AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Ole Juul Andersen, Michael Wulff Risør, Emil Christian Poulsen, Niels Chr. Nielsen, Yinglong Miao, Jan J. Enghild, and Birgit Schiøtt (2017). The reactive center loop insertion in α -1-antitrypsin captured by accelerated molecular dynamics simulation. In *Biochemistry* 56 (4), pp 634–646.

DOI: <https://doi.org/10.1021/acs.biochem.6b00839>

Publication metadata

Title: The reactive center loop insertion in α -1-antitrypsin captured by accelerated molecular dynamics simulation

Author(s): Ole Juul Andersen, Michael Wulff Risør, Emil Christian Poulsen, Niels Chr. Nielsen, Yinglong Miao, Jan J. Enghild, and Birgit Schiøtt

Journal: *Biochemistry*

DOI/Link: <https://doi.org/10.1021/acs.biochem.6b00839>

Document version: Accepted manuscript (post-print)

© The authors 2017. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Biochemistry* 56 (4), pp 634–646.

<https://doi.org/10.1021/acs.biochem.6b00839>

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

The reactive center loop insertion in α -1-antitrypsin captured by accelerated molecular dynamics simulation

Journal:	<i>Biochemistry</i>
Manuscript ID	bi-2016-00839z.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Andersen, Ole; Aarhus University, Department of Chemistry Risør, Michael; Aarhus University, Department of Molecular Biology and Genetics Poulsen, Emil; Aarhus University, Department of Molecular Biology and Genetics Nielsen, Niels; University of Aarhus, Interdisciplinary Nanoscience Center (iNANO) and Department of Chemistry Miao, Yinglong; Howard Hughes Medical Institute, Enghild, Jan; University of Aarhus, Molecular Biology Schiøtt, Birgit; Aarhus University, Department of Chemistry

SCHOLARONE™
Manuscripts

The reactive center loop insertion in α -1-antitrypsin captured by accelerated molecular dynamics simulation.

Funding information

This work was supported by grants from the Danish Council for Independent Research | Technology and Production Sciences (FTP 11-105010), the Danish National Research Foundation (DNRF59), the Program Commission on Strategic Growth Technologies, Innovation Fund Denmark (0603-00439B), the NSF (grant MCB1020765), the NIH (grant GM31749), the Howard Hughes Medical Institute, and the National Biomedical Computation Resource. Computations were possible mainly through allocations of time at the Centre for Scientific Computing Aarhus. Part of the initial research was carried out on the Triton Shared Computing Cluster and Gordon supercomputer (XSEDE award project TG-MCA93S013) at the San Diego Supercomputer Center. O.J.A. was supported by a grant from the Novo Scholarship Program and M.W.R. was supported by a travel grant from the Carlsberg foundation.

Ole Juul Andersen,^{†§¶} Michael Wulff Risør,^{†‡¶} Emil Christian Poulsen,^{†‡} Niels Chr. Nielsen,^{†§} Yinglong Miao,^Δ Jan J. Enghild,^{†‡} and Birgit Schiøtt^{†§}*

[†]Center for Insoluble Protein Structures (inSPIN) and the Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark

[§]Department of Chemistry, Aarhus University, Aarhus, Denmark

[‡]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

^ΔHoward Hughes Medical Institute and Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093, United States

AUTHOR INFORMATION

*Corresponding author: Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, DK, Tel.: +45 8715 5975, Fax: +45 8619 6199, E-mail: birgit@chem.au.dk

[†]Both authors contributed equally and should be considered as first authors.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABBREVIATIONS AND TEXTUAL FOOTNOTES

α_1 AT, α -1-antitrypsin; AATD, α -1-antitrypsin deficiency; aMD, accelerated molecular dynamics; CD, circular dichroism; cMD, conventional molecular dynamics; hF, helix F; hI, helix I; HNE, human neutrophil elastase; MD, molecular dynamics; PDB, protein data bank; PPE, porcine pancreatic elastase; RCL, reactive center loop; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SE, standard error of the mean; SI, stoichiometry of inhibition; thFs3A, turn connecting helix F with strand 3A; WT, wild type.

ABSTRACT

Protease inhibition by metastable serine protease inhibitors (serpins) is mediated by one of the largest functional intra-domain conformational changes known in biology. In this extensive structural rearrangement, protease-serpin complex formation triggers cleavage of the serpin reactive center loop (RCL), its subsequent insertion into the central β -sheet A, and covalent trapping of the target protease. In this study, we present the first detailed accelerated molecular dynamics simulation of the insertion of the fully cleaved RCL in α_1 -antitrypsin (α_1 AT), the archetypal member of the human serpins. Our results revealed internal water pathways that enable the initial incorporation of RCL residue side chains into the protein interior. We observed structural plasticity of the helix F (hF) element that blocks the RCL path in the native state which is in excellent agreement with previous experimental reports. Furthermore, the simulation suggested a novel role of hF and the connected turn (thFs3A) as chaperones that support the insertion process by reducing the conformational space available to the RCL. Transient electrostatic interactions of RCL residues potentially fine-tune the serpin inhibitory activity. Based on our simulation, we generated the α_1 AT mutants K168E, E346K and K168E/E346K, and analyzed their inhibitory activity along with their intrinsic stability and heat-induced polymerization. Remarkably, the E346K mutation exhibited enhanced inhibitory activity along with an increased rate of premature structural collapse (polymerization), suggesting a significant role of E346 in gate-keeping the strain in the metastable native state.

INTRODUCTION

The inhibitory members of the serpin superfamily are effective regulators of proteolytic cascades through an evolutionarily conserved suicide substrate mechanism that leads to the covalent trapping and structural distortion of the target proteases.¹ Key to this process is a native metastable (stressed) state that upon protease binding and acyl-enzyme formation at the reactive center loop (RCL) (Figure 1A) structurally rearranges to a thermodynamically more favored (relaxed) state, during which the RCL inserts into the central β -sheet A and the target protease translocates by more than 70 Å to the distal end of the serpin (Figure 1B).² Experimental and clinical evidence strongly indicate that the intrinsic metastability of the native state entails mutational vulnerability, as a high number of substitutions causes misfolding, premature structural collapse, and self-polymerization with resulting deficiency diseases collectively known as serpinopathies.^{1,3} The most prevalent of the serpinopathies is α -1-antitrypsin deficiency (AATD), where 1 in 2,000 humans carries a single mutation (E342K, referred to as Z- α_1 AT) in α_1 AT. The E342K mutation results in an accumulation of polymeric protein inclusions in the endoplasmic reticulum of hepatocytes leading to blood deficiency.⁴ α_1 AT is the archetypal member of the serpin superfamily and protects the lungs against excessive human neutrophil elastase (HNE) activity during inflammatory bursts.⁵⁻⁷ The α_1 AT serpin α/β -fold is comprised of three β -sheets (A-C) and nine α -helices (A-I) with residues 344-362 forming the RCL (P15-P4' in the protease-substrate nomenclature of Schechter and Berger)⁸ and encompassing the protease recognition site (Figure 1A). Upon cleavage of the P1-P1' scissile bond, the kinetic trap is released, and the RCL initiates its insertion into the breach region (top of β -sheet A) by a zipper-like motion⁹ of

the proximal hinge (residues 342-350). In fact, such a partial insertion is seen in the native inactivated form of antithrombin that requires interaction with heparin to reach full exposure of the RCL for protease binding.¹ The subsequent full insertion of the RCL to form β -strand s4A involves the gradual opening of β -sheet A and displacement of helix F (hF) that blocks the free passage in the native serpin structure, although the nature of this rearrangement has remained elusive. Release of the native strain by RCL cleavage and insertion happens on a timescale sufficiently short to irreversibly inhibit the target protease before the deacylation step occurs. The delicate kinetic tuning of this multi-step process controls the partition between stable serpin-enzyme complexes and cleaved serpins and has been described in numerous mutational and biophysical studies.^{2,10-16}

The framework for investigating structural features of metastability, inhibitory action, intermediate states, and self-polymerization is based on native, cleaved and protease-complexed α_1 AT crystal structures and more recent data from hydrogen-deuterium exchange mass spectrometry, and NMR spectroscopy.¹⁷⁻²⁵ However, as the rate of loop insertion and protease translocation is on the order of 1,000 per second, structural and dynamic information on the individual steps of the insertion is largely missing due to obvious experimental challenges.²⁶ Molecular dynamics (MD) simulations supply the theoretical tools to gauge molecular transitions at an atomic level but require enhanced sampling algorithms to reach the 1 ms timescale of the large intramolecular serpin reorganization.

