

OPEN

Residues W215, E217 and E192 control the allosteric E*-E equilibrium of thrombin

Leslie A. Pelc, Sarah K. Koester, Zhiwei Chen, Noah E. Gistover & Enrico Di Cera

A pre-existing, allosteric equilibrium between closed (E*) and open (E) conformations of the active site influences the level of activity in the trypsin fold and defines ligand binding according to the mechanism of conformational selection. Using the clotting protease thrombin as a model system, we investigate the molecular determinants of the E*-E equilibrium through rapid kinetics and X-ray structural biology. The equilibrium is controlled by three residues positioned around the active site. W215 on the 215–217 segment defining the west wall of the active site controls the rate of transition from E to E* through hydrophobic interaction with F227. E192 on the opposite 190–193 segment defining the east wall of the active site controls the rate of transition from E* to E through electrostatic repulsion of E217. The side chain of E217 acts as a lever that moves the entire 215–217 segment in the E*-E equilibrium. Removal of this side chain converts binding to the active site to a simple lock-and-key mechanism and freezes the conformation in a state intermediate between E* and E. These findings reveal a simple framework to understand the molecular basis of a key allosteric property of the trypsin fold.

Trypsin-like proteases utilize a catalytic triad for activity, composed of the highly conserved residues H57, D102 and S195. Catalysis is assisted by several residues within the active site^{1–3}. D189 at the bottom of the primary specificity pocket engages the Arg/Lys residue at the P1 position of substrate⁴. The oxyanion hole defined by the backbone N atoms of S195 and G193 stabilizes the developing partial charge on the tetrahedral intermediate during the catalytic cycle. The 215–217 segment defines the west wall of the active site and provides additional anchor points for substrate residues immediately upstream of the peptide bond to be cleaved. A peculiar property of this segment borne out from analysis of hundreds of crystal structures deposited in the Protein Data Bank (PDB) is that it can assume alternative conformations that directly influence access to the primary specificity pocket^{5,6}. The D216G mutant of α I-trypsin crystallizes in the free form with the 215–217 segment in equilibrium between open and closed conformations in the same crystal⁷. The clotting protease thrombin crystallizes in the open or closed conformation depending on solution conditions⁸. A high resolution structure of chymotrypsinogen was the first to reveal two distinct conformations of the 215–217 segment in two molecules of the asymmetric unit⁹. Crystals of prethrombin-2 harvested from the same well document alternative arrangements of the 215–217 segment¹⁰.

The flexibility of the 215–217 segment is an intrinsic property of the trypsin fold and has functional consequences. Binding of a ligand to the active site requires the 215–217 segment to assume an “open” configuration and is precluded in the “closed” one when side chains and backbone shift to occlude access to the primary specificity pocket. Consistent with this scenario, recent rapid kinetics studies of proteases like chymotrypsin, thrombin, factor Xa and activated protein C^{8,11–16} have shown that ligand binding to the active site does not obey “induced fit”¹⁷, where a conformational rearrangement of the complex follows the initial binding step, but rather obeys the alternative mechanism of “conformational selection”¹⁸, where the ligand selects optimal conformations from a pre-existing equilibrium of closed (E*) and open (E) forms that precedes the binding step. The mechanism of conformational selection also applies to ligand binding to the zymogen^{15,19}. The E* form predominates in the zymogen and is replaced by the E form in the protease. The replacement is gradual along the activation pathway, as illustrated by a recent investigation of the conversion of prothrombin to thrombin via the intermediates prethrombin-2 and meizothrombin¹⁹. The celebrated Huber-Bode mechanism of zymogen activation²⁰ envisions a transition triggered by proteolytic cleavage in the conserved activation domain and subsequent organization of the active site region. The E*-E equilibrium layers on top of this mechanism and casts activation as a shift along a pre-existing spectrum of conformations. The new paradigm links activity to the intrinsic dynamics of the fold,

Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, St. Louis, MO, 63104, USA. Correspondence and requests for materials should be addressed to E.D.C. (email: enrico@slu.edu)

which raises important questions about the structural determinants of this linkage and the mechanism under-scoring the interconversion between E^* and E .

A number of candidates emerge from analysis of the current structural database and three residues deserve particular attention. The highly conserved residue W215 shuttles in and out of the active site entrance and functions as a lid that closes and opens access to the primary specificity pocket^{5,6}. This residue has long been considered the major structural determinant of the E^* - E equilibrium, but recent studies support involvement of additional residues²¹. In the clotting protease thrombin²², E217 preferentially contributes to procoagulant and prothrombotic activities^{23,24}, along with W215^{25,26}, and may influence the E^* - E equilibrium through its H-bonding interaction with the side chain of K224 that stabilizes the 215–217 segment and the E form in thrombin²⁷, prethrombin-2¹⁰ and prothrombin^{28–30}. On the other hand, E217 may stabilize the E^* form by H-bonding to the active site S195 and occluding access to the primary specificity pocket, as reported recently for plasma kallikrein³¹. There is currently no crystal structure of the W215A or E217A mutants, but the double mutant W215A/E217A of thrombin crystallizes in a collapsed form similar to E^* ³² and so does the E217K mutant³³. Residue E192 is positioned on the east wall of the active site, across from E217 and the 215–217 segment defining the west wall^{27,34}. The side chain of E192 is an uncompensated negative charge that may influence the dynamics of neighbor side chains such as E217 through electrostatic coupling³⁵. The conformation of the 215–217 backbone is important in catalysis³⁶ and contributes to correct orientation of substrate in the active site^{1–3}. It is unclear how the conformation of this backbone is linked to the orientation of the side chains of W215 and E217. Residue 216 of the 215–217 segment is a Gly.

