Supporting Information for

Why does a structurally complex protein fold to its thermodynamically less stable active state?

V. V. Hemanth Giri Rao[1], Shachi Gosavi[1]

[1]Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, TIFR, Bellary Road, Bangalore, India, 560065.

Corresponding author: Shachi Gosavi, Tel: +91-80-23666105, Fax: +91-80-23636662, Email: shachi@ncbs.res.in.

Keywords: α1-antitrypsin, Reactive center loop, Protein folding simulations, Kinetic trap, Alternative folded conformations.

SI Text

Domain definitions of active and latent conformations of α1-AT (Fig. 2 and Fig. S1).

According to the CATH database(1), both conformations have two domains each. Each domain is made up of two discontinuous segments of the protein. In the active conformation, domain 1 is made up of residues 195 to 289 and residues 344 to 394, and the rest of the protein constitutes domain 2. In case of the latent conformation, domain 1 is made up of residues 195 to 289 and residues 359 to 394, while the rest of the protein constitutes domain 2. Note that the residue numbering here is according to the sequence of WT α1-AT which has 394 residues(2, 3). The domain definitions according to the numbering in our SBMs can be obtained by subtracting 22 (see Methods section in the main text) from all numbers.

SI Figures

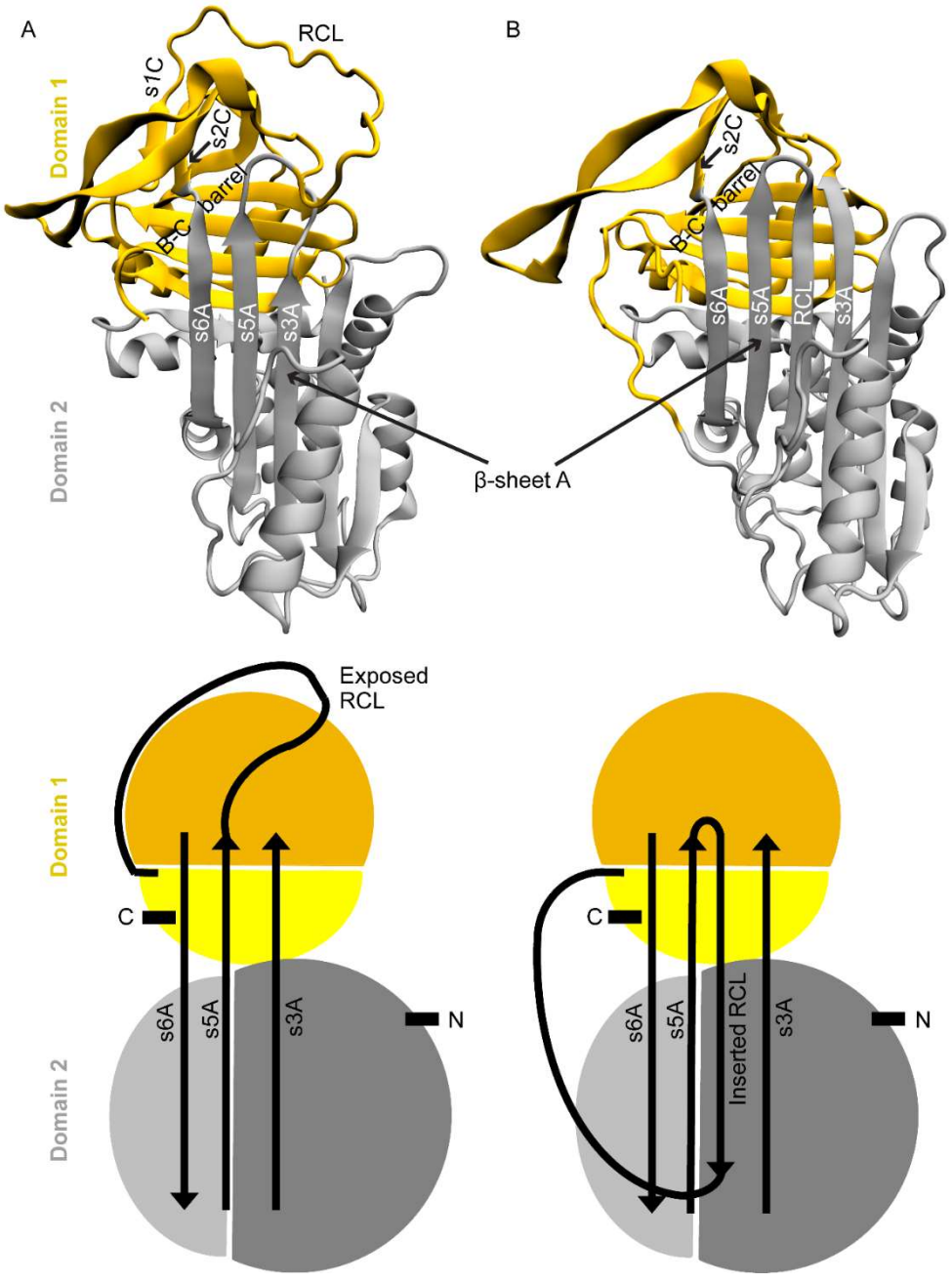

Figure S1 Domain definitions of the active and latent conformations of α1-AT. The structures of the (A, top) active and (B, top) latent conformations of α1-AT colored according to their domains are reproduced from Fig. 2. Simplified cartoon representations of the (A, bottom) active and (B, bottom) latent conformations are shown. Domains 1 and 2 are made up of two sets of polypeptide segments. These are colored using two different shades of yellow for domain 1, and two different shades of grey for domain 2. In both conformations, the N- and C-termini are in domains 2 and 1 respectively. Some of the key structural

elements such as the RCL, strands s3A, s5A and s6A are also marked. The RCL is exposed in the active conformation (A, bottom) while it is inserted between strands s5A and s3A in the latent conformation (B, bottom). Importantly, the RCL is in domain 1 in the active conformation while it is a part of domain 2 in the latent conformation.

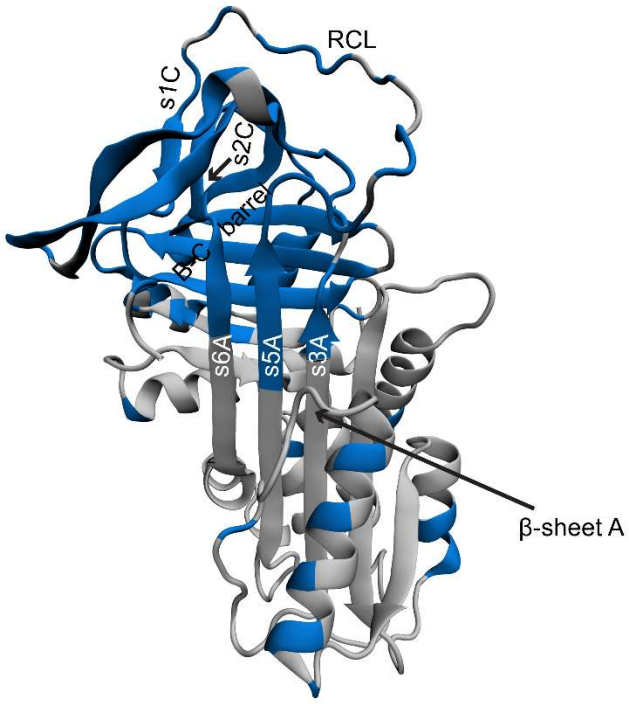

Figure S2. Average structure of the I_{active} intermediate populated in folding to the active conformation of α1-AT. The average contact map of I_{active} (Fig. 4B) was used to identify residues which are folded in this ensemble. For every residue, we obtain the average fraction of native contacts formed by considering only native contacts of that residue. When this average is greater than 0.4 i.e. 40% of the native contacts of that residue are formed on average, we consider that residue to be folded. Such folded residues are marked in blue on the structure of the active conformation while the unfolded residues are marked in grey. We find that domain 1 is folded, while domain 2 remains unfolded.

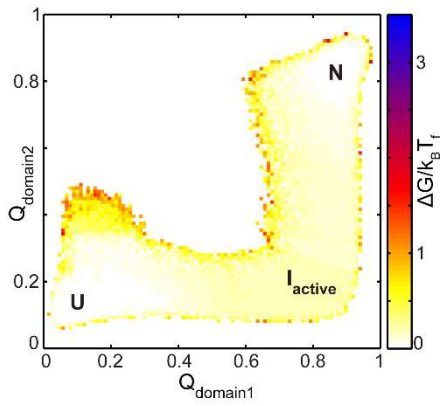

Figure S3. Errors on the 2DFES of folding to active α1-AT. The 2DFES of folding to the active conformation plotted in Fig. 4C shows the average free energy along two reaction coordinates $Q_{domain2}$ and $Q_{domain1}$. For the same 2DFES, we plot the standard deviation of the free energy obtained from four independent replicate simulations. The colors show the standard deviation of the free energy in k_BT_f. We find that the errors in the region connecting the N and U states is less than 1 k_BT_f suggesting sufficient sampling in our simulations.

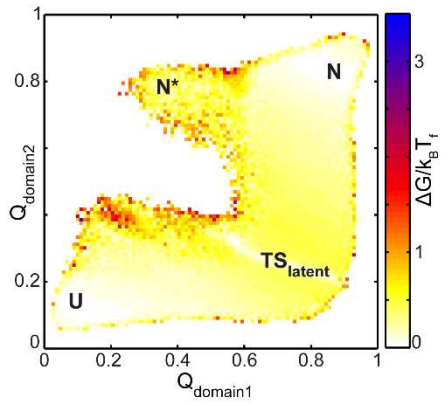

Figure S4. Errors on the 2DFES of folding to latent α1-AT. The 2DFES of folding to the latent conformation plotted in Fig. 6C shows the average free energy along two reaction coordinates $Q_{domain2}$ and $Q_{domain1}$. For the same 2DFES, we plot the standard deviation of the free energy obtained from four independent replicate simulations. The colors show the standard deviation of the free energy in k_BT_f. We find that the errors in the region connecting the N and U states is less than 1 k_BT_f suggesting sufficient sampling in our simulations.