While two previous studies have used MD techniques to investigate the non-inhibitory latency transition of serpins,^{27,28} herein we present the first fully atomistic description of the insertion of the fully cleaved RCL of α_1 AT by means of an accelerated molecular

dynamics²⁹ (aMD) simulation. The simulation provides clues to the role of various structural elements in the insertion process and the concatenated mechanical events that are triggered by cleavage of the P1-P1' bond. In particular, we discuss the initial opening of β -sheet A in the breach region, the pair-wise electrostatic interactions throughout the insertion process, the roles of various conserved polar residues in the serpin family, and a novel role of hF as a chaperone for RCL insertion. We further present biochemical data on the charge-reversal mutants K168E, E346K, and K168E/E346K, which were constructed to analyze the functional role of two highly conserved charged residues (K168 and E346) that potentially act as electrostatic modulators of the RCL insertion.

Figure 1. Crystal structures of the (A) stressed and (B) relaxed states of α_1 AT in complex with bovine trypsin (gray surface) [Protein Data Bank (PDB) entries 1OPH and 1EZK, respectively].^{19,30} The RCL including the proximal hinge is shown in orange.

EXPERIMENTAL PROCEDURES

Modeling. Simulations were performed on a crystal structure of the uncomplexed stressed state of α_1 AT [Protein Data Bank (PDB) entry 3NE4]⁷ with the following modifications: (i) the crystal structure lacks a stretch of 23 N-terminal residues and the C-terminal K394. The N- and C-termini were therefore capped with an acetyl group and an N-methylamide group, respectively, to mimic the neutral nature of a continuous peptide chain; (ii) the crystal structure contains two rotameric states with 50% occupancy of V216 and I340. Neither of the two states could be identified as being more favorable than the other, so state A was chosen for both residues; (iii) the scissile bond between M358 and S359 in the RCL was cleaved *in silico* and the M358 residue was capped with an N-methylamide group to mimic the neutral nature of the acyl-enzyme intermediate linkage, while the S359 residue was capped with a positively charged amine group. The simulations were performed without a protease attached to the RCL to speed up the calculations by reducing the number of atoms in the system. Including the protease would increase the system size about 2.5-fold rendering it infeasible to perform the simulation with the computational resources currently available to the general research community.

Protonation states of ionizable residues were assigned using PROPKA3.0.³¹ All histidines were modeled as neutral. H93, H231, and H262 were modeled as the N ϵ tautomer, while the rest were modeled as the N δ tautomer. The hydrogen bond network was optimized using the Protein Preparation Wizard^{32,33} included in the Schrödinger Suite version 2013-2. Energy minimization was then performed to relax the protein structure, during which the heavy atoms were constrained to a root-mean-square deviation (RMSD) of 0.3 Å relative to the crystal structure. The system was solvated using the CHARMM variant of the TIP3P water model^{34,35} (commonly referred to as TIPS3P) in a cubic box with sides

of length 110 Å, corresponding to a minimum spacing of 20 Å between the protein and the edge of the box. This yielded a system of ~133,000 atoms. NaCl was added to a concentration of 0.15 M and the system was neutralized.

Conventional MD simulation. As part of this study, we performed a 2 μs conventional MD (cMD) simulation in which we observed only insertion of the first RCL residue, T345 (system and simulation setup provided in Supporting Information). Together with the estimated rate constant ($\sim 1,000 \text{ s}^{-1}$),²⁶ this showed the need for applying an enhanced sampling technique. To this end, aMD that provides unconstrained enhanced sampling without the need to set predefined reaction coordinates was chosen as an appropriate method for investigating the RCL insertion process. The aMD simulation was executed in a manually enforced ratchet-and-pawl^{27,36} type fashion in which the simulation was restarted from a previous checkpoint if an RCL residue was observed to leave β-sheet A. By selectively extending the simulation in this manner we favored exploration of a conformational space in which the RCL did not leave β-sheet A as described in detail in the following section.

Accelerated MD simulation. The NAMD2.9³⁷ package was used for simulation with the CHARMM36 force field parameters.³⁸⁻⁴⁰ The temperature was kept constant at 310 K by Langevin dynamics with a damping coefficient of 2 ps^{-1} . The pressure was maintained at 1 atm using the Nosé-Hoover Langevin piston method^{41,42} with a period of 200 fs and a decay of 100 fs. The Particle Mesh Ewald method⁴³ was used to treat full electrostatics, while van der Waals interactions were truncated at a cut-off distance of 12 Å using a switching function from 10 Å. Bonds to hydrogen atoms were constrained using SHAKE⁴⁴ enabling a 2 fs time step, with non-bonded interactions being updated every 2

fs, and full electrostatics updated every 4 fs. Initial system minimization for 2,000 steps was performed four times, fixing first all the protein atoms, then the backbone atoms, then all C α atoms, and finally with all atoms moving freely. System equilibration was performed by heating from 1 K to 310 K for 5 ns at a pressure of 1 atm constraining all C α atoms, simulating for 5 ns constraining all C α atoms, and final 5 ns of unconstrained equilibration.

The aMD simulation was performed using the dual-boost,²⁹ in which a bias is applied to both the dihedral energetic terms and the total potential energy of the system. This effectively decreases the energy barriers connecting different conformational states of the protein, without altering positions of the local energy minima. The bias is applied according to:

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \frac{(E - V(r))^2}{\alpha + (E - V(r))}, & V(r) < E \end{cases} \quad (1)$$

where $\Delta V(r)$ is the bias potential, $V(r)$ is the original potential, E is the threshold energy, and α is an acceleration parameter. The values of E and α were calculated according to Miao et al.⁴⁵:

$$E_{dihed} = \langle V_{dihed} \rangle + a_1 \cdot N_{res}, \quad \alpha_{dihed} = a_2 \cdot N_{res}/5 \quad (2)$$

$$E_{total} = \langle V_{total} \rangle + b_1 \cdot N_{atoms}, \quad \alpha_{total} = b_2 \cdot N_{atoms} \quad (3)$$

where N_{res} is the number of protein residues, N_{atoms} is the total number of atoms, and $\langle V_{dihed} \rangle$ and $\langle V_{total} \rangle$ are the average dihedral and total potential energies of the system,

1
2
3 which we obtained from a 20 ns unbiased MD simulation of the starting structure. Miao
4
5 et al. identified $a_1=a_2=3.5$ and $b_1=b_2=0.175$ as optimal values for accelerating protein
6
7 folding for two of their four studied fast-folding proteins.⁴⁵ These values were chosen for
8
9 the current study with the exception of b_2 , which was inadvertently set to 0.035.
10
11 Decreasing b_2 corresponds to increasing the boost applied to the total potential, leading to
12
13 increased protein dynamics, increased solvent diffusion, and weakening of non-bonded
14
15 interactions. The increased total boost potential caused by a lower b_2 is within the
16
17 acceptable range for the studied system because several highly stable non-bonded
18
19 interactions were observed throughout the aMD simulation and the cMD simulation
20
21 backed our findings on the functional role of water in the RCL insertion mechanism.
22
23
24
25
26

27
28 A total length of 700 ns aMD simulation that combines continuous 20 ns extension
29
30 trajectories was obtained. An extension was aborted if the RCL was seen to move out of
31
32 β -sheet A. In this case, up to ten parallel aMD runs were performed from the coordinates
33
34 and velocities stored at the previous checkpoint. The aMD run visually showing the
35
36 greatest promise for further RCL insertion was selected for further extension. This
37
38 procedure was applied five times during the entire simulation at 340 ns, 460 ns, 480 ns,
39
40 500 ns and 680 ns. The divergence of the parallel runs originated from the stochastic
41
42 nature of the thermostat and the barostat. Performing the simulation in this ratchet-and-
43
44 pawl^{27,36} type fashion does not force the system toward the RCL-inserted state, but
45
46 prevents it from backtracking toward the initial state. We note that release of T345 from
47
48 β -sheet A was observed in the final part of the cMD simulation; however, it is our belief
49
50 that while an insert-release-insert scenario is definitely possible, it is not an essential part
51
52 of the inhibitory mechanism. It is important to note that the timescale on which the events
53
54
55
56
57
58
59
60

in the simulation occur does not correspond to that of the biological timescale, as the aMD boost potential greatly accelerates the insertion process.