In this study we investigate the role of the side chains of W215, E217 and E192 through Ala substitutions and rapid kinetics of binding of the irreversible inhibitor H-D-Phe-Pro-Arg-CH₂Cl (PPACK). The structure of PPACK is similar to that of the highly specific chromogenic substrate H-D-Phe-Pro-Arg-p-nitroanilide (FPR), except for replacement of the leaving group with a CH₂Cl that alkylates the active site residues S195 and H57. The results reveal a simple mechanism for the E^* - E equilibrium and provide the starting point for additional analysis.

Results

PPACK binding to wild-type obeys conformational selection. An informative approach to the study of ligand binding to the active site of protease or zymogen has been developed recently based on stopped flow measurements^{15,19}. Extension of this approach to the case of the irreversible inhibitor PPACK allows for the study of the properties of wild-type and mutants without the need to prevent catalysis with the additional S195A replacement. Figure 1A shows the value of the slow relaxation (α_2) for PPACK binding to wild-type thrombin as a function of PPACK concentration. Unlike the case of FPR binding to the S195A mutant of thrombin, where both the fast and slow relaxations in Eq. 2 could be resolved experimentally^{15,19}, the fast relaxation (α_1) for PPACK binding could not be detected because it is too fast to be resolved within the dead time of the stopped flow apparatus (0.5–1 ms) or spectroscopically silent. Because $\alpha_2(0) = k_{off}$ in the mechanism of conformational selection^{15,16,19,21}, a value of zero for this lower asymptote confirms the irreversible nature of PPACK inhibition. The upper asymptotic value $\alpha_2(\infty) = k_{12}$ measures the rate for opening of the active site in the $E^* \rightarrow E$ transition. The value $56 \pm 3 \text{ s}^{-1}$ measured for this rate constant is significantly faster than the value of $16 \pm 1 \text{ s}^{-1}$ measured for FPR binding to the S195A mutant. The difference underscores intrinsic changes in dynamics of the enzyme caused by replacement of the active site residue S195, whose importance as an end-point of allosteric transduction has been documented recently³⁷. The lack of information on the fast relaxation makes it difficult to estimate the other two independent parameters k_{on} and k_{21} in Eq. 2. However, after constraining the value of k_{on} within a range consistent with the value measured recently for FPR binding to S195A^{15,19}, a local minimum in parameter space yields a best-fit estimate for $k_{21} = 8.8 \pm 0.3 \text{ s}^{-1}$ (Table 1). This implies that wild-type thrombin exists in equilibrium between two conformations, E^* and E , that exchange over a time scale $\tau = (k_{12} + k_{21})^{-1} = 15 \text{ ms}$ in a 1:6 ratio. FPR binding to S195A thrombin reveals an equilibrium between two conformations, E^* and E , that exchange over a longer time scale $\tau = (k_{12} + k_{21})^{-1} = 56 \text{ ms}$ in a 1:4 ratio. Replacement of S195 with Ala abrogates catalytic activity but also influences the dynamics of exchange between E^* and E . Importantly, the relative distribution of these conformations is affected little and confirms a prevalence of the E conformation for the mature protease.

To rule out differences in the mechanism of interaction between FPR and PPACK, that share the same peptide sequence but a different C-terminal blocking group, we measured the binding of PPACK to the S195A mutant of thrombin. In this case, acylation of S195 is not possible and PPACK behaves as a reversible binder with a finite value of $k_{off} = 4.6 \pm 0.2 \text{ s}^{-1}$ (Table 1) similar to that measured for the chromogenic substrate FPR. Importantly, this result holds over the short (ms) time scale of stopped flow measurements, because longer incubation (weeks) produces acylation of H57, as demonstrated by a crystal structure of the S195A mutant of meizothrombin desF1 solved in the presence of PPACK (Fig. S1 and Table S1). Binding of PPACK to the S195A mutant produces two relaxations (Fig. 1B) that can be fit to the same values of $k_{12} = 16 \pm 1 \text{ s}^{-1}$ and $k_{21} = 3.8 \pm 0.8 \text{ s}^{-1}$ reported for FPR binding under identical solution conditions¹⁹. Hence, FPR and PPACK bind with the same mechanism and with identical values of rate constants that reflect the interconversion of E^* and E in the free protein.

To further validate the results obtained with PPACK binding to wild type thrombin, we measured the interaction as a function of temperature using a strategy developed for the analysis of steady state kinetics³⁸. Because binding and conformational transitions in Eq. 1 have distinct activation energies, their contribution to the observed values of the relaxations in Eq. 2 changes with temperature and enables unequivocal resolution of all independent parameters involved from a global fit of the data. Data collected in the temperature range 5–30 °C (Fig. 1C) were simultaneously fit to Eq. 2 with the rate constants expressed according to the Arrhenius Eq. 3 and allowed resolution of all rate constants and their activation energies. The values obtained at the reference temperature of 15 °C are in excellent agreement with those reported in Table 1 from analysis of the data in Fig. 1A obtained independently under identical solution conditions. Resolution of the activation energies from the data in Fig. 1C provides additional details about the E^* - E equilibrium, with the $E^*:E$ ratio increasing with temperature