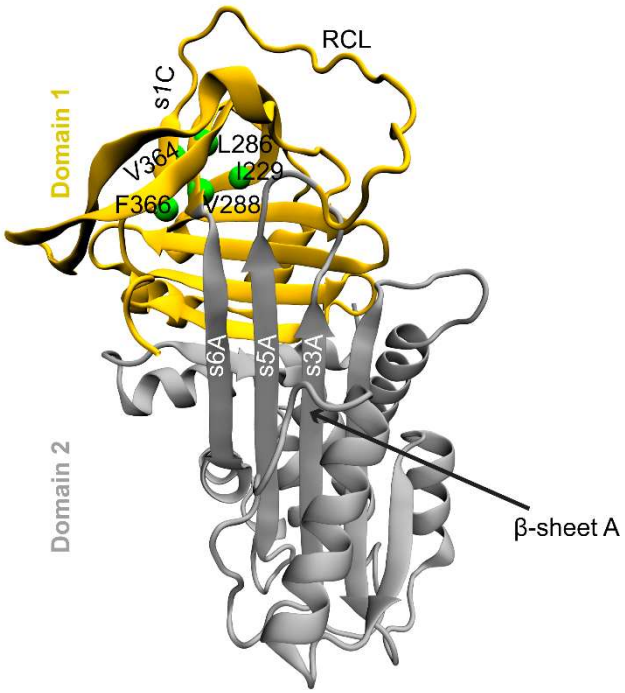

Figure S5. Mutations which transform WT α1-AT to the latent conformation are present in domain 1. C-α atoms of residues where mutations transform WT α1-AT to the latent conformation are shown on the structure of active α1-AT. These residues correspond to residue numbers 229, 286, 288, 364, and 366(4).

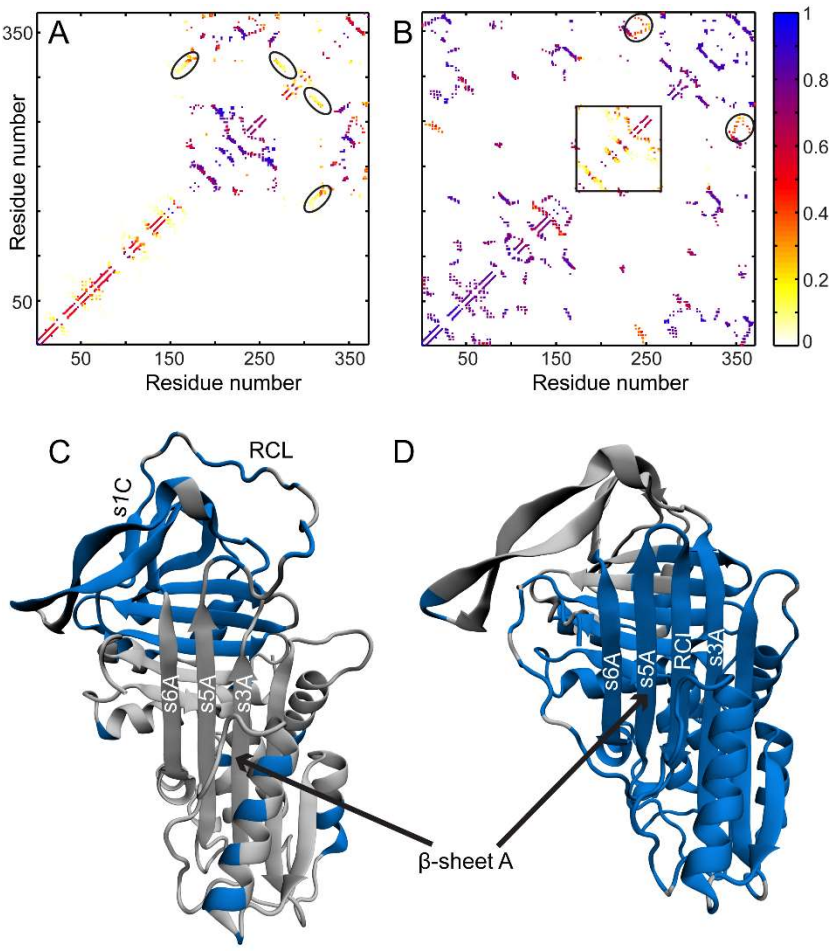

Figure S6. Structural ensembles from the folding landscapes of the active and latent conformations of α1-AT than can potentially facilitate domain swapping and polymerization. (A) The average contact map of the structural ensemble at Q~0.34 in the folding free energy profile of the active conformation (Fig. 4A). This ensemble is structurally similar to I_{active} (Fig. 4B). However, at Q~0.34, in addition to the regions of α1-AT that are unfolded in I_{active}, the regions marked in black ovals are also unfolded. The black ovals whose long axes are parallel to the diagonal correspond to β-strands s5A/s3A, and the remaining two ovals whose long axes are perpendicular to the diagonal correspond to β-strands s5A/s6A. The contact map is symmetric about the diagonal and the colors represent the probability of contact formation. The average structure of this ensemble is shown in C. (B) The average contact map of the structural ensemble marked as N* in the 2DFES of the latent conformation (see Fig. 6C). In N*, several contacts of domain 1 are unfolded (contacts marked by the large black square and the two ovals), while domain 2 is folded. The average structure of this ensemble is shown in D. (C) The average structure of the ensemble at Q~0.34 obtained from the average contact map shown in A. (D) The average structure of the N* ensemble obtained from the average contact map shown in B. In C and D, a residue is folded when 40% of its native contacts

are formed on average in that ensemble. The residues which are folded are colored in blue, while those which are unfolded are colored in grey.

SI Methods

C-α SBMs of active and latent conformations of α1-AT.

In this study, we employed C-α SBMs in which each residue is represented by a bead at its C-α atom(5). We obtained the native structures of active and latent conformations of α1-AT from PDB IDs 1QLP and 1IZ2 respectively (Fig. 1A, B). The sequence of WT α1-AT has 394 residues(2, 3). However, residues 1 to 22 are missing in PDB 1QLP, and 1 to 23 are missing in PDB 1IZ2. We renumbered α1-AT so that residue 23 is residue 1, and thus, 1QLP has residues numbered 1 to 372 and 1IZ2 has residues numbered 2 to 372. Using these structures, we obtained the native contact maps for both conformations from the Contacts of Structural Units (CSU) software(6). According to CSU, if non-hydrogen atoms belonging to different residues (i and j, such that $|i-j|>3$) have intersecting solvent accessible surfaces, the corresponding pair of C-α atoms can be considered as a native contact. The native contact maps of the active and latent conformations have 1115 and 1137 contacts respectively (shown in Fig. 3). These contact maps are listed in the Supporting Information below. The coordinates of the C-α atoms from the PDB file and its corresponding native contact map were given as inputs to the SMOG webserver(7) to generate the C-α SBM for that conformation. Using this procedure, we constructed separate C-α SBMs for the active and the latent conformations of α1-AT. The energy scale for both C-α SBMs was set to 1 kJ/mol.

Replica exchange umbrella sampling (REUS) simulations.

We performed canonical (NVT ensemble) molecular dynamics (MD) simulations of the C-α SBMs of both active and latent α1-AT to obtain their respective folding free energy landscapes at their respective folding temperatures (T_fs). The folding temperature (T_f) is the temperature at which both native and unfolded states are equally populated. Using conventional MD simulations, we did

not observe either folding or unfolding transitions at T_f in a computationally reasonable time. We resolved this issue by enhancing the sampling of protein conformations having large free energies by using replica exchange umbrella sampling (REUS)(8, 9). In REUS, the system is constrained to sample specific regions of the free energy landscape between the native and the unfolded states and the fraction of native contacts formed (Q) is a suitable reaction coordinate to capture the folding process(10). Q is obtained by suitably normalizing the number of native contacts formed by the total number of native contacts in the protein(11).

We chose 64 values of Q between 0 and 1 (termed Q_k, for k = 1, 2, ...,64), and initiated an equal number of simulations (termed as replicas) in parallel, each starting with a structure having a Q close to the corresponding Q_k (Eq. 0). The simulation temperature for all the replicas was the same, and was chosen to be a temperature close to the T_f. For each replica, an external potential (termed as an umbrella) was applied on the potential energy of the system which constrained the sampling to protein conformations having values of Q close to the corresponding value of Q_k. Here, the umbrella potential was chosen to be a harmonic function of Q with an equilibrium value at the corresponding value of Q_k, and having a force constant, K_Q(8, 9).

$$E^{k}_{Umbrella}(Q, Q_k) = \frac{1}{2} K_Q (Q - Q_k)^2 \qquad \text{Eq. 0}$$

In Eq. 0, $E^{k}_{Umbrella}(Q, Q_k)$ is the harmonic umbrella potential for the k^{th} simulation, K_Q is the strength of the umbrella bias potential in kJ/mol, and Q_k is the value of Q at which the umbrella bias potential is centred.

The choice of K_Q was such that it allows for significant overlap between structural ensembles of adjacent replicas (i.e. between replicas k and k+1, see Fig. S7)(8, 9). This overlap is a necessary criterion to obtain reliable free energy profiles using the weighted histogram analysis method(8, 9,

12) (WHAM). To also ensure that there are folding and unfolding transitions, REUS requires adjacently located replicas to exchange structures at fixed time intervals during the simulation. The acceptance probability for an attempt to exchange structures is determined by the Metropolis criterion(8, 9). Previous studies have shown that acceptance probabilities between 0.2 and 0.4 correspond to sufficient sampling and overlap between the structural ensembles of adjacent replicas(8, 9, 13).

We performed all MD simulations in GROMACS 4.5.4(14). Since, REUS along Q is not implemented in any version of GROMACS, we resolved this issue by using a modified version of the PLUMED 1.3 plugin along with GROMACS 4.5.4. The PLUMED 1.3 plugin extends the capabilities of GROMACS by implementing several reaction coordinates and enhanced sampling techniques suitable for capturing protein dynamics(15). We combined the replica exchange functionality of GROMACS, with umbrella sampling and bias exchange functionalities of PLUMED 1.3 to implement REUS. However, since the bias exchange module attempts exchanges between randomly chosen replicas by default, we modified it so that it attempts exchanges only between adjacent replicas (i.e. between replicas k and k+1) instead.

We performed four independent replicate REUS simulations for each of the two C-α SBMs of α1-AT. These replicates had different sets of starting structures and were at different temperatures close to the T_f (see Table S1). All simulations were performed using the stochastic dynamics integrator, with a time step of 0.0005 ps while saving structures every 2000 steps, and with non-bonded interactions cut-off at 3 nm. The REUS parameters were as follows. The force constant K_Q was 0.03 kJ/mol for all replicas across all replicates for both C-α SBMs, the collective variable i.e. the reaction coordinate along which the umbrella potentials were applied, was the 'Contact Map' CV (which is equivalent to the total number of native contacts formed(11)). We chose 64 equally

spaced Q_k values between and including 0.05 to 0.95 each defining a unique umbrella potential – one for each of the 64 replicas. The time interval for attempting replica exchange was once every 10000 time steps, and each replicate was simulated for 4×10^8 time steps.

