 10865-10872 (2006)]. The reported error in SI is the observed standard error from three independent measurements.

Supplementary Table 2. Equilibrium stabilities of single cysteine α_1AT mutants

Variant	$\Delta G°_{IN}$ [a] (kcal/mol)	$\Delta\Delta G°_{IN}$ [b] (kcal/mol)	$\Delta G°_{UI}$ [a] (kcal/mol)	$\Delta\Delta G°_{UI}$ [b] (kcal/mol)
WT	4.0	-	4.0	-
A34C	3.9	-0.1	3.8	-0.2
V55C	3.7	-0.3	3.9	-0.1
S65C	3.5	-0.4	3.9	-0.1
V134C	4.4	0.5	3.8	-0.2
A153C	3.4	-0.6	3.9	-0.1
L172C	5.5	1.5	3.9	-0.1
A183C	5.5	1.6	4.0	-0.1
I188C	5.0	1.0	3.8	-0.2
S237C	4.3	0.3	4.3	0.3
T249C	4.1	-0.1	4.4	0.4
I251C	3.8	-0.2	3.6	-0.4
A284C	4.8	0.8	3.9	-0.1
V302C	4.5	0.5	4.0	0.0
A332C	3.9	-0.1	3.8	-0.2
A336C	3.5	-0.4	4.1	0.1
L338C	4.4	0.5	4.2	0.2
S381C	4.5	0.5	4.1	0.1

a. Obtained by fitting the GdmCl unfolding of cysteine mutants monitored by CD (Fig. 2b). The reported values for the cysteine mutants derive from fits with fixed m_{IN} (-5.67 kcal/mol/M) and m_{UI} (-1.52 kcal/mol/M).

b. Estimated from the free energy difference ($\Delta\Delta G°$) between the cysteine mutant and the wild type protein for the native to intermediate (IN) and intermediate to unfolded (UI) transitions.

Supplementary Table 3. Comparison of α_1AT denaturation monitored by CD and site accessibility monitored by PEGylation

Variant	Equilibrium Stability		Thiol Accessibility by PEGylation		
	$C_{m\,IN}$[a] (M)	$C_{m\,UI}$[a] (M)	$C_{m\,IN}$[b] (M)	$C'_{m\,UI}$[b] (M)	$C_{m\,UI}$[b] (M)
A34C	0.7	2.5	0.7	1.7	-
V55C	0.7	2.5	0.5	1.8	-
S65C	0.6	2.6	0.6	-	2.3
V134C	0.8	2.5	0.7	1.6	-
A153C	0.6	2.6	0.4	-	-
L172C	1.0	2.6	0.9	-	-
A183C	1.0	2.6	0.9	-	2.5
I188C	0.9	2.5	0.8	-	2.1
S237C	0.8	2.9	-	-	2.2
T249C	0.7	2.9	-	-	2.4
I251C	0.7	2.4	-	1.2	-
A284C	0.7	2.6	0.8	-	-
V302C	0.8	2.6	0.7	-	-
A332C	0.7	2.5	0.6	-	-
A336C	0.6	2.7	0.6	-	-
L338C	0.8	2.7	0.7	-	-
S381C	0.8	2.8	0.8	-	-

a. GdmCl unfolding of cysteine mutants monitored by CD (data in Fig. 2b).
b. Midpoints for transitions in fits of extent of cysteine modification by PEGylation for the various mutants (Fig. 2b). The values reported here were obtained from fitting the PEGylation data with fixed m-values of m_{IN} and m_{UI} of -5.67 kcal/mol/M and -1.52 kcal/mol/M, respectively, obtained from the CD unfolding measurements.