Computational analyses. All computational analyses were performed with VMD1.9.1⁴⁶ on the 700 ns trajectory containing snapshots saved every 0.5 ns. Secondary structure analysis of the snapshots was performed using the Timeline plugin in VMD1.9.1.⁴⁶ Residues were designated as having β -structure if they were part of either an extended (E) or isolated bridge (B) segment. RMSD and root-mean-square fluctuation (RMSF) analyses were performed on protein C α atoms. All residues except T345-M358 were used for alignment to the fully inserted crystal structure of α_1 AT (PDB entry 1EZ X).¹⁹ Ionic interactions were counted as an electrostatic contact if the distance between the ionizable groups was below 5 Å. The distance was measured using the carbon atom in the carboxyl group of Asp and Glu residues, the side chain nitrogen in Lys residues, and the central carbon atom in the guanidinium group of Arg residues.

Production of α_1 AT WT and mutants. The α_1 AT coding sequence was amplified from human liver cDNA and inserted into the pET-M11 vector by directional cloning using Nco1/Not1 restriction sites. Mutants were constructed by site-directed mutagenesis. See Supporting Information for details on expression and purification. Prior to the kinetic and functional analyses, each variant was purified by size-exclusion chromatography on a Superose12 column (equilibrated in PBS) to ensure a true monomeric and native α_1 AT pool, validated by running a 10% reducing non-denaturing PAGE.

Analysis of heat-induced α_1 AT polymerization. Polymerization of the α_1 AT variants was performed by incubation at 55 °C for 0, 15, 30 or 45 min at a protein concentration

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

of 0.2 mg/mL in PBS. The presence of polymers was analyzed by reducing non-denaturing polyacrylamide gel electrophoresis using 10% polyacrylamide gels. Protein bands were visualized with Coomassie Brilliant Blue. The residual monomer amount for each time point was quantified by densitometry and normalized to the zero hour value for each variant. Polymerization rates were determined by linear regression analysis on the disappeared monomer fraction at each time point as stated in the Supporting Information.

SDS-evaluation of inhibitor:protease complex formation. α_1 AT variants at 0.2 mg/mL in PBS were incubated at 0.1, 0.2 and 0.4 molar ratios of inhibitor:PPE (porcine pancreatic elastase) for 10 min at 37° C. Complex formation and amount of cleaved α_1 AT were evaluated by SDS-PAGE using 10% polyacrylamide gels. For each sample, SDS-sample buffer was added, followed by immediate boiling to quench the reaction. The bands corresponding to cleaved and complexed serpin, and control and residual PPE were quantified by densitometry for the 0.2 and 0.4 molar ratio reactions.

Stoichiometry of inhibition toward PPE. The concentration of PPE was estimated by UV280 absorption and assumed to be 100% active. SI was determined by incubating α_1 AT variants at concentrations of 0, 10, 20, 30 and 50 nM with 100 nM PPE and assay buffer (50 mM phosphate, 137 mM NaCl, 0.05% Triton X-100). Each reaction was set up in triplicate in a total volume of 100 μ L using a 96-well plate. After incubation for 10 min at RT and 10 min at 37 °C, 50 μ L of assay buffer and 50 μ L of 1 mM N-Succinyl-(Ala)₃-*p*-nitroanilide were added to each well. The residual protease activity was examined by observing the rate of *p*-nitroaniline release per hour at 405 nm using a Fluostar Omega plate reader. A linear fit of the residual protease activity as a function of inhibitor concentration was used to estimate the stoichiometry of inhibition (SI) for each

variant. Relative activities were calculated as $SI_{WT}/SI_{variant}$ using data from eight independent experiments.

Stoichiometry of inhibition toward trypsin. The active trypsin concentration was measured by active site titration in 20 mM Hepes (pH 8.0) using 4-nitrophenyl 4-guanidinobenzoate hydrochloride (pNPGB, Sigma). SI was determined by incubating WT α_1 AT at concentrations of 0, 2, 4, 6, 9, nM with 20 nM Trypsin in assay buffer (50 mM phosphate, 137 mM NaCl, 0.05% Triton X-100). Each reaction was set up in triplicate in a total volume of 100 μ L using a 96-well plate. After incubation for 10 min at RT and 10 min at 37 °C, 50 μ L of assay buffer and 50 μ L of 0.5 mM H-D-PFR-*p*-nitroanilide (Kabi Diagnostica) were added to each well. The residual protease activity and SI were determined as indicated for PPE above using data from ten independent experiments.

Thermal denaturation by circular dichroism. Circular dichroism (CD) spectroscopy was performed on a Jasco J-800 spectropolarimeter using a 0.1 cm quartz cuvette and a protein concentration of 0.2 mg/mL in PBS containing 1 mM dithiothreitol (DTT). Thermal denaturation was monitored at 222 nm from 20-95 °C with a gradient of 1 °C/min. Each variant was assayed in triplicates, and the midpoint transition temperature (T_m) was extracted by fitting the triplicate mean data to a two-state unfolding model using KaleidaGraph 4.1 (Synergy Software).

Inhibitory human serpins and alignment details. Multiple sequence alignment of the 27 known inhibitory human serpins was carried out using CLC Main Workbench 6.5.⁴⁷ The alignment tool used a progressive algorithm with a gap open cost of 10 and a gap

extension cost of 1. The 27 inhibitory serpins (included) and the non-inhibitory or uncharacterized (excluded) serpins are listed in Table S1.

RESULTS AND DISCUSSION

RCL insertion captured by aMD. Insertion of the α_1 AT RCL into β -sheet A was simulated by aMD using an RCL-cleaved version of the native α_1 AT structure (see experimental procedures). The simulation captured the RCL insertion up to F352 during a total simulation time of 700 ns. Figure 2 shows a selection of 15 equally spaced snapshots from the simulation. Full RCL insertion up to M358 was not obtained because the final stretch of the RCL (L353-M358) formed a strong network of salt bridges with residues in β -sheet A and the turn connecting the top of hF with s3A (thFs3A) (Figure 3). Further extension of the simulation to 900 ns did not result in additional RCL insertion due to the stability of the many ionic interactions. Supporting Information includes three animations that show i) the insertion of the RCL as seen from the front of α_1 AT (Movie S1), ii) a detailed view of specific events occurring during the insertion process (Movie S2) and iii), the unfolding/folding of helix F during the time window from 400 to 500 ns (Movie S3). In Movies S1-S2 the visualization of the protein secondary structure is updated for only the RCL and the unfolding of hF is therefore not reflected. We verified that our simulation of α_1 AT progressed from the stressed towards the relaxed state by calculating the difference in RMSD to either state as a function of simulation time (Figure S1).

Figure 2. Snapshots from the aMD simulation showing RCL insertion up to F352. The RCL is shown in orange and hF in blue. s3A and s5A are shown in lime to aid the visualization of the separation of β -sheet A. Extending the simulation to 900 ns did not

lead to further RCL insertion. The full animation (Movie S1) is provided in Supporting Information.

Figure 3. The RCL conformation as seen in the final snapshot of the simulation. E175 and D177 form salt bridges with K331, and E354 forms salt bridges with K174 and K335.

Tracking mobile elements with transient structural plasticity. To quantify the flexibility of the structural elements in α_1 AT we calculated the per residue RMSD with respect to the fully inserted crystal structure as a function of time (Figure 4A) together with the RMSF (Figure 4B). Helix F clearly stood out as the most flexible structural element second only to the RCL. This is in excellent agreement with hydrogen-deuterium exchange studies and kinetic measurements particularly highlighting the N- and C-terminal parts as inherently flexible.^{25,48,49} The plastic behavior of hF was further exemplified by its large downward movement that accompanied the insertion of A348-

G349 in the aMD simulation (Figure 2, 350-450 ns). Based on fluorescence analysis of the equilibrium unfolding behavior of a Y160W (hF) variant of α_1 AT, hF has been suggested to undergo an unfolding/folding reaction during translocation of the protease.⁵⁰ Such behavior is consistent with the simulation presented here, in which the structural integrity of the helix was compromised during a change in the position of the free part of the RCL (Movie S3 and Figure 2, 450 ns).