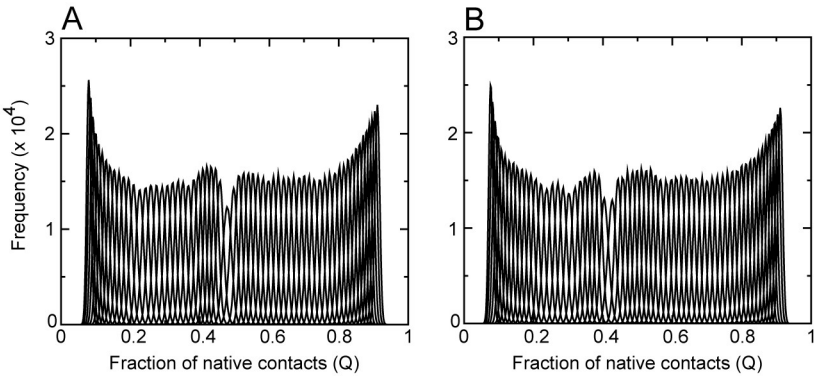

Figure S7. The overlap between adjacent histograms of Q in an REUS simulation. For each REUS replicate, we obtained 64 trajectories i.e. one from each of the 64 replicas. We computed Q for each simulation frame that was saved (2×10^5 per trajectory), and obtained the time evolution of Q for each of the 64 replicas. The value of Q was observed to fluctuate about the value of Q_k (see Eq. 0) defined for the k^{th} replica. The histograms of Q from an REUS simulation for the SBM of (A) the active conformation and (B) the latent conformation of α1-AT are shown and are observed to have sufficient overlap between adjacent replicas. This criterion is necessary to obtain convergence during WHAM analysis(8, 9, 12) (see Fig. S10).

WHAM analysis to obtain unbiased estimates of the folding free energy profiles and the folding mechanisms.

Using the simulation parameters specified in the previous section, we could obtain (i) sufficient structural overlap between adjacent replicas (Fig. S7), (ii) acceptance probabilities in the range of 0.2 to 0.4 for all pairs of adjacent replicas (Fig. S8), and (iii) at least 15 folding and unfolding transitions in each replicate REUS simulation (see Fig. S9, and Table S1), for all four independent replicates of both C-α SBMs. These three conditions ensure that the underlying landscape has been

sampled sufficiently(8, 9). Since REUS enhances the sampling of the free energy landscape by adding umbrella potentials (Eq. 0) to the system's potential energy function, the distributions of the conformational states observed from REUS simulations must be reweighted to remove the effect of the umbrella potentials in order to obtain the unbiased distribution of conformational states(8, 9). The weighted histogram analysis method or WHAM(8, 9, 12), provides a set of equations that must be self consistently solved to obtain the unbiased probabilities of the conformational states defined along the reaction coordinate Q at the temperature of the simulation. Since WHAM can also be used to unbias the effect of temperature, the unbiased distribution can be obtained at the T_f, using data from simulations that were carried out at a close enough temperature (i.e. within ± 0.01663 reduced units of the T_f, where 1 K = 0.008314 reduced units). Finally, the negative logarithm of the unbiased distribution of conformational states along Q, at the T_f, represents the folding free energy profile. An important criterion to obtain reliable folding free energy profiles is the convergence of the WHAM algorithm(12) and is shown in Fig. S10. For improved reliability, we also stipulated that the folding free energy profiles and the T_fs also converge within ± 1 k_BT_f for the free energy and ± 0.008314 reduced units for the temperature (Table S1) across the four independent replicates of each C-α SBM. We report the unbiased folding free energy profiles at T_f averaged over the four independent replicates, and the associated standard deviation as the error bar on the free energy. Since each independent replicate also had a corresponding T_f, we report the average of these as the T_f for a given C-α SBM and the associated standard deviation as the error in estimating the T_f. WHAM analysis can also be performed on multi-dimensional distributions (i.e. distribution of conformational states along multiple reaction coordinates), and we used it to obtain two-dimensional free energy surfaces (2DFESs), as well as

average contact maps along Q which shows the average probability of contact formation for all native contacts during folding or unfolding.

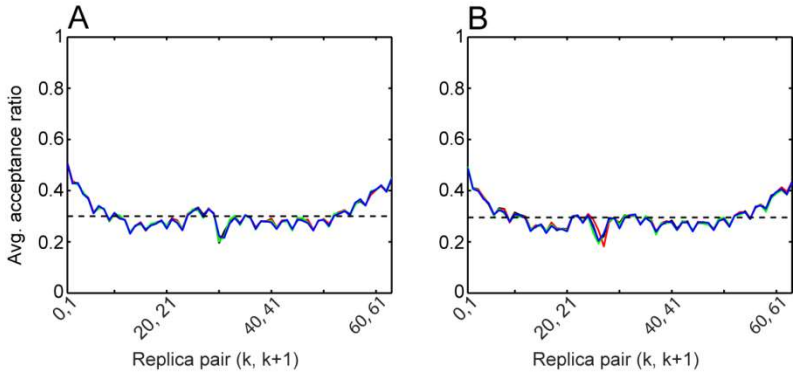

Figure S8. Average acceptance ratio for all REUS simulations. These are plotted for all the four replicates (colored black, red, blue and green) for the SBM of (A) the active conformation and (B) the latent conformation of α1-AT. The average acceptance ratio for a pair of adjacently located replicas is defined as the ratio of 'the number of exchanges that have taken place' to 'the total number of exchanges that have been attempted' during the entire simulation. A low average acceptance ratio (~0) implies that the adjacent replicas are sampling disconnected regions of the energy landscape or that the force constants of the harmonic potentials are so large that they are over constraining the sampling to a small region around the corresponding Q_ks. On the other hand, a large average acceptance ratio (~1) implies that the adjacent replicas are sampling the same region of the energy landscape. In accordance with the recommendations from previous studies(8, 9) to achieve optimum sampling, we find that the average acceptance ratio is between 0.2-0.4 and is approximately constant (~0.3, dashed line) for all pairs of replicas that are between and including the unfolded and native basins of the folding free energy profile, i.e. replica pairs (4, 5; 5, 6; ...; 55, 56). The remaining replica pairs at the left and right extremes of the graph correspond to structural ensembles of the completely unfolded and the completely folded states respectively. Therefore, their respective average acceptance ratios for exchange are higher.

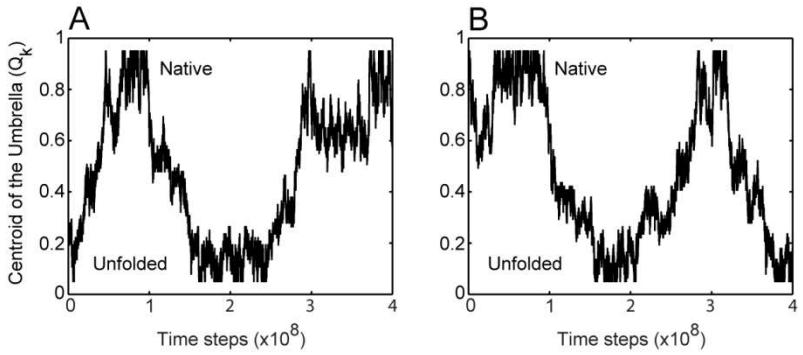

Figure S9. Replica exchange facilitates the sampling of folding and unfolding transitions of α1-AT. The initial structure of a replica is followed in replica space for one of the replicas from the REUS simulation of the SBM of (A) the active conformation and (B) the latent conformation of α1-AT. The centre of the umbrella potential (Q_k) corresponding to the replica in which the structure exchanges into is plotted versus the number of time steps. In (A) the initial structure (at time step = 0) belongs to a replica that has a Q_k of ~0.25. This structure begins by exchanging into replicas having lower values of Q_k i.e. it unfolds. Subsequently, it exchanges into replicas having higher values of Q_k and folds completely to reach the native state. We define this as a folding transition. Similarly, exchanges from replicas corresponding to higher values of Q_k into those having lower values of Q_k are termed as unfolding transitions. We find that there are two folding transitions and one unfolding transition in (A), while there are two unfolding transitions and one folding transition in (B). The total numbers of folding and unfolding transitions obtained by following all 64 initial structures through replica space as the simulations proceed are summarized in Table S1 for all the independent replicate simulations for both SBMs.

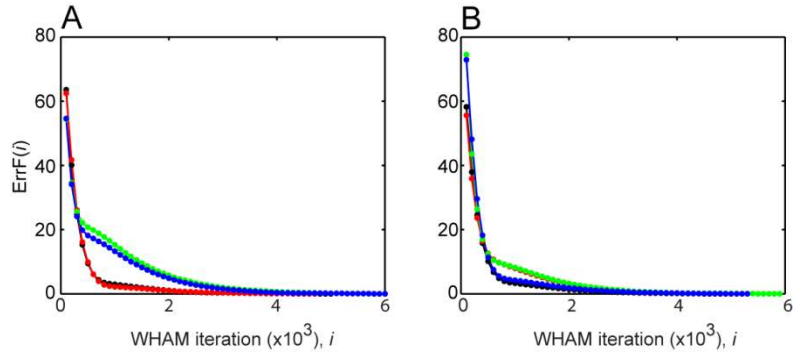

Figure S10. Convergence of the WHAM analysis to obtain the free energy profiles. The error function, errF(*i*) (see Eq. 1) measures the difference between the free energy profile obtained in the final WHAM iteration (n_{iter}) and that obtained at any other previous WHAM iteration *i*. When this difference reaches 0, the WHAM algorithm has converged. The convergence of the WHAM algorithm is shown for all four

replicates (colored black, red, blue and green) for the SBM of (A) the active conformation and (B) the latent conformation of α1-AT.

$$ErrF(i) = \sqrt{\sum_{Q}(F(Q,i) - F(Q,n_{iter}))^2} \quad \text{Eq. 1}$$

In Eq. 1, ErrF(i) is the error function, Q is the fraction of native contacts, F(Q,i) is the free energy profile (as a function of Q) obtained at the i^{th} WHAM iteration, and F(Q,n_{iter}) is the free energy profile (as a function of Q) obtained at the last WHAM iteration, n_{iter}.

Calculating the relative enthalpy and entropy changes during folding to the active conformation versus folding to the latent conformation.

For each replicate simulation, we used WHAM to obtain the unbiased two-dimensional probability distribution as a function of the potential energy change ($\Delta H_{\xi-U}$, which is the potential energy of state ξ with reference to the unfolded state) and Q. This was used to obtain the average potential energy change ($\langle\Delta H_{\xi-U}\rangle$) versus Q. Next, we obtained the average entropy change ($\langle\Delta S_{\xi-U}\rangle$) versus Q using Eq. 2.

$$T\langle\Delta S_{\xi-U}\rangle = \langle\Delta H_{\xi-U}\rangle - \langle\Delta G_{\xi-U}\rangle \quad \text{Eq. 2}$$

In Eq. 2, $\langle\Delta G_{\xi-U}\rangle$ is the average free energy change of state ξ with respect to the unfolded state, T is the temperature and, $\langle\Delta H_{\xi-U}\rangle$ and $\langle\Delta S_{\xi-U}\rangle$ are as defined above.

Subsequently, at every value of Q, we subtracted the average potential energy change obtained from a replicate of the C-α SBM of the latent conformation ($\langle\Delta H_{\xi}^{latent}{}_{-U}\rangle$, where ξ^{latent} is a conformational state in the energy landscape of folding to the latent conformation) from that obtained from a replicate of the C-α SBM of the active conformation ($\langle\Delta H_{\xi}^{active}{}_{-U}\rangle$, where ξ^{active} is a conformational state in the energy landscape of folding to the active conformation) to obtain the difference in the average potential energy change in folding to ξ^{active} versus folding to ξ^{latent}

from the unfolded state. This difference is denoted as ΔΔH(active−latent) and is a function of Q. Note that during this subtraction, the ξ^{active} and ξ^{latent} conformational states that are at the same value of Q in their respective folding landscapes, i.e. ξ^{active} and ξ^{latent} have folded to the same extent from their unfolded states. Similarly by subtracting ($<\Delta S_{\xi}^{latent}{}_{-U}>$) from ($<\Delta S_{\xi}^{active}{}_{-U}>$) at every value of Q, we obtained ΔΔS(active−latent) versus Q. Since there were four replicates of both C-α SBMs, we obtained the average ΔΔH(active−latent) and average ΔΔS(active−latent) versus Q, where each was averaged over the 16 pairings of independent replicates of active and latent C-α SBMs. The standard deviation across these 16 pairings of replicates is reported as the error on the averages obtained.

Changes made to biasexchange.c code of PLUMED v1.3 to implement Replica Exchange Umbrella Sampling (REUS).

In bias exchange implemented in PLUMED v1.3, multiple MD simulations (or replicas) with different external bias potentials (conceptually similar to the umbrella potentials) are carried out in parallel. At regular intervals, the structures between two randomly chosen replicas are exchanged using the Metropolis criterion. However, for REUS we require that such replicas be adjacent to each other in Q space, as only conformations from replicas having adjacent Q_k values will have sufficient structural overlap to satisfy the Metropolis criterion(8, 9). We therefore modified the biasexchange.c code of PLUMED v1.3, by commenting out the 'for loop' that randomizes the indices of the replicas before attempting an exchange of structures, by a line of code that retains the original indices of the replicas.

Original code of biasexchange.c	Modified code of biasexchange.c for use with REUS.

#include "metadyn.h"
 	
void bias_exchange_traj(int nrepl, int *seed, int *ind)
{
#ifdef PLUMED_GROMACS
int stat, i, j;
 	
for(i=0;i<nrepl;i++) {
stat=1;
while(stat) {
stat=0;
ind[i] = rando(seed)*nrepl;
for(j=0;j<i;j++) if(ind[i]==ind[j]) stat=1;
}
}
#endif
}
	#include "metadyn.h"
	
	void bias_exchange_traj(int nrepl, int *seed, int *ind)
	{
	#ifdef PLUMED_GROMACS
	int stat, i, j;
	
	for(i=0;i<nrepl;i++) {
	/* stat=1;
	while(stat) {
	stat=0;
	ind[i] = rando(seed)*nrepl;
	for(j=0;j<i;j++) if(ind[i]==ind[j]) stat=1;
	} */
	
	ind[i]=i;
	}
	#endif
	}

SI Results

Comparison of experimental folding times with refolding free energy barriers in folding to the active conformation.

Experiments find that under strongly refolding conditions (0.6 M GdnHCl), folding to the active conformation of α1-AT has time constants of τ_{fast} ~1000 sec and τ_{slow} ~1500 sec (2). The three state fit to the experimentally observed equilibrium unfolding transition of the active conformation of α1-AT has been published previously (16). By assuming a linear dependence of protein stability with denaturant concentration, they find that

$$\Delta G_{NU} = \Delta G_{NI}(H_2O) - m_{NI}[D] + \Delta G_{IU}(H_2O) - m_{IU}[D] \quad \text{Eq. 3}$$

where, ΔG_{NU} is the free energy difference between the unfolded and the native state at a denaturant concentration [D], $\Delta G_{NI}(H_2O)$ is the free energy difference between the intermediate state and the native state, m_{NI} is the steepness of the transition from native to the intermediate state, $\Delta G_{IU}(H_2O)$ is the free energy difference between the unfolded state and the intermediate state and, m_{NI} is the steepness of the transition from the intermediate to the unfolded state. Using values reported in Ref (16) for wild type α1-AT, we find that,

$$\Delta G_{NU} \text{ (in kcal/mol)} = 4.63 - 5.84[D] + 6.4 - 2.4[D] \quad \text{Eq. 4}$$

At denaturant concentration ([D]) of 0.6 M GdnHCl, we find that ΔG_{NU} is 6.1 kcal/mol or 10.3 RT (T = 298 K and R = 1.98 cal/mol-K). From our simulations we obtain the folding free energy barriers at T_f. We then reweight the free energy profile to a temperature lower than T_f (temperature of $0.988T_f$) at which the free energy difference between the native and unfolded states is ~10.3

k_BT (Fig. S11). Note that, since we are simulating a single molecule, k_B replaces R. After this reweighting, the folding free energy barrier is 16.4 k_BT (Fig. S11).

In order to convert this free energy barrier into folding rate, we use the data presented in Fig B1 in the supporting information of Ref (17) and find that

$$\Delta G^{\ddagger}/k_BT = -2.75*\log(k_f) + 8.25 \qquad \text{Eq. 5}$$

where, $\Delta G^{\ddagger}/k_BT$ is the folding free energy barrier from simulation and k_f is also the folding rate from simulation obtained by averaging molecular dynamics timesteps.

This equation has been derived from simulation data on several two state folding proteins, spanning a large range of folding free energy barriers (from 0 to 10 k_BT) using the same Cα SBM that is used in the present study. From the above equation, we find that for a $\Delta G^{\ddagger}/k_BT$ of 16.4, as is the case for α1-AT under native conditions, the $\log(k_f)$ is ~-3. Therefore, our simulations predict that the time constant τ for folding is ~1000 sec. This is within an order of magnitude of the experimentally observed folding time constant of ~1000 s.

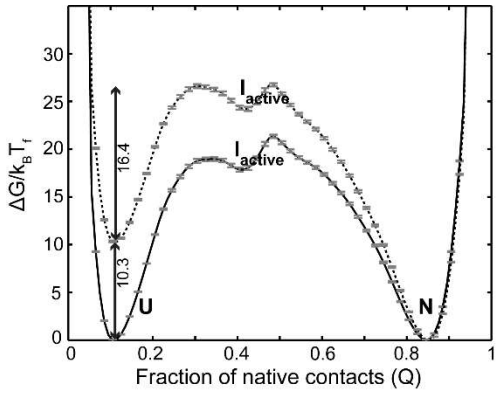

Fig S11 Folding free energy barrier of the active conformation for comparison with experimentally observed refolding rates. The FEP plotted as a function of the fraction of native contacts (Q) at the folding temperature (T_f) is shown as a solid line (reproduced from Fig. 4A), and the FEP at a temperature of $0.988T_f$ is shown as a dotted line. At $0.988T_f$, the difference in free energy between the N and U states is 10.3 (in

k_BT units) and this is similar to the stability of the native state at 0.6 M GdnHCl(16). When the stability is 10.3 (in kBT units), the refolding free energy barrier is 16.4 (in k_BT units). The FEPs and the error bars represent the average and standard deviation from four independent replicates. The native ensemble, N, is at Q~0.84, the intermediate ensemble, I_{active}, is at Q~0.4, and the unfolded ensemble, U, is at Q~0.1.

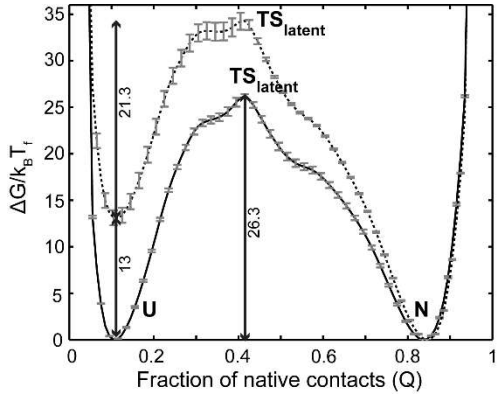

Fig S12 Estimating the increased stability of the latent conformation over the stability of the active conformation. The FEP plotted as a function of the fraction of native contacts (Q) at the folding temperature (T_f) is shown as a solid line (reproduced from Fig. 5A), and the FEP at a temperature of $0.984T_f$ is shown as a dotted line. At $0.984T_f$, the refolding free energy barrier for the latent conformation is 21.3 (in k_BT units) and is similar to the folding free energy barrier of the active conformation at its T_f. When the refolding free energy barrier for the latent conformation is 21.3 (in k_BT units), its stability is 13 (in k_BT units). The FEP and the error bars represent the average and standard deviation from four independent replicates. The native ensemble, N, is at Q~0.84, the transition state ensemble, TS_{latent}, is at Q~0.4, and the unfolded ensemble, U, is at Q~0.1.

SI Discussion

Disease linked polymerization in serpins could arise from both active and latent conformations.

Previous studies have shown, using both experiments(18, 19) and simulations(20, 21), that partially folded ensembles (either intermediates or transition states) can be used to predict the structure of domain-swapped multimers. In this context, we discuss the connection between partially folded ensembles obtained from the present simulations and serpin polymerization.

Serpins are prone to forming polymers *in vitro* and *in vivo*(22). Several human diseases, collectively termed as serpinopathies have been linked to the formation of these polymers. Two different models of serpin polymers have been proposed. These are, the s5A/RCL domain swap model(23) and the C-terminal β-hairpin domain swap model(24). Interestingly, both of them involve the insertion of the RCL into β-sheet A either within the same molecule (intramolecular insertion) or into another serpin molecule (intermolecular insertion)(25). A recent AFM study showed that there is not one but many types of polymers formed by α1-AT, both *in vitro* and *in vivo*(22). However, the structural details of the mechanism by which these polymers form remains unclear. A third model of polymerization is the loop A-sheet model(26), which has been contested owing to structural restrictions involved in forming such a polymer(27). We also find that unlike the two domain swap models, there does not exist direct structural evidence that an uncleaved RCL can intermolecularly insert into β-sheet A. Therefore, we choose to discuss only the domain swap models for serpin polymerization.