The flexibility analysis of the aMD simulation also highlighted inherently rigid elements during the RCL insertion process, including helix I (hI) that exhibited low RMSF (~ 1.5 -2 Å) and small RMSD values. The structural rigidity of this element at the base of the protein, as captured by our simulation, has been suggested to be important for the inhibitory mechanism.⁵¹ In contrast, serpin polymerization, which also involves structural transition of the RCL, is suggested to proceed through population of a partially unfolded flexible intermediate state. Events such as loss of β -sheet A integrity,⁵² partial hI unfolding,^{51,52} and loosening of β -sheet C⁵³ are observed depending on conditions for polymer formation. None of these events were present during the aMD simulation; therefore, our results could indicate that α_1 AT does not need to go through the polymerization-prone intermediate to achieve functional RCL insertion. Due to our ratchet-and-pawl simulation approach and the lack of full RCL insertion, we cannot rule out the existence of an alternative RCL insertion pathway that may involve more elaborate insert-release-insert events.

Figure 4. (A) Heatmap showing the C α RMSD as a function of time for each individual residue when compared to the fully inserted crystal structure (PDB entry 1EZ X).¹⁹ To aid interpretation, selected helices and the RCL have been denoted on the right and enclosed by dashed lines in the heatmap. The residues spanned by β -sheet A are: 140-146 (s1A), 109-121 (s2A), 181-194 (s3A), 326-342 (s5A) and 290-299 (s6A). (B) The fully inserted crystal structure (PDB entry 1EZ X)¹⁹ colored according to the C α RMSF.

Insertion of the proximal hinge into the breach region. The RCL is converted from a loop structure to a β -strand of β -sheet A with one face packed against the core of α_1 AT during the insertion process. This entails that the side chain of every other residue in the RCL is buried upon complete insertion (Figure 5A). The first residue side chain to insert into the breach region is T345. Based on crystal structures of the stressed native state, water molecules are found to occupy the T345-accepting pocket.⁷ These water molecules hinder RCL insertion and must be displaced by the inserting T345 side chain blocking their exit to the exterior (Figure 5B). Interestingly, our simulation data revealed that the hindrance may be resolved by an intricate network of water pathways connecting a series of small pockets in the protein transiently occupied by water throughout the simulation (Figure 5C). T345 can insert into the pocket as transient water molecules exit through

pathways running behind β -sheet A. Both the aMD and the cMD simulations support this observation. The same scenario is seen upon insertion of the next large side chain, M351. Internal water pathways appear to be a key feature for allowing side chain insertion into the protein core, and hence for the inhibitory mechanism of serpins in general.

Figure 5. (A) Structure of the fully inserted RCL seen from the side, shown together with hF (PDB entry 1EZ_X).¹⁹ (B) Snapshot from the aMD simulation. Water molecules (red spheres) occupy the pocket where T345 is meant to insert, seemingly hindering insertion. (C) Examples of water pathways (red dotted lines) observed in the simulation. The pathways connect a series of pockets transiently populated during the simulation by water molecules (red spheres).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Helix F acts as a chaperone for RCL insertion - Interactions at the shutter. The functional role of hF has been a subject of much speculation.^{10,48,50,54-56} At first sight, hF seems to sterically block RCL insertion by its position in front of β -sheet A (Figure 1A). Here, it may protect serpins against non-functional loop-sheet polymerization by obstructing the *trans* insertion of an RCL from a second serpin.⁴⁸ Another proposal suggests that hF acts as a spring for storing energy when pushed aside during RCL insertion.⁵⁴ In this way, the subsequent release of hF back to its position in front of β -sheet A may be responsible for the active site distortion in the protease as the protease is pushed into its final position.⁵⁴ It is not possible to support or dispute these ideas based on the aMD simulation, because there is only a single α_1 AT and no protease present. We do, however, observe events that hint at another possible role of hF.

During the simulated RCL insertion, thFs3A was observed to interact in several ways with the RCL. An illustrative example is the 250 ns snapshot (Figure 2) in which a stretch of the RCL is seen to partake in an anti-parallel β -sheet interaction with thFs3A. The presence of β -structure was monitored in the RCL and thFs3A throughout the simulation to quantify the interactions (Figure 6). We see that the insertion of E346 occurs at the same time as the backbone of the RCL reaches the top of hF (~200-240 ns). Shortly after, a β -sheet interaction is created between I169 and F352 (P7) and is stable from ~240-450 ns. This interaction is broken when the RCL flips to the other side of hF (Figure 2, 450 ns). Next, an interaction is created between I169 and M351 (P8), and this remains stable until M351 is fully inserted into β -sheet A. The conservation of non-polar residues at P7 and P8 in serpins supports the role of such a hydrophobic interaction. The interaction reduces the conformational space readily available to the RCL. Only a short segment of

the RCL, the part that is about to insert, is left unrestricted for conformational sampling. Therefore, it may be hypothesized that the conformation needed for further insertion of residues in the shutter region may be reached faster. Based on these observations, we propose that hF and thFs3A act together as chaperones that speed up the insertion process. This is in contrast to the intuitive assumption that hF slows down the insertion by blocking access to β -sheet A.

The ability of hF to interact with the RCL during the insertion is aided by its flexibility. In particular, the insertion of A348-G349 is accompanied by a large downward movement of hF (Figure 2, 350-450 ns). Kinetic data on a disulfide-locked variant between K168C at the top of thFs3A and F189C at the top of s3A seem to suggest that such flexibility is not necessary for the serpin inhibitory function against porcine pancreatic elastase (PPE).¹⁰ However, a more recent study of the K168C-F189C double mutant found an eight-fold increase of the stoichiometry of inhibition (SI) against bovine α -chymotrypsin in the presence of the disulfide bond.⁵⁵ The SI reflects the partition between cleaved serpins and formation of stable serpin-enzyme complexes, assuming insertion and protease translocation is necessary for trapping of the acyl-enzyme intermediate.^{11,17,57} Consequently, the reported increase in SI of the K168C-F189C disulfide-locked mutant suggests retardation of loop insertion by which a larger fraction of the serpins acts as mere substrates. Therefore, locking hF to s3A seems to have a significant effect on the inhibitory mechanism. Based on our observations in the aMD simulation, we suggest that flexibility is an essential and intrinsic feature of hF/thFs3A.

Figure 6. β -structure observed in thFs3A residues (blue) and RCL residues (orange). The black dots for residues T345-F352 indicate the point of complete insertion into β -sheet A. The presence of β -structure for residues E354-P357 in the final part of the simulation is due to the interactions with thFs3A described in Figure 3. A stable interaction is observed between I169 and F352 during ~240-450 ns, after which the RCL flips to the other side of hF (Figure 2), and I169 participates in a β -sheet interaction with M351 until this inserts fully into β -sheet A.

Electrostatic contacts formed during the RCL insertion. Two acidic residues (E346, E354) within the RCL have the potential to engage in electrostatic interactions along the insertion pathway. The aMD simulation of the near-complete RCL insertion presented here allowed us to extract information on the transient electrostatic contacts formed by these two residues (Figure 7A). Additionally, contacts between charged residues of hF/thFs3A and the rest of the protein were also extracted, as shown in Figure 7B (basic residues) and Figure 7C (acidic residues).

The first electrostatic contact to be formed during insertion of the proximal hinge is between E346 and K168 at the top of thFs3A (Figure 8A and Movie S2). The same interaction is also observed in the 2 μ s cMD simulation. This contact may facilitate the downward movement of the RCL by exerting a pulling force on the E346 backbone. The interaction is maintained until the backbone of the RCL comes into contact with the backbone of thFs3A (Figure 8B and Movie S2). The breakage of the K168-E346 interaction is directly followed by a reorientation of E346 with subsequent insertion (Figure 8C and Movie S2). After insertion, E346 engages in an electrostatic interaction with K191 for the remainder of the simulation. This is not surprising because the two residues are located next to each other in s4A and s3A, respectively.

Figure 7. Overview of salt bridge interactions. Salt bridges involving RCL residues are colored orange. Charged hF/thFs3A residues that do not form salt bridges with non-hF/thFs3A residues have been omitted for clarity. (A) Salt bridges between acidic RCL residues and the rest of the protein. (B) Salt bridges between basic hF/thFs3A residues and the rest of the protein. (C) Salt bridges between acidic hF/thFs3A residues and the rest of the protein.