Previous studies suggest that exposure to mild denaturant i.e. 1 M GdnHCl leads to the formation of a folding intermediate capable of polymerization(28). We find that in the folding intermediate

I_{active}, domain 1 remains folded but domain 2 has unfolded (Fig. 4C). We propose that the structure of I_{active} can explain the underlying mechanism of the s5A/RCL domain swap model. This model suggests that both s5A and the RCL form a β-hairpin (similar to that observed in the latent conformation) and this inserts into another serpin molecule (Fig. S13B). The structural ensemble at Q~0.34 is similar in stability, in terms of $\Delta G/k_BT_f$, to that of I_{active} (Fig. 4A). The ensemble at Q~0.34 also has an average contact map similar to that of I_{active} (Fig. S6A). However, interactions corresponding to strand s5A are additionally absent. Therefore, the entire polypeptide segment from the N-terminus of strand s5A to the C-terminus of the RCL is unfolded. This segment can also form the local interactions required for the s5A/RCL β-hairpin. We propose that if the s5A/RCL β-hairpin were to form and interact with the unfolded domain 2 of another serpin from the same structural ensemble, it could lead to the insertion of the s5A/RCL β-hairpin from one serpin molecule into another leading to the formation of a dimer in which only one monomer is completely folded (Fig. S13A). Subsequent addition of monomers to the unfolded domain 2 of the dimer could lead to polymerization. Therefore, a structural ensemble similar to I_{active} can explain how polymerization could be initiated in α1-AT. In fact, several protein folding or misfolding intermediates have been shown to domain swap because the partially folded structure corresponds to polypeptide segments which undergo domain swapping(20, 21, 29, 30).

Serpin polymerization can also be induced *in vitro* by a moderate increase in temperature(22). It has been suggested that heating induces a conformational transition to a polymerogenic species(31, 32). Structural evidence suggests that the resulting polymers have an intramolecularly inserted RCL (into β-sheet A), while the C-terminal β-hairpin comprising strands s4B and s5B domain swap intermolecularly into an adjacent serpin molecule (Fig. S13D)(24). Interestingly, the conformational transition of the active to latent α1-AT is also facilitated by heating(33). Given the

observation of an intramolecular RCL insertion, exactly as seen in the latent conformation, we are tempted to propose that the C-terminal β-hairpin domain swapped polymer arises from the latent conformation. Our simulations of the latent conformation suggest the presence of an off-pathway intermediate (termed N*, see Fig. 6C). The average contact map of N* shows that domain 1 i.e. the C-terminal domain, is unfolded while domain 2 remains folded (Fig. S6B). Strands s4B and s5B have an unusually large packing fraction (i.e. the number of native contacts per residue) of ~7.3 (189 contacts for residues 344 to 369) as compared to that of the packing fraction of ~3 for both the active and latent conformations of α1-AT. Higher packing fractions are expected to indicate regions of high stability(34) and therefore, domain swapping of strands s4B and s5B must be facilitated by local unfolding. We suggest that the N* ensemble with an unfolded C-terminal domain i.e. domain 1, likely facilitates the unfolding of strands s4B and s5B. Interaction of such structures with one another and subsequent refolding to the latent conformation could be accompanied by the insertion of strands s4B and s5B into another serpin molecule resulting in the formation of a dimer (Fig. S13C). This dimer can polymerize in a similar fashion to form higher order polymers. Polymerization in α1-AT has been observed experimentally when the latent conformation is heated(33) however no structural details of this polymerization are known. Taken together with the observation of the N* state in the folding landscape of the latent conformation, we suggest that latent α1-AT may lead to polymerization according to the C-terminal β-hairpin domain swap model.

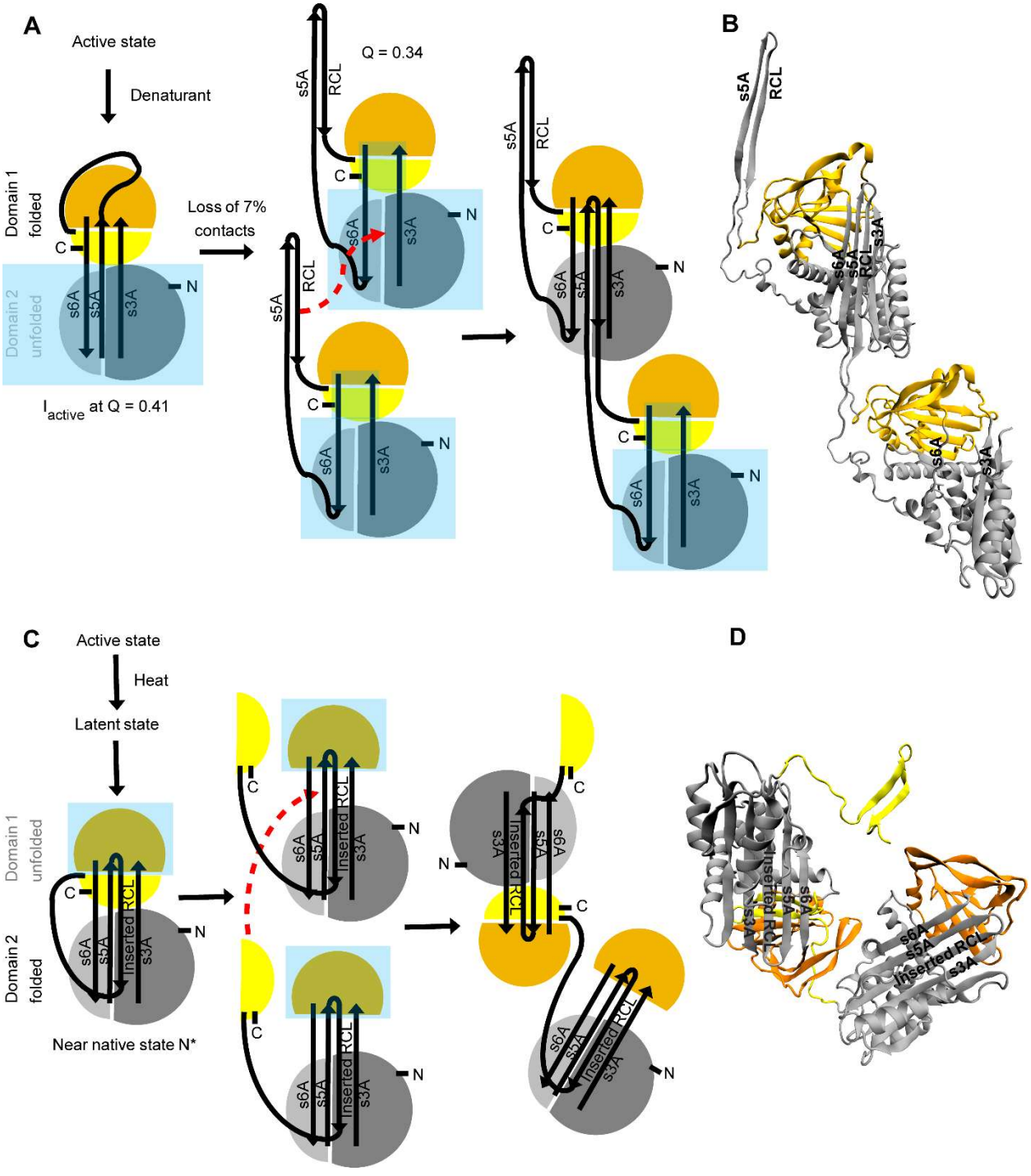

Figure S13 Proposed mechanism of serpin polymerization indicated from the folding simulations. (A-B) s5A/RCL domain swap model. (A) The schematic shows that the addition of denaturant can transform the active conformation into the I_{active} intermediate. The structures of active α1-AT are depicted using cartoons with the two domains colored in two shades of yellow (domain 1) and two shades of grey (domain

2, see Fig. S1A). Cyan colored filled boxes drawn over the cartoons mark only those regions which are unfolded. We find that further unfolding of I_{active} leads to the complete loss of interactions between strands s3A/s5A and s5A/s6A (ensemble at Q~0.34 in Fig. 4A). Thus, strands s6A, s5A, the RCL and their intervening loops are unfolded, and local interactions can facilitate the formation of the s5A/RCL β-hairpin. We propose that when two such structures having the s5A/RCL β-hairpin and an unfolded domain 2 interact, the s5A/RCL β-hairpin of one serpin can insert into domain 2 of another, leading to domain swapping of the s5A/RCL β-hairpin. (B) A model of the s5A/RCL domain swap structure obtained from x-ray crystallography (reconstructed from PDB 2ZNH according to Ref(23)). (C-D) C-terminal β-hairpin domain swap model. (C) The schematic shows that heating can transform the active conformation into the latent conformation(33). Our study shows that the latent conformation populates near native N* state in which domain 1 is partially unfolded. The structures of latent α1-AT are depicted using cartoons with the two domains colored in two shades of yellow (domain 1) and two shades of grey (domain 2, see Fig. S1B). Cyan colored filled boxes drawn over the cartoons mark only those regions which are unfolded. We propose that partial unfolding of domain 1 can facilitate the separation of the C-terminal β-hairpin from the rest of the serpin. When two such structures having domain 1 unfolded and a free C-terminal β-hairpin interact, the C-terminal β-hairpin of one serpin can insert into domain 1 of another, leading to domain swapping of the C-terminal β-hairpin. (D) A model of the C-terminal β-hairpin domain swap structure obtained from x-ray crystallography (PDB 3TIP(24)).

SI Table

Table S1. Table showing the T_f obtained after WHAM analysis and the numbers of unfolding and folding transitions observed, for each REUS replicate for both active and latent conformations of α1-AT. All temperatures are in reduced units (1 K=0.008314 reduced units). For canonical MD simulations, unfolding transitions were initiated from the native state, and folding transitions were initiated from an unfolded state at the temperatures specified below. These simulations were performed using the stochastic dynamics integrator, with a time step of 0.0005 ps while saving structures every 2000 steps, with non-bonded interactions cut-off at 3 nm and a typical simulation time of 1×10^8 time steps.