Figure 8. Salt bridges formed by E346. (A) A salt bridge is initially formed between E346 and K168. (B) The salt bridge between E346 and K168 is maintained until the backbone of the RCL reaches thFs3A. (C) After E346 reorients to insert into β -sheet A, it forms a stable salt bridge with the neighboring K191. K193 and K343 are also within reach of E346.

The electrostatic interaction pattern of the second Glu of the RCL, E354, is slightly more intricate (Figure 7A). In the beginning, the side chain of E354 is placed on top of the protein making electrostatic contacts to R196 and R223. Based on kinetic measurements and mutational studies, these interactions are considered important for regulating the rate of protease inhibition by influencing the formation of the encounter complex with the

target protease.^{11,58} As the salt bridges break and the RCL starts its downward movement, E354 travels past K343 and K193. It then reaches K168 in thFs3A. This salt bridge is present from ~220-450 ns, the same time frame in which a β -sheet interaction is observed between I169 and F352 (Figure 6, Figure 7, Figure 9A and Movie S2). Interestingly, mutation of I169 to the smaller Val residue is linked to a decrease in the α_1 AT inhibitory function supporting the possible importance of the hydrophobic contacts to F352, although the primary suggested role of this well-conserved residue has been ascribed to the hydrophobic environment in the native state.⁵⁹

In the final part of the simulation, the RCL rearranges and electrostatic contacts are formed between E354, K174, and K335. Upon insertion of M351 after ~650 ns, the RCL reaches the final state as shown in Figure 3 (Movie S2).

Figure 9. (A) A salt bridge between E354 and K168 is present from ~220-450 ns. At the same time, a β -sheet interaction is present between I169 and F352, stabilized by the interaction between the two hydrophobic side chains. Together, these two interactions form a stable interface between the RCL and thFs3A. (B) Residues involved in salt bridges between thFs3A and the body of the protein (PDB entry 3NE4).⁷

The most noteworthy electrostatic contacts formed by basic hF/thFs3A residues, besides those already mentioned, are from K168 to E141 (located at the top of s1A) towards the end of the simulation, and from E122 to R178 (located at the bottom of s2A and at the bottom of thFs3A, respectively). For the acidic hF/thFs3A residues (D171, E175, D177, and D179), a number of contacts are observed to residues in helix I (K300) and s5A (K328, K331, and K335) (Figure 9B). These contacts are present throughout most of the simulation. All must be transiently broken for RCL insertion to occur and constitute an energy barrier for the insertion process. However, it has been shown that the D171N mutation does not significantly affect the inhibitory activity, indicating that the interaction between D171 and K335 is easily broken and the related energy barrier is of negligible size.⁵⁹

Multiple charged residues along the insertion pathway are largely conserved. The serpin fold is one of the most conserved complex protein structures throughout evolution although the pairwise identity can be lower than 25%.⁶⁰ The high degree of individual residue conservation relates to the delicate folding landscape leading to the metastable state, the hydrophobic core packing, and the inherent flexibility of the structural elements that allows for loop insertion and acquisition of the stable, relaxed state.^{23,61} In the region for loop insertion, several positions display >70% conservation among 219 serpin sequences, including proximal hinge residues E342, G344, A347, breach residues G192, W194, F198, Y244, as well as shutter residues L184, N186, H334 (β -sheet A) and I157, N158, V161, T165, G167 (hF/thFs3A).⁶¹ Of these, most are involved in hydrophobic packing motifs and/or hydrogen bonding patterns involving β -sheet A, RCL and hF. Mutations herein alter native state metastability, the folding pathway or inhibitory

function, which are not *per se* interdependent entities.^{48,62-66} Slight sequence variations from one serpin to another fine-tune this machinery, necessitating a deeper understanding of the molecular transitions and the forces controlling them.^{11,67} For α_1 AT, we sought to elaborate on the importance of the electrostatic contacts formed during loop insertion, as observed in the presented simulation. Several of the charged residues are largely conserved between the various clades of human serpins, including the strictly conserved E342 but also E346 (proximal hinge), K335 (s5A), K191 (s3A) and K168 (thFs3A) with a cut-off at 80% (Figure 10). Of these residues, the role of E342 has been studied intensely because the α_1 AT Z-variant (an E342K substitution) is the most common pathological disease mutation causing AATD.⁶⁸ This Glu residue is a gate-keeper of the native state by forming an electrostatic contact to K290 keeping the breach region closed and preventing premature opening of β -sheet A.⁶⁹ In fact, this position, along with E346 and K335, stand out as the slowest evolving charged residues along the insertion pathway⁷⁰ and represent positive functional selection. The role of K335 has been studied extensively through a number of residue substitutions hinting at a critical role in serpin metastability.^{59,71} This polar residue is buried in the native state and decreases serpin stability, which at this position is critical for the kinetics of loop insertion. Slowing the insertion rate down by stabilizing substitutions at K335 alters the partition between substrate and inhibitor pathways and results in decreased efficiency in forming the covalently trapped serpin-enzyme complexes.⁷¹

The role of E346 (P13) has, to our knowledge, not been explored in detail for α_1 AT, nor has the importance of the electrostatic contacts suggestively formed during the RCL insertion.

Figure 10. Conservation and position of charged residues along the insertion pathway of the RCL. (A) Conservation of charge polarity (R/K or D/E) at selected residues (α_1 AT numbering) by alignment of the 27 inhibitory human serpin sequences from the clades A to I (Table S1). A score >80% is colored red (negative charge) or blue (positive charge). Connecting lines indicate the major electrostatic pairings of the conserved charged residues. The red bar indicates the transient contact formed between K168 and E346, as observed in the aMD simulation. (B) Indication of the position of the charged residues along the insertion pathway of α_1 AT. Negatively charged residues <80% or >80% conservation are represented by light red spheres and red spheres, respectively. Positively charged residues are represented by light blue or blue, accordingly. The scissile bond P1 Met is colored orange.

Consequences of charge substitution at K168 and E346 for α_1 AT. We prepared constructs for the α_1 AT single and double mutants K168E, E346K, and K168E/E346K to investigate the role of E346 and the electrostatic contact potentially formed to the thFs3A

residue K168. Each variant was recombinantly expressed in *E. coli*, purified and isolated by size exclusion chromatography before every analysis to ensure a completely monomeric preparation of the correctly folded metastable serpin. We hypothesized that a charge reversal of either residue would impede loop insertion whereas the double mutant would be functionally rescued regarding that particular electrostatic interaction. The stability of each variant was assessed by thermal denaturation (Figure S2), resulting in estimates of midpoint transition temperatures (T_m) as indicated in Table 1. All mutants were found to have T_m values within 0.7 °C of WT, indicating that the residue substitutions did not cause a major change in thermal unfolding. K168E T_m was 0.5 °C higher, and E346K T_m was 0.7 °C lower. K168E/E346K was 0.2 °C lower but was not significant. Although the mutational alterations to the electrostatic surface only had a small effect on the native state stability, the energy landscape for the structural transitions of α_1 AT could still be affected. Such a relationship has been observed for another RCL substitution, the Z-variant of α_1 AT (E342K). Here, the mutation caused only a small difference in thermodynamic stability but lowered the kinetic barrier for formation of a polymerogenic intermediate.⁷²

Table 1. Stability, activity and polymerization rates of α_1 AT variants K168E, E346K and K168E/E346K^a

Variant	T_m^b	Relative Activity ^c	$V_{poly} (\mu M/h)^d$
WT	58.6±0.2	1.00±0.01	3.27±0.16
K168E	59.1±0.1*	0.91±0.02*	2.82±0.25
E346K	57.9±0.1*	1.60±0.04*	6.67±0.37*
K168E/E346K	58.4±0.1	1.38±0.03*	6.17±0.24*

^a Asterisks indicate paired T-test p-values <0.05 for T_m and <0.001 for Relative Activity and V_{poly} compared to WT.

^b Thermal denaturation melting points (T_m) $\pm$ SE for WT and α_1 AT mutants assessed by CD spectroscopy at 222 nm (n=3) (Figure S2).