Protein (PDB ID)	Replicate number and REUS simulation temperature	64 starting structures are obtained from a canonical MD simulation at the specified temperature	T_f obtained after WHAM analysis	Number of unfolding transitions from all 64 replicas of REUS	Number of folding transitions from all 64 replicas of REUS
Active conformation of α1-AT (PDB 1QLP)	Replicate 1 at 1.2056	Unfolding transition at 1.2638	1.2048	14	11
	Replicate 2 at 1.2056	Folding transition at 1.1308	1.2041	13	18
	Replicate 3 at 1.1973	Unfolding transition at 1.2721	1.2050	17	17

	Replicate 4 at 1.1973	Folding transition at 1.1058	1.2039	14	16
Latent conformation of α1-AT (PDB 1IZ2)	Replicate 1 at 1.2222	Unfolding transition at 1.2846	1.2125	9	9
	Replicate 2 at 1.2222	Folding transition at 1.0851	1.2098	10	11
	Replicate 3 at 1.2139	Unfolding transition at 1.2763	1.2135	10	8
	Replicate 4 at 1.2139	Folding transition at 1.0767	1.2112	9	10

SI Lists of native contacts used to construct the SBMs of active and latent α1-AT

List S1. Native contacts of active α1-AT (1115 contacts)

(1,5), (1,356), (1,357), (1,358), (2,6), (2,73), (2,76), (2,77), (2,357), (2,359), (3,7), (3,64), (3,65), (3,69), (3,72), (4,8), (4,60), (4,65), (4,69), (4,70), (4,73), (4,74), (4,77), (5,9), (5,247), (5,357), (5,358), (5,359), (5,360), (6,10), (6,247), (7,11), (7,58), (7,59), (7,60), (7,61), (7,65), (8,12), (8,32), (8,33), (8,36), (8,58), (8,60), (8,74), (8,77), (8,359), (8,360), (9,13), (9,247), (9,360), (10,14), (10,17), (10,59), (10,246), (10,247), (11,15), (11,32), (11,35), (11,36), (11,57), (11,58), (11,59), (11,281), (11,284), (12,16), (12,30), (12,32), (12,351), (12,360), (12,363), (13,17), (13,245), (13,246), (13,247), (13,250), (13,351), (13,363), (14,18), (14,59), (14,284), (15,19), (15,30), (15,32), (15,35), (15,277), (15,280), (15,281), (15,284), (16,20), (16,28), (16,30), (16,242), (16,245), (16,349), (16,363), (16,364), (16,365), (17,21), (17,242), (17,243), (17,245), (18,22), (18,280), (18,283), (18,284), (19,23), (19,28), (19,30), (19,275), (19,280), (20,24), (20,28), (20,242), (20,365), (21,239), (22,275), (22,279), (22,280), (22,283), (23,28), (24,238), (24,239), (24,242), (24,365), (24,367), (25,367), (25,371), (26,365), (26,367), (26,371), (27,271), (27,272), (27,275), (27,365), (27,366), (27,367), (27,368), (27,371), (28,271), (28,275), (28,364), (28,365), (29,271), (29,272), (29,273), (29,275), (29,314), (29,315), (29,316), (29,350), (29,362), (29,363), (30,35), (30,275), (30,312), (30,314), (30,362), (30,363), (31,35), (31,166), (31,312), (31,314), (31,361), (31,362), (32,36), (32,360), (32,361), (32,362), (32,363), (33,37), (33,74), (33,77), (33,78), (33,81), (33,359), (33,360), (33,361), (34,38), (34,92), (34,164), (34,166), (34,312), (35,39), (35,58), (35,164), (35,275), (35,277), (35,281), (35,312), (36,40), (36,58), (36,74), (37,41), (37,74), (37,77), (37,78), (37,116), (38,42), (38,94), (38,162), (38,164), (38,277), (39,43), (39,54), (39,58), (39,162), (39,277), (39,281), (39,286), (39,289), (39,290), (39,307), (39,310), (40,44), (40,55), (40,58), (40,70), (40,71), (40,74), (40,116), (41,94), (41,96), (41,108), (41,112), (41,116), (41,118), (41,162), (42,46), (42,54), (42,96), (42,108), (42,160), (42,162), (42,289), (42,296), (42,307), (43,51), (43,54), (43,55), (43,71), (44,51), (44,71), (44,108), (44,111), (44,115), (44,116), (44,298), (45,108), (45,296), (45,298), (45,299), (46,50), (46,51), (46,54), (46,289), (46,295), (46,296), (47,51), (47,295), (47,297), (47,298), (48,52), (48,295), (49,53), (49,288), (50,54), (50,288), (50,289), (50,293), (50,294), (50,295), (51,55), (51,62), (51,67), (51,71), (52,56), (52,62), (52,67), (53,57), (53,285), (53,286), (53,287), (53,288), (53,289), (54,58), (54,286), (54,289), (55,60), (55,62), (55,67), (55,70), (55,71), (56,60), (56,61), (56,62), (57,284), (57,286), (58,281), (58,286), (60,65), (60,70), (60,74), (61,65), (62,66), (62,67), (62,70), (65,69), (65,70), (66,70), (67,71), (68,72), (69,73), (70,74), (71,75), (71,115), (71,116), (72,76), (72,79), (73,77), (74,78), (74,116), (75,79), (75,115), (75,116), (76,80), (76,357), (77,81), (77,357), (77,359), (78,82), (78,115), (78,116), (78,117), (79,83), (79,115), (79,117), (80,225), (80,354), (80,356), (80,357), (81,86), (81,88), (81,90), (81,92), (81,225), (81,354), (81,361), (82,117), (83,87), (83,356), (86,224), (86,225), (87,171), (87,223), (87,224), (88,168), (88,170), (88,171), (88,222), (88,223), (88,224), (88,225), (88,226), (88,361), (89,167), (89,168), (89,169), (89,170), (89,171), (90,166), (90,167), (90,168), (90,361), (90,362), (91,142), (91,143), (91,144), (91,165), (91,166), (91,167), (92,164), (92,165), (92,166), (93,119), (93,138), (93,164), (93,165), (94,118), (94,119), (94,138), (94,162), (94,163), (94,164), (95,119), (95,121), (95,138), (95,161), (95,162), (95,163), (96,108), (96,112), (96,118), (96,119), (96,120), (96,121), (96,160), (96,161), (96,162), (97,121), (97,123), (97,125), (97,134), (97,135), (97,138), (97,160), (97,161), (97,163), (98,102), (98,104), (98,108), (98,109), (98,120), (98,121), (98,122), (98,123), (98,125), (98,159), (98,160), (98,299), (98,305), (99,104), (99,122), (99,123), (99,125), (99,159), (100,122),

(100,123), (100,124), (100,125), (102,159), (102,300), (102,303), (102,304), (102,305), (103,300), (103,301), (103,302), (104,109), (104,120), (104,121), (104,122), (104,299), (104,300), (105,109), (105,297), (105,298), (105,299), (105,301), (106,110), (107,111), (107,115), (108,112), (108,160), (108,298), (108,299), (109,113), (109,120), (110,114), (111,115), (111,116), (112,116), (112,117), (112,118), (112,119), (112,120), (113,117), (113,118), (113,119), (113,120), (119,138), (119,165), (121,138), (122,134), (123,131), (123,134), (123,135), (124,130), (124,131), (124,134), (125,131), (125,135), (125,154), (125,156), (125,158), (125,159), (125,161), (127,131), (128,132), (128,154), (128,155), (128,156), (129,133), (130,134), (131,135), (131,154), (132,136), (132,149), (132,151), (132,152), (132,153), (132,154), (133,137), (134,138), (135,139), (135,151), (135,154), (135,161), (135,163), (136,140), (136,147), (136,149), (136,150), (136,151), (137,141), (138,142), (138,163), (138,165), (139,143), (139,147), (139,150), (139,163), (139,165), (140,144), (140,145), (140,147), (140,148), (140,149), (142,165), (143,147), (143,165), (143,167), (146,167), (146,270), (146,272), (146,313), (146,315), (146,317), (146,372), (147,165), (147,166), (147,167), (147,313), (147,315), (148,152), (148,313), (150,163), (150,164), (150,165), (150,274), (150,311), (150,313), (151,163), (151,309), (151,311), (152,309), (153,309), (154,158), (154,161), (154,309), (155,278), (155,308), (155,309), (157,292), (157,306), (157,308), (158,308), (158,309), (159,304), (159,305), (159,306), (159,308), (160,296), (160,299), (160,305), (160,306), (160,307), (160,308), (160,309), (161,307), (161,309), (162,277), (162,307), (162,309), (162,310), (162,311), (163,309), (163,311), (164,277), (164,311), (164,312), (164,313), (165,313), (166,312), (166,313), (166,314), (166,315), (166,361), (166,362), (167,315), (167,317), (168,222), (168,226), (168,315), (168,316), (168,317), (168,318), (168,352), (168,362), (169,317), (169,319), (169,322), (169,323), (170,222), (170,317), (170,318), (170,319), (170,322), (171,222), (171,223), (171,322), (171,323), (172,176), (172,199), (172,220), (172,221), (172,222), (172,223), (172,228), (172,230), (172,266), (172,318), (172,319), (172,320), (172,321), (172,322), (173,221), (173,222), (173,223), (173,224), (173,330), (173,355), (174,201), (174,204), (174,220), (174,221), (174,259), (174,330), (174,332), (175,201), (175,320), (175,321), (175,327), (175,330), (176,199), (176,200), (176,201), (176,205), (176,220), (176,264), (176,266), (176,320), (176,321), (177,181), (177,200), (177,201), (177,320), (178,268), (178,320), (178,321), (180,198), (180,200), (180,201), (181,197), (181,198), (181,199), (181,268), (181,320), (182,197), (182,198), (182,200), (183,195), (183,196), (183,197), (184,195), (184,196), (184,198), (184,200), (184,265), (184,343), (184,345), (185,193), (185,194), (185,195), (185,198), (185,345), (186,193), (186,194), (186,196), (186,198), (186,267), (186,344), (186,345), (186,346), (186,347), (186,348), (186,366), (186,367), (186,368), (186,369), (187,191), (187,192), (187,193), (187,345), (187,346), (187,347), (188,192), (188,193), (188,194), (188,347), (188,367), (188,368), (190,235), (190,236), (194,368), (194,370), (195,370), (196,267), (196,368), (196,369), (196,370), (197,267), (197,268), (197,369), (197,370), (198,265), (198,266), (198,267), (198,344), (198,345), (199,265), (199,266), (199,267), (199,268), (199,318), (199,319), (199,320), (200,264), (200,265), (201,205), (201,220), (201,259), (201,262), (201,263), (201,264), (201,330), (201,332), (202,262), (202,263), (202,332), (202,334), (202,339), (203,260), (203,261), (203,262), (203,332), (203,334), (204,259), (204,260), (204,261), (204,332), (204,333), (204,334), (204,335), (205,218), (205,220), (205,259), (205,260), (205,262), (205,264), (206,218), (206,219), (206,253), (206,256), (206,257), (206,258), (206,259), (207,216), (207,217), (207,218), (207,258), (207,260), (207,262), (207,264), (207,342), (207,344), (208,216), (208,217), (208,219), (208,253), (208,256), (208,258), (208,260), (209,214), (209,215), (209,216), (209,234), (209,260), (210,214), (210,215), (210,217), (210,249), (210,253), (211,240), (212,240), (212,244), (212,249), (213,217), (213,231), (213,237), (213,240), (213,241), (213,244), (213,245), (213,249), (213,253), (214,233), (214,234), (214,237), (214,240), (215,231), (215,232), (215,233), (215,234), (215,241), (216,230), (216,231), (216,232), (216,234), (216,260), (216,340),

(216,341), (216,342), (216,344), (217,229), (217,230), (217,231), (217,253), (218,228), (218,229), (218,230), (218,264), (218,266), (218,344), (219,227), (219,228), (219,229), (219,253), (219,254), (219,256), (219,259), (220,227), (220,228), (220,230), (220,264), (220,266), (221,226), (221,227), (221,257), (221,259), (221,355), (222,226), (222,228), (222,318), (222,352), (222,355), (224,355), (225,354), (225,355), (225,356), (226,352), (226,353), (226,355), (226,361), (227,254), (227,351), (227,352), (227,353), (227,355), (228,351), (228,352), (229,250), (229,253), (229,254), (229,350), (229,351), (229,353), (230,266), (230,318), (230,344), (230,348), (230,349), (230,350), (230,352), (231,241), (231,245), (231,250), (231,253), (231,344), (231,348), (231,349), (231,350), (231,351), (232,342), (232,343), (232,344), (232,346), (232,348), (233,237), (233,238), (233,241), (233,346), (233,347), (233,348), (234,346), (235,346), (236,240), (237,241), (238,242), (238,347), (238,349), (238,365), (238,367), (239,243), (240,244), (241,245), (241,349), (242,349), (242,365), (244,249), (245,249), (245,250), (245,349), (245,351), (245,363), (246,250), (247,251), (247,358), (248,252), (249,253), (250,254), (250,351), (250,353), (251,255), (251,358), (252,256), (254,353), (254,355), (254,358), (255,358), (260,340), (261,334), (261,335), (261,337), (261,339), (261,340), (262,339), (262,340), (262,342), (263,339), (263,340), (263,341), (263,342), (264,342), (264,344), (265,341), (265,342), (265,343), (265,344), (266,318), (266,344), (266,348), (267,318), (267,344), (267,348), (267,369), (268,318), (268,319), (268,320), (268,369), (269,316), (269,317), (269,318), (269,348), (269,366), (269,369), (270,315), (270,316), (270,317), (270,369), (270,372), (271,315), (271,316), (271,350), (271,365), (271,366), (271,369), (271,371), (271,372), (272,313), (272,314), (272,315), (272,372), (273,312), (273,313), (273,314), (274,311), (274,312), (274,313), (275,280), (275,311), (275,312), (276,280), (276,309), (276,310), (276,311), (277,281), (277,290), (277,310), (277,311), (277,312), (278,282), (278,290), (278,306), (278,307), (278,308), (278,310), (279,283), (280,284), (281,285), (281,286), (281,290), (282,286), (282,287), (286,290), (287,291), (288,293), (288,294), (289,294), (289,296), (289,307), (290,306), (290,307), (290,308), (290,310), (291,306), (292,304), (292,305), (292,306), (294,305), (294,306), (296,300), (296,305), (296,306), (296,307), (297,301), (299,305), (300,305), (314,362), (315,362), (315,372), (316,350), (316,352), (318,352), (321,325), (321,326), (321,327), (334,339), (344,348), (347,365), (347,366), (347,367), (348,365), (348,366), (349,363), (349,364), (349,365), (350,362), (350,363), (350,364), (350,365), (350,366), (351,360), (351,362), (351,363), (352,360), (352,361), (352,362), (353,358), (353,359), (353,360), (353,361), (354,358), (354,359), (354,361), (367,371).

List S2. Native contacts of latent α1-AT (1137 contacts)

(2,6), (2,77), (2,356), (2,357), (2,358), (3,7), (3,65), (3,69), (3,73), (4,8), (4,60), (4,65), (4,69), (4,70), (4,73), (4,74), (4,77), (5,9), (5,77), (5,247), (5,357), (5,358), (5,359), (5,360), (6,10), (6,247), (7,11), (7,58), (7,59), (7,60), (7,61), (7,65), (8,12), (8,32), (8,33), (8,36), (8,58), (8,60), (8,74), (8,77), (8,359), (8,360), (9,13), (9,247), (9,358), (9,360), (10,14), (10,17), (10,59), (10,246), (10,247), (11,15), (11,32), (11,35), (11,36), (11,57), (11,58), (11,59), (11,281), (11,284), (12,16), (12,30), (12,32), (12,351), (12,360), (12,363), (13,17), (13,242), (13,245), (13,246), (13,247), (13,250), (13,351), (13,363), (14,18), (14,59), (14,284), (15,19), (15,30), (15,35), (15,277), (15,280), (15,281), (15,284), (16,20), (16,28), (16,30), (16,242), (16,245), (16,349), (16,363), (16,364), (16,365), (17,21), (17,242), (17,243), (17,245), (17,246), (18,22), (18,280), (18,283), (18,284), (19,23), (19,28), (19,30), (19,275), (19,277), (19,280), (19,365), (20,24), (20,28), (20,239), (20,242), (20,365), (21,243), (21,283), (22,279), (22,280), (23,341), (23,365), (24,238), (24,239), (24,341), (24,343), (24,365), (24,367), (25,340), (25,341), (25,342), (25,343),

(25,367), (25,368), (26,339), (26,340), (26,341), (26,365), (26,367), (26,368), (27,271), (27,272), (27,275), (27,339), (27,340), (27,365), (27,366), (27,367), (27,368), (27,370), (28,271), (28,272), (28,275), (28,364), (28,365), (29,271), (29,272), (29,273), (29,275), (29,312), (29,314), (29,315), (29,316), (29,350), (29,362), (29,363), (30,35), (30,275), (30,312), (30,314), (30,362), (30,363), (31,35), (31,166), (31,312), (31,327), (31,361), (31,362), (32,36), (32,360), (32,361), (32,362), (32,363), (33,37), (33,74), (33,77), (33,78), (33,81), (33,92), (33,166), (33,359), (33,360), (33,361), (34,38), (34,92), (34,164), (34,166), (34,312), (34,328), (34,329), (35,39), (35,58), (35,277), (35,281), (35,312), (35,329), (36,40), (36,58), (36,60), (36,74), (37,41), (37,74), (37,78), (37,92), (37,94), (37,116), (37,164), (38,42), (38,94), (38,162), (38,164), (38,329), (38,331), (39,43), (39,54), (39,58), (39,281), (39,286), (39,289), (39,290), (39,307), (39,310), (39,329), (39,331), (40,44), (40,55), (40,58), (40,70), (40,71), (40,74), (40,116), (41,94), (41,95), (41,96), (41,108), (41,112), (41,116), (41,118), (41,162), (42,46), (42,54), (42,108), (42,160), (42,162), (42,289), (42,296), (42,299), (42,307), (42,331), (43,51), (43,54), (43,55), (43,67), (43,71), (44,67), (44,71), (44,107), (44,108), (44,111), (44,112), (44,115), (44,116), (44,298), (45,108), (45,296), (45,298), (45,299), (46,50), (46,51), (46,54), (46,295), (46,296), (47,51), (47,295), (47,297), (47,298), (48,52), (49,53), (49,288), (50,54), (50,286), (50,288), (50,289), (50,293), (50,294), (50,295), (50,296), (51,55), (51,62), (51,67), (52,56), (52,62), (53,57), (53,285), (53,286), (53,287), (53,288), (53,289), (54,58), (54,286), (54,289), (55,60), (55,62), (55,67), (55,70), (55,71), (56,60), (56,61), (56,62), (57,284), (57,286), (58,281), (58,286), (60,65), (60,70), (60,74), (61,65), (62,66), (62,67), (62,70), (65,69), (65,70), (66,70), (67,71), (68,72), (69,73), (70,74), (70,116), (71,75), (71,115), (71,116), (72,76), (72,115), (73,77), (74,78), (74,116), (75,79), (75,114), (75,115), (75,117), (76,80), (77,81), (77,357), (77,359), (78,82), (78,92), (78,94), (78,116), (78,117), (79,83), (79,117), (80,354), (80,356), (80,357), (81,90), (81,91), (81,92), (81,166), (81,354), (81,357), (81,359), (81,361), (82,90), (82,91), (82,92), (82,117), (83,90), (83,225), (83,356), (84,90), (85,224), (85,225), (87,170), (87,171), (87,223), (87,224), (88,168), (88,169), (88,170), (88,222), (88,223), (88,225), (88,226), (89,144), (89,167), (89,168), (89,169), (90,166), (90,167), (90,168), (90,225), (90,354), (90,361), (91,142), (91,165), (91,166), (91,167), (92,164), (92,165), (92,166), (93,138), (93,164), (93,165), (94,116), (94,117), (94,118), (94,119), (94,138), (94,163), (94,164), (95,118), (95,119), (95,121), (95,138), (95,162), (95,163), (96,108), (96,109), (96,112), (96,118), (96,119), (96,120), (96,121), (96,160), (96,161), (96,162), (97,121), (97,123), (97,125), (97,134), (97,135), (97,138), (97,160), (97,161), (97,163), (98,102), (98,104), (98,108), (98,109), (98,120), (98,121), (98,122), (98,123), (98,125), (98,159), (98,160), (99,104), (99,122), (99,123), (99,125), (99,158), (99,159), (100,122), (100,123), (100,124), (102,159), (102,160), (102,300), (102,302), (102,303), (102,305), (103,300), (103,302), (104,108), (104,109), (104,122), (104,160), (104,299), (104,300), (104,302), (105,109), (105,297), (105,298), (105,299), (105,300), (105,301), (106,110), (107,111), (108,112), (108,160), (108,162), (108,298), (108,299), (109,113), (109,120), (110,114), (111,115), (112,116), (112,117), (112,118), (113,117), (113,118), (113,120), (119,138), (121,134), (121,138), (122,134), (123,131), (123,134), (123,135), (124,130), (124,131), (124,134), (125,131), (125,135), (125,154), (125,155), (125,156), (125,158), (125,159), (125,161), (126,156), (127,131), (128,132), (128,154), (128,155), (128,156), (129,133),