^c Relative activity $\pm$ SE reported as $SI_{WT}/SI_{variant}$ toward PPE (n=8). The absolute SI values $\pm$ standard deviation for WT was 1.1 ± 0.04 toward trypsin and 1.66 ± 0.14 toward PPE in agreement with reported SI values toward PPE $> 1.6^{62,73}$.

^d Polymerization rates $\pm$ SE were calculated by linear regression on the normalized monomer disappearance based on non-denaturing PAGE gels (n $\geq$ 3) after incubation at 55 °C for indicated times (Figure S3).

To evaluate the effect of the selected charge reversals on the overall α_1 AT inhibition process, we determined their ability to inhibit PPE relative to WT with kinetic assays for SI (Table 1). For K168E, the charge substitution resulted in a slightly decreased inhibition to a relative activity of 0.91. Both E346K and K168E/E346K displayed more potent inhibition with higher relative activities of 1.60 and 1.38, respectively. The E346K substitution thus significantly enhanced the inhibitory activity of α_1 AT.

We also investigated the amount of cleaved and complexed α_1 AT by SDS-PAGE after 10 min incubation at various enzyme:inhibitor ratios (Figure 11A). We used the gel band intensities to estimate SI values because the amount of cleaved and complexed serpin directly reflects the relationship between the rate constants for the substrate ($k_{substrate}$) and inhibitor ($k_{inhibitory}$) pathways. The E346K substitution led to an increased ratio of

complexed α_1 AT-PPE to cleaved α_1 AT compared to WT, whereas the K168E substitution resulted in a slight decrease. The K168E/E346K double mutant resembled the E346K substitution. Compared to WT, the relative activity estimates by gel analysis were 0.9, 1.5 and 1.5 for K168E, E346K and K168/E346K, respectively, which agrees with our kinetic determination and indicates that the mutations affect the rate of loop insertion.

The opening of β -sheet A facilitates not only loop insertion but also serpin polymerization, because all current models of the thermodynamically stable polymers involve a complete 6-stranded sheet A with an RCL inserted in either *cis* or *trans*.^{21,74,75} To assess the effect of charge substitutions at K168 and E346 on serpin polymerization, the variants were subjected to heat-induced polymerization at 55 °C and evaluated by native gel electrophoresis (Figure 11B and Figure S3). Substitution of the charges was clearly visible by the altered mobility of the natively folded monomeric α_1 AT. Of the four variants, the extent of polymerization was largest for the E346K mutant, followed by K168E/E346K, WT and K168E. Polymerization rates determined from time-resolved monomer loss corroborated these findings (Table 1 and Figure S3). WT and K168E had similar polymerization rates but the charge substitution at E346 caused an almost two-fold increase for the E346K and K168E/E346K mutants, suggesting the native E346 plays a role in attenuating polymerization.

Figure 11. Functional evaluation of α_1 AT variants K168E, E346K and K168E/E346K. (A) SDS-PAGE of WT and α_1 AT mutants after incubation for 10 min at indicated molar ratios with PPE. The gel band intensities of cleaved and complexed α_1 AT, corrected for protease intensity contribution, were used to determine gel-based SI values for the 0.2 and 0.4 molar ratio reactions, $SI = (I_{\text{complex}} + I_{\text{cleaved}}) / I_{\text{complex}}$. The average relative activity estimate ($SI_{\text{WT}} / SI_{\text{variant}}$) was 0.9 for K168E, 1.5 for E346K and 1.5 for K168E/E346K. (B) Reducing non-denaturing PAGE analysis showing the WT and the three mutants under native conditions (left) and after incubation for 30 minutes at 55 °C (right). Mobility shifts are consistent with the mutational change in overall pI for each variant. The extent of polymerization was increased for E346K and K168E/E346K (K+E), whereas K168E was similar to WT (see also Figure S3).

The role of charged residues in gate-keeping the native strain and inhibitory function. Serpins are evolutionary refined molecular machines tailored to exactly balance a delicate metastable native state with the risk of premature structural collapse on the one hand and proteolytic substrate cleavage of the RCL on the other. The gate-keepers involved in these transitions are many, of which both E342 and K335 are important members acting in opposing ways. Whereas K335 ensures metastability, E342 prevents premature collapse, as previously mentioned. The roles of K168 on the top of hF and E346 in the P13 position of the RCL are more elusive and represent fine-tuning of the kinetic machinery. Our results showed that the substitution of K168 to a negatively charged Glu slightly decreased the inhibitory activity of α_1 AT. This is in line with previous mutational studies that linked substitution of K168 (K168I, K168A) to an activity loss.⁵⁹ Based on our simulation results (Figure 8), the effect of the K168 charge reversal would compromise the electrostatic interaction to E346, but our mutational data on K168 and E346 indicates that this interaction is not critical for ensuring rapid loop insertion. In fact, loss of the negative charge displayed at the P13 position and substitution to a positive Lys residue clearly enhanced the rate of loop insertion and correlated to both an increase in polymerization rate and a slight decrease in thermal stability. The latter two effects have been observed for various substitutions of the charged E342 residue (P17) as well; however, this has been linked to a dramatic decrease in inhibitory activity of more than 40%^{69,76} and thus differs from the effect of the E346 residue discussed here. The importance of the P13 residue in regulating the rate of loop insertion has to our knowledge only been investigated for PAI-1 that has a Val at this

position. Here, latency transition is clearly decreased by a Val→Glu substitution,⁷⁷ supporting the role of a P13 Glu in gate-keeping the native metastable state.

Based on our experimental data, we suggest the electrostatic interaction between E346 and K168 is not a primary driver of loop insertion in α_1 AT in its complex formation with PPE. The double mutant K168E/E346K did not result in a WT-like appearance but largely retained the behavior of the E346K mutant. The dominant effect of the E346 substitution highlighted the role of this residue in maintaining a structural strain in the native conformation. It has previously been hypothesized that the salt bridge between K168 and E346 is important for the stability of the loop-inserted protease-serpin complex and hence for serpin inhibitory function.⁵⁹ By inference, this should have resulted in less efficient inhibition by the E346K mutant which was not the case in the performed experiments. In our simulation of loop insertion, another potential interaction of K168 with E354 was observed; however, as E354 is not a conserved residue across the serpin family this is unlikely to be an interaction of primary importance and was therefore not tested experimentally.

CONCLUSION

In this work, we have successfully applied aMD to gain insight into the insertion of the fully cleaved RCL in α_1 AT, the archetypal member of the serpin superfamily. The observed molecular rearrangements highlight important features not previously described, including i) internal water pathways enabling insertion of RCL residues into the protein core, ii) transient unfolding and movement of hF during insertion allowing for RCL passage, and iii) chaperoning of the RCL by hF/thFs3A through β -sheet

interactions, proposing that hF and thFs3A act in union to effectively reduce the conformational space available to the loop. In addition, the simulation was used to explore existing and transient electrostatic contacts, of which the K168-E346 interaction during insertion of the proximal hinge had a potential importance in guiding the RCL insertion. Our experimental data on charge-reversed mutants at these largely conserved positions showed that the E346K mutant was a more potent inhibitor than WT α_1 AT, whereas the K168E mutant slightly decreased the inhibitory activity. It is evident that interactions between K168 and E346 are not pivotal for the insertion mechanism, because the loss of this contact did not abrogate inhibitory function. Mutations also affect other important features such as serpin metastability and the complex nature of the serpin fold complicates a complete delineation of the mechanism underlying RCL insertion. Based on the enhanced inhibitory activity and larger polymerization propensity of the E346K mutant, together with a noteworthy conservation of the E346 residue, we propose that this residue, amongst others, acts as a gate-keeper of the strain in the metastable native state.

In conclusion, we have presented a computational aMD study of the conformational transition of α_1 AT from the biologically active and metastable state to the loop-inserted thermodynamically relaxed state. This study provides new clues to key elements in the serpin inhibitory function and should find its application in delineating the fine details differing serpin polymerization from inhibition; a process key to the design of effective pharmaceuticals aimed at stabilizing the native state without compromising function.⁷⁸⁻⁸²

ACKNOWLEDGEMENTS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

We thank Prof. J. A. McCammon for valuable discussions regarding the aMD simulation.
O.J.A. would like to thank Prof. J. A. McCammon for support during his stay at UCSD.

SUPPORTING INFORMATION

- The Supporting Information is available free of charge on the ACS Publications website.
- Animations of the aMD simulation.
- Simulation setup of the 2 μ s cMD simulation.
- Details on the expression and purification of α_1 AT WT and mutants.
- List of inhibitory and non-inhibitory human serpins.
- Supporting Figures.

REFERENCES

1. Huntington, J. A. (2006) Shape-shifting serpins - advantages of a mobile mechanism. *Trends Biochem. Sci.* 31, 427-435.
2. Stratikos, E., and Gettins, P. G. W. (1999) Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into β -sheet A. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4808-4813.
3. Gooptu, B., and Lomas, D. A. (2008) Polymers and inflammation: Disease mechanisms of the serpinopathies. *J. Exp. Med.* 205, 1529-1534.
4. Perlmutter, D. H. (2004) Alpha-1-antitrypsin deficiency: Diagnosis and treatment. *Clin. Liver Dis.* 8, 839-859.
5. Nyon, M. P., and Gooptu, B. (2014) Therapeutic targeting of misfolding and conformational change in α_1 -antitrypsin deficiency. *Future Med. Chem.* 6, 1047-1065.
6. McElvaney, O. J., Bella, A. M. E., and McElvaney, N. G. (2015) α_1 -antitrypsin deficiency: Current and future treatment options. *Expert Opin. Orphan Drugs.* 3, 5-14.
7. Patschull, A. O. M., Segu, L., Nyon, M. P., Lomas, D. A., Nobeli, I., Barrett, T. E., and Gooptu, B. (2011) Therapeutic target-site variability in α_1 -antitrypsin characterized at high resolution. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 67, 1492-1497.
8. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157-162.
9. Gettins, P. G. W. (2002) Serpin structure, mechanism, and function. *Chem. Rev.* 102, 4751-4804.
10. Baek, J., Im, H., Kang, U., Seong, K. M., Lee, C., Kim, J., and Yu, M. (2007) Probing the local conformational change of α_1 -antitrypsin. *Protein Sci.* 16, 1842-1850.
11. Kang, U., Baek, J., Ryu, S., Kim, J., Yu, M., and Lee, C. (2004) Kinetic mechanism of protease inhibition by α_1 -antitrypsin. *Biochem. Biophys. Res. Commun.* 323, 409-415.
12. Im, H., Woo, M. S., Hwang, K. Y., and Yu, M. H. (2002) Interactions causing the kinetic trap in serpin protein folding. *J. Biol. Chem.* 277, 46347-46354.
13. Shin, J. S., and Yu, M. H. (2002) Kinetic dissection of α_1 -antitrypsin inhibition mechanism. *J. Biol. Chem.* 277, 11629-11635.

14. Zhou, A., Carrell, R. W., and Huntington, J. A. (2001) The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. *J. Biol. Chem.* 276, 27541-27547.
15. Im, H., Ahn, H., and Yu, M. (2000) Bypassing the kinetic trap of serpin protein folding by loop extension. *Protein Sci.* 9, 1497-1502.
16. Hopkins, P. C. R., Carrell, R. W., and Stone, S. R. (1993) Effects of mutations in the hinge region of serpins. *Biochemistry.* 32, 7650-7657.
17. Dobó, J., and Gettins, P. G. W. (2004) α_1 -proteinase inhibitor forms initial non-covalent and final covalent complexes with elastase analogously to other serpin-proteinase pairs, suggesting a common mechanism of inhibition. *J. Biol. Chem.* 279, 9264-9269.
18. Elliott, P. R., Pei, X. Y., Dafforn, T. R., and Lomas, D. A. (2000) Topography of a 2.0 Å structure of α_1 -antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci.* 9, 1274-1281.
19. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature.* 407, 923-926.
20. Elliott, P. R., Abrahams, J. P., and Lomas, D. A. (1998) Wild-type α_1 -antitrypsin is in the canonical inhibitory conformation. *J. Mol. Biol.* 275, 419-425.
21. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C., and Huntington, J. A. (2011) Molecular basis of α_1 -antitrypsin deficiency revealed by the structure of a domain - swapped trimer. *EMBO Rep.* 12, 1011-1017.
22. Peterson, F. C., and Gettins, P. G. W. (2001) Insight into the mechanism of serpin-proteinase inhibition from 2D [^1H - ^{15}N] NMR studies of the 69 kDa α_1 -proteinase inhibitor Pittsburgh-Trypsin covalent complex. *Biochemistry.* 40, 6284-6292.
23. Tsutsui, Y., Dela Cruz, R., and Wintrode, P. L. (2012) Folding mechanism of the metastable serpin α_1 -antitrypsin. *Proc. Natl. Acad. Sci. U. S. A.* 109, 4467-4472.
24. Zheng, X., Wintrode, P. L., and Chance, M. R. (2008) Complementary structural mass spectrometry techniques reveal local dynamics in functionally important regions of a metastable serpin. *Structure.* 16, 38-51.
25. Tsutsui, Y., Liu, L., Gershenson, A., and Wintrode, P. L. (2006) The conformational dynamics of a metastable serpin studied by hydrogen exchange and mass spectrometry. *Biochemistry.* 45, 6561-6569.

26. Maddur, A. A., Swanson, R., Izaguirre, G., Gettins, P. G. W., and Olson, S. T. (2013) Kinetic intermediates en route to the final serpin-protease complex: Studies of complexes of α_1 -protease inhibitor with trypsin. *J. Biol. Chem.* 288, 32020-32035.
27. Cazzolli, G., Wang, F., a Beccara, S., Gershenson, A., Faccioli, P., and Wintrode, P. L. (2014) Serpin latency transition at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15414-15419.
28. Krüger, P., Verheyden, S., Declerck, P. J., and Engelborghs, Y. (2001) Extending the capabilities of targeted molecular dynamics: Simulation of a large conformational transition in plasminogen activator inhibitor 1. *Protein Sci.* 10, 798-808.
29. Hamelberg, D., de Oliveira, C. A. F., and McCammon, J. A. (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.* 127, 155102.
30. Dementiev, A., Simonovic, M., Volz, K., and Gettins, P. G. W. (2003) Canonical inhibitor-like interactions explain reactivity of α_1 -proteinase inhibitor Pittsburgh and antithrombin with proteinases. *J. Biol. Chem.* 278, 37881-37887.
31. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* 7, 525-537.
32. Protein Preparation Wizard 2013-2: Epik version 2.4; Impact version 5.9; Prime version 3.2, 2013, Schrödinger, LLC, New York, NY.
33. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* 27, 221-234.
34. Durell, S. R., Brooks, B. R., and Ben-Naim, A. (1994) Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* 98, 2198-2202.
35. Neria, E., Fischer, S., and Karplus, M. (1996) Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105, 1902-1921.
36. Paci, E., and Karplus, M. (1999) Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *J. Mol. Biol.* 288, 441-459.
37. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781-1802.
38. MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L.,

- Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102, 3586-3616.
39. MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., III (2004) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126, 698-699.
40. Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D., Jr. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8, 3257-3273.
41. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101, 4177-4189.
42. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103, 4613-4621.
43. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089-10092.
44. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n -alkanes. *J. Comput. Phys.* 23, 327-341.
45. Miao, Y., Feixas, F., Eun, C., and McCammon, J. A. (2015) Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* 36, 1536-1549.
46. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics.* 14, 33-38.
47. CLC Main Workbench 6.5, CLC bio, <http://www.clcbio.com> (accessed November 4, 2016).
48. Cabrita, L. D., Dai, W., and Bottomley, S. P. (2004) Different conformational changes within the F-helix occur during serpin folding, polymerization, and proteinase inhibition. *Biochemistry.* 43, 9834-9839.
49. Baek, J., Yang, W. S., Lee, C., and Yu, M. (2009) Functional unfolding of α_1 -antitrypsin probed by hydrogen-deuterium exchange coupled with mass spectrometry. *Mol. Cell. Proteomics.* 8, 1072-1081.

50. Cabrita, L. D., Whisstock, J. C., and Bottomley, S. P. (2002) Probing the role of the F-helix in serpin stability through a single tryptophan substitution. *Biochemistry*. 41, 4575-4581.
51. Knaupp, A. S., Keleher, S., Yang, L., Dai, W., Bottomley, S. P., and Pearce, M. C. (2013) The roles of helix I and strand 5A in the folding, function and misfolding of α_1 -antitrypsin. *PLoS ONE*. 8, e54766.
52. Krishnan, B., and Gierasch, L. M. (2011) Dynamic local unfolding in the serpin α_1 -antitrypsin provides a mechanism for loop insertion and polymerization. *Nat. Struct. Mol. Biol.* 18, 222-226.
53. Tsutsui, Y., Kuri, B., Sengupta, T., and Wintrode, P. L. (2008) The structural basis of serpin polymerization studied by hydrogen/deuterium exchange and mass spectrometry. *J. Biol. Chem.* 283, 30804-30811.
54. Gettins, P. G. W. (2002) The F-helix of serpins plays an essential, active role in the proteinase inhibition mechanism. *FEBS Lett.* 523, 2-6.
55. Irving, J., Haq, I., Dickens, J., Faull, S., and Lomas, D. (2014) Altered native stability is the dominant basis for susceptibility of α_1 -antitrypsin mutants to polymerization. *Biochem. J.* 460, 103-119.
56. Gooptu, B., Hazes, B., Chang, W. W., Dafforn, T. R., Carrell, R. W., Read, R. J., and Lomas, D. A. (2000) Inactive conformation of the serpin α_1 -antichymotrypsin indicates two-stage insertion of the reactive loop: Implications for inhibitory function and conformational disease. *Proc. Natl. Acad. Sci. U. S. A.* 97, 67-72.
57. Schechter, N. M., and Plotnick, M. I. (2004) Measurement of the kinetic parameters mediating protease-serpin inhibition. *Methods*. 32, 159-168.
58. Chaillan-Huntington, C. E., and Patston, P. A. (1998) Influence of the P₅ residue on α_1 -proteinase inhibitor mechanism. *J. Biol. Chem.* 273, 4569-4573.
59. Im, H., Seo, E. J., and Yu, M. (1999) Metastability in the inhibitory mechanism of human α_1 -antitrypsin. *J. Biol. Chem.* 274, 11072-11077.
60. van Gent, D., Sharp, P., Morgan, K., and Kalsheker, N. (2003) Serpins: Structure, function and molecular evolution. *Int. J. Biochem. Cell Biol.* 35, 1536-1547.
61. Irving, J. A., Pike, R., Lesk, A. M., and Whisstock, J. C. (2000) Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. *Genome Res.* 10, 1845-1864.

62. Seo, E. J., Im, H., Maeng, J.-S., Kim, K. E., and Yu, M.-H. (2000) Distribution of the native strain in human α_1 -antitrypsin and its association with protease inhibitor function. *J. Biol. Chem.* 275, 16904-16909.
63. Knaupp, A. S., and Bottomley, S. P. (2011) Structural change in β -sheet A of Z α_1 -antitrypsin is responsible for accelerated polymerization and disease. *J. Mol. Biol.* 413, 888-898.
64. Pearce, M. C., Cabrita, L. D., Ellisdon, A. M., and Bottomley, S. P. (2007) The loss of tryptophan 194 in antichymotrypsin lowers the kinetic barrier to misfolding. *FEBS J.* 274, 3622-3632.
65. Lomas, D. A., and Carrell, R. W. (2002) Serpinopathies and the conformational dementias. *Nat. Rev. Genet.* 3, 759-768.
66. Haq, I., Irving, J. A., Saleh, A. D., Dron, L., Regan-Mochrie, G., Motamedi-Shad, N., Hurst, J. R., Gooptu, B., and Lomas, D. A. (2016) Deficiency mutations of α_1 -antitrypsin differentially affect folding, function and polymerization. *Am. J. Respir. Cell Mol. Biol.* 54, 71-80.
67. Lee, K. N., Im, H., Kang, S. W., and Yu, M. H. (1998) Characterization of a human α_1 -antitrypsin variant that is as stable as ovalbumin. *J. Biol. Chem.* 273, 2509-2516.
68. Crystal, R. G. (1990) α_1 -antitrypsin deficiency, emphysema, and liver disease - genetic basis and strategies for therapy. *J. Clin. Invest.* 85, 1343-1352.
69. Kass, I., Knaupp, A., Bottomley, S., and Buckle, A. (2012) Conformational properties of the disease-causing Z variant of α_1 -antitrypsin revealed by theory and experiment. *Biophys. J.* 102, 2856-2865.
70. Buck, M. J., and Atchley, W. R. (2005) Networks of coevolving sites in structural and functional domains of serpin proteins. *Mol. Biol. Evol.* 22, 1627-1634.
71. Im, H., and Yu, M. H. (2000) Role of Lys335 in the metastability and function of inhibitory serpins. *Protein Sci.* 9, 934-941.
72. Knaupp, A. S., Levina, V., Robertson, A. L., Pearce, M. C., and Bottomley, S. P. (2010) Kinetic instability of the serpin Z α_1 -antitrypsin promotes aggregation. *J. Mol. Biol.* 396, 375-383.
73. Seo, E. J., Lee, C., and Yu, M.-H. (2002) Concerted regulation of inhibitory activity of α_1 -antitrypsin by the native strain distributed throughout the molecule. *J. Biol. Chem.* 277, 14216-14220.

74. Yamasaki, M., Sendall, T. J., Harris, L. E., Lewis, G. M. W., and Huntington, J. A. (2010) Loop-sheet mechanism of serpin polymerization tested by reactive center loop mutations. *J. Biol. Chem.* 285, 30752-30758.
75. Huntington, J. A., Pannu, N. S., Hazes, B., Read, R. J., Lomas, D. A., and Carrell, R. W. (1999) A 2.6 Å structure of a serpin polymer and implications for conformational disease. *J. Mol. Biol.* 293, 449-455.
76. Agarwal, S., Jha, S., Sanyal, I., and Amla, D. V. (2010) Expression and purification of recombinant human α_1 -proteinase inhibitor and its single amino acid substituted variants in *Escherichia coli* for enhanced stability and biological activity. *J. Biotechnol.* 147, 64-72.
77. Gils, A., Vleugels, N., Tobback, K., Knockaert, I., and Declerck, P. J. (1998) Characterization of plasminogen activator inhibitor 1 mutants containing the P13 to P10 region of ovalbumin or antithrombin III: Evidence that the P13 residue contributes significantly to the active to substrate transition. *Biochim. Biophys. Acta.* 1387, 291-297.
78. McNab, G. L., Dafforn, T. R., Wood, A., Sapey, E., and Stockley, R. A. (2012) A novel model and molecular therapy for Z α_1 -antitrypsin deficiency. *Mamm. Genome.* 23, 241-249.
79. Patschull, A. O. M., Gooptu, B., Ashford, P., Daviter, T., and Nobeli, I. (2012) *In silico* assessment of potential druggable pockets on the surface of α_1 -antitrypsin conformers. *PLoS ONE.* 7, e36612.
80. Mallya, M., Phillips, R. L., Saldanha, S. A., Gooptu, B., Leigh Brown, S. C., Termine, D. J., Shirvani, A. M., Wu, Y., Sifers, R. N., Abagyan, R., and Lomas, D. A. (2007) Small molecules block the polymerization of Z α_1 -antitrypsin and increase the clearance of intracellular aggregates. *J. Med. Chem.* 50, 5357-5363.
81. Burrows, J. A. J., Willis, L. K., and Perlmutter, D. H. (2000) Chemical chaperones mediate increased secretion of mutant α_1 -antitrypsin (α_1 -AT) Z: A potential pharmacological strategy for prevention of liver injury and emphysema in α_1 -AT deficiency. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1796-1801.
82. Ordóñez, A., Pérez, J., Tan, L., Dickens, J. A., Motamedi-Shad, N., Irving, J. A., Haq, I., Ekeowa, U., Marciniak, S. J., Miranda, E., and Lomas, D. A. (2015) A single-chain variable fragment intrabody prevents intracellular polymerization of Z α_1 -antitrypsin while allowing its antiproteinase activity. *FASEB J.* 29, 2667-2678.

For Table of Contents Use Only

The reactive center loop insertion in α -1-antitrypsin captured by accelerated molecular dynamics simulation

Ole Juul Andersen, Michael Wulff Risør, Emil Christian Poulsen, Niels Chr. Nielsen, Yinglong Miao, Jan J. Enghild, and Birgit Schiøtt