(130,134), (131,135), (131,154), (132,136), (132,151), (132,152), (132,153), (132,154), (133,137), (134,138), (135,139), (135,151), (135,154), (135,161), (135,163), (135,330), (136,140), (136,147), (136,149), (136,150), (136,151), (137,141), (138,142), (138,163), (138,165), (139,143), (139,147), (139,150), (139,163), (139,165), (139,328), (140,144), (140,145), (140,147), (140,148), (140,149), (142,165), (142,167), (143,147), (143,165), (143,166), (143,167), (143,326), (143,327), (144,167), (146,167), (146,315), (146,316), (146,317), (146,324), (146,326), (147,313), (147,315), (147,326), (147,327), (147,328), (148,152), (150,163), (150,313), (150,328), (150,330), (150,334), (150,335), (151,330), (151,333), (151,334), (152,334), (152,335), (153,330), (153,333), (153,335), (154,158), (154,161), (154,330), (154,332), (155,332), (158,332), (159,303), (159,305), (160,296), (160,299), (160,305), (160,307), (160,331), (160,332), (161,330), (161,331), (161,332), (162,329), (162,330), (162,331), (163,329), (163,330), (164,328), (164,329), (165,327), (166,325), (166,326), (166,327), (166,361), (166,362), (167,324), (167,325), (167,326), (168,222), (168,226), (168,323), (168,324), (168,325), (168,352), (168,361), (168,362), (169,323), (169,324), (170,222), (170,322), (170,323), (171,222), (171,223), (171,321), (171,322), (172,176), (172,220), (172,221), (172,222), (172,223), (172,228), (172,230), (172,266), (172,318), (172,319), (172,320), (172,321), (172,322), (172,323), (173,221), (173,222), (173,223), (173,224), (174,201), (174,204), (174,205), (174,206), (174,220), (174,221), (174,259), (175,201), (175,320), (175,321), (176,181), (176,199), (176,200), (176,201), (176,205), (176,220), (176,264), (176,266), (176,319), (176,320), (176,321), (177,181), (177,200), (177,201), (177,320), (178,268), (178,320), (178,321), (180,200), (180,201), (181,197), (181,198), (181,199), (181,268), (181,320), (182,197), (182,198), (182,200), (183,195), (183,196), (183,197), (184,194), (184,195), (184,196), (184,198), (184,200), (184,265), (185,193), (185,194), (185,195), (185,196), (186,193), (186,194), (186,196), (186,342), (186,343), (186,344), (186,368), (186,369), (187,191), (187,192), (187,193), (187,342), (188,192), (188,194), (188,340), (188,341), (188,342), (188,368), (189,342), (196,267), (196,369), (196,371), (197,267), (197,268), (197,371), (198,265), (198,266), (198,267), (198,344), (198,369), (199,264), (199,265), (199,266), (199,267), (199,268), (199,318), (199,319), (199,320), (200,263), (200,264), (200,265), (201,205), (201,263), (201,264), (202,261), (202,262), (202,263), (203,261), (203,262), (204,259), (204,260), (204,261), (205,218), (205,219), (205,220), (205,259), (205,260), (205,262), (205,264), (206,217), (206,218), (206,219), (206,221), (206,256), (206,257), (206,258), (206,259), (207,216), (207,217), (207,218), (207,256), (207,258), (207,260), (207,262), (208,216), (208,217), (208,219), (208,252), (208,253), (208,256), (208,260), (209,214), (209,215), (209,216), (209,234), (209,253), (209,260), (210,214), (210,215), (210,217), (210,249), (210,253), (211,240), (212,240), (212,244), (212,249), (213,217), (213,231), (213,237), (213,240), (213,241), (213,244), (213,245), (213,249), (213,253), (214,233), (214,234), (214,237), (214,240), (215,231), (215,232), (215,233), (215,234), (215,241), (216,230), (216,231), (216,232), (216,234), (216,260), (217,229), (217,230), (217,231), (217,253), (218,228), (218,229), (218,230), (218,264), (218,266), (219,227), (219,228), (219,229), (219,253), (219,254), (219,256), (220,227), (220,228), (220,230), (220,264), (220,266), (221,226), (221,227), (221,257), (221,259), (221,355), (222,226), (222,227), (222,228), (222,323), (222,352), (224,355), (225,354), (225,355), (225,356), (226,352), (226,353), (226,354), (226,355), (226,361), (227,254), (227,351), (227,352), (227,353), (227,355), (228,350), (228,351), (228,352), (229,250), (229,253), (229,254), (229,350),

(229,351), (229,353), (230,266), (230,318), (230,348), (230,349), (230,350), (231,241), (231,245), (231,249), (231,250), (231,253), (231,348), (231,349), (231,350), (231,351), (232,346), (232,347), (232,348), (233,237), (233,238), (233,241), (233,346), (233,347), (233,348), (234,238), (235,345), (237,241), (238,242), (238,343), (238,347), (238,349), (238,365), (239,243), (239,343), (240,244), (241,245), (241,349), (242,349), (242,365), (244,249), (245,250), (245,349), (245,351), (245,363), (246,250), (247,251), (247,358), (248,252), (249,253), (250,254), (250,351), (250,353), (251,255), (251,358), (252,256), (254,353), (254,355), (254,358), (266,318), (266,348), (267,318), (267,344), (267,348), (267,366), (267,369), (268,318), (268,319), (268,320), (268,371), (269,316), (269,317), (269,318), (269,348), (269,366), (269,368), (269,369), (269,370), (269,371), (270,316), (270,317), (270,370), (270,371), (270,372), (271,315), (271,316), (271,350), (271,365), (271,366), (271,368), (271,370), (271,372), (272,313), (272,314), (272,315), (272,337), (272,372), (273,312), (273,313), (273,314), (273,337), (274,311), (274,312), (274,313), (274,335), (274,336), (274,337), (275,280), (275,311), (275,312), (275,336), (276,280), (276,309), (276,310), (276,311), (276,336), (277,281), (277,290), (277,310), (277,312), (277,329), (278,282), (278,290), (278,306), (278,307), (278,308), (278,310), (279,283), (280,284), (281,286), (281,290), (282,286), (282,287), (286,290), (287,291), (288,293), (288,294), (289,294), (289,296), (289,306), (289,307), (290,306), (290,307), (290,308), (290,309), (290,310), (291,306), (292,306), (294,304), (294,305), (294,306), (296,300), (296,303), (296,305), (296,306), (296,307), (297,301), (299,305), (300,305), (307,331), (308,331), (308,332), (309,331), (309,332), (309,333), (309,334), (309,335), (309,336), (310,329), (310,330), (310,331), (311,328), (311,329), (311,330), (311,332), (311,333), (311,334), (311,335), (311,336), (312,327), (312,328), (312,329), (313,327), (313,328), (313,334), (313,335), (313,337), (314,326), (314,327), (314,362), (315,325), (315,326), (315,362), (315,372), (316,323), (316,324), (316,325), (316,350), (316,352), (316,362), (317,322), (317,323), (317,324), (318,322), (318,323), (318,348), (318,350), (318,352), (323,352), (325,352), (325,362), (326,362), (327,362), (330,334), (340,368), (340,370), (342,367), (342,368), (343,347), (343,367), (344,348), (344,366), (344,367), (344,368), (344,369), (347,365), (347,366), (347,367), (348,365), (348,366), (349,363), (349,364), (349,365), (350,362), (350,363), (350,364), (350,365), (350,366), (351,360), (351,362), (351,363), (352,360), (352,361), (352,362), (353,358), (353,359), (353,360), (353,361), (354,358), (354,359), (354,361).

SI References

1. Sillitoe I, et al. (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* 43(Database issue):D376-81.

2. Tsutsui Y, Dela Cruz R, Wintrode PL (2012) Folding mechanism of the metastable serpin α1-antitrypsin. *Proc Natl Acad Sci U S A* 109(12):4467–72.

3. Stocks BB, Sarkar A, Wintrode PL, Konermann L (2012) Early hydrophobic collapse of α_1-antitrypsin facilitates formation of a metastable state: insights from oxidative labeling and mass spectrometry. *J Mol Biol* 423(5):789–99.

4. Im H, Woo M-S, Hwang KY, Yu M-H (2002) Interactions causing the kinetic trap in serpin protein folding. *J Biol Chem* 277(48):46347–54.

5. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298(5):937–53.

6. Sobolev V, Sorokine A, Prilusky J, Abola E, Edelman M (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15(4):327–332.

7. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* 38(Web Server issue):W657-61.

8. Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 113(15):6042.

9. Murata K, Sugita Y, Okamoto Y (2004) Free energy calculations for DNA base stacking by replica-exchange umbrella sampling. *Chem Phys Lett* 385(1):1–7.

10. Best RB (2013) How well does a funneled energy landscape capture the folding mechanism of spectrin domains? *J Phys Chem B* 117(42):13235–44.

11. Giri Rao VVH, Gosavi S (2015) Structural Perturbations Present in the Folding Cores of Interleukin-33 and Interleukin-1β Correlate to Differences in Their Function. *J Phys Chem B* 119(34):11203–14.

12. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13(8):1011–1021.

13. Predescu C, Predescu M, Ciobanu C V (2005) On the efficiency of exchange in parallel tempering monte carlo simulations. *J Phys Chem B* 109(9):4189–96.

14. Pronk S, et al. (2013) GROMACS 4.5: A High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* 29(7):845–854.

15. Bonomi M, et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* 180(10):1961–1972.

16. James EL, Whisstock JC, Gore MG, Bottomley SP (1999) Probing the unfolding pathway of alpha1-antitrypsin. *J Biol Chem* 274(14):9482–8.

17. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc* 126(27):8426–32.

18. Kim DE, Fisher C, Baker D (2000) A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* 298(5):971–984.

19. Moschen T, et al. (2014) A kinetic study of domain swapping of Protein L. *Phys Chem Chem Phys* 16(14):6383.

20. Mascarenhas NM, Gosavi S (2016) Protein Domain-Swapping Can Be a Consequence of Functional Residues. *J Phys Chem B* 120(28):6929–6938.

21. Zhuravlev PI, Reddy G, Straub JE, Thirumalai D (2014) Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins. *J Mol Biol* 426(14):2653–2666.

22. Gaczynska M, et al. (2016) AFM Imaging Reveals Topographic Diversity of Wild Type and Z Variant Polymers of Human α1-Proteinase Inhibitor. *PLoS One* 11(3):e0151902.

23. Yamasaki M, Li W, Johnson DJD, Huntington JA (2008) Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* 455(7217):1255–8.

24. Yamasaki M, Sendall TJ, Pearce MC, Whisstock JC, Huntington JA (2011) Molecular basis of α1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep* 12(10):1011–1017.

25. Knaupp AS, et al. (2013) The Roles of Helix I and Strand 5A in the Folding, Function and Misfolding of α1-Antitrypsin. *PLoS One* 8(1):e54766.

26. Sivasothy P, Dafforn TR, Gettins PG, Lomas DA (2000) Pathogenic alpha 1-antitrypsin polymers are formed by reactive loop-beta-sheet A linkage. *J Biol Chem* 275(43):33663–8.

27. Huntington JA, Whisstock JC (2010) Molecular contortionism - on the physical limits of serpin "loop-sheet" polymers. *Biol Chem* 391(8):973–82.

28. Koloczek H, Guz A, Kaszycki P (1996) Fluorescence-detected polymerization kinetics of human alpha 1-antitrypsin. *J Protein Chem* 15(5):447–54.

29. Yang S, et al. (2004) Domain swapping is a consequence of minimal frustration. *Proc Natl Acad Sci U S A* 101(38):13786–91.

30. Tian P, Best RB (2016) Structural Determinants of Misfolding in Multidomain Proteins. *PLOS Comput Biol* 12(5):e1004933.

31. Dafforn TR, Mahadeva R, Elliott PR, Sivasothy P, Lomas DA (1999) A kinetic mechanism for the polymerization of alpha1-antitrypsin. *J Biol Chem* 274(14):9548–55.

32. James EL, Bottomley SP (1998) The Mechanism of α1-Antitrypsin Polymerization Probed by Fluorescence Spectroscopy. *Arch Biochem Biophys* 356(2):296–300.

33. Lomas DA, Elliott PR, Chang WS, Wardell MR, Carrell RW (1995) Preparation and characterization of latent alpha 1-antitrypsin. *J Biol Chem* 270(10):5282–8.

34. Ghosh K, Dill KA (2009) Computing protein stabilities from their chain lengths. *Proc Natl Acad Sci U S A* 106(26):10649–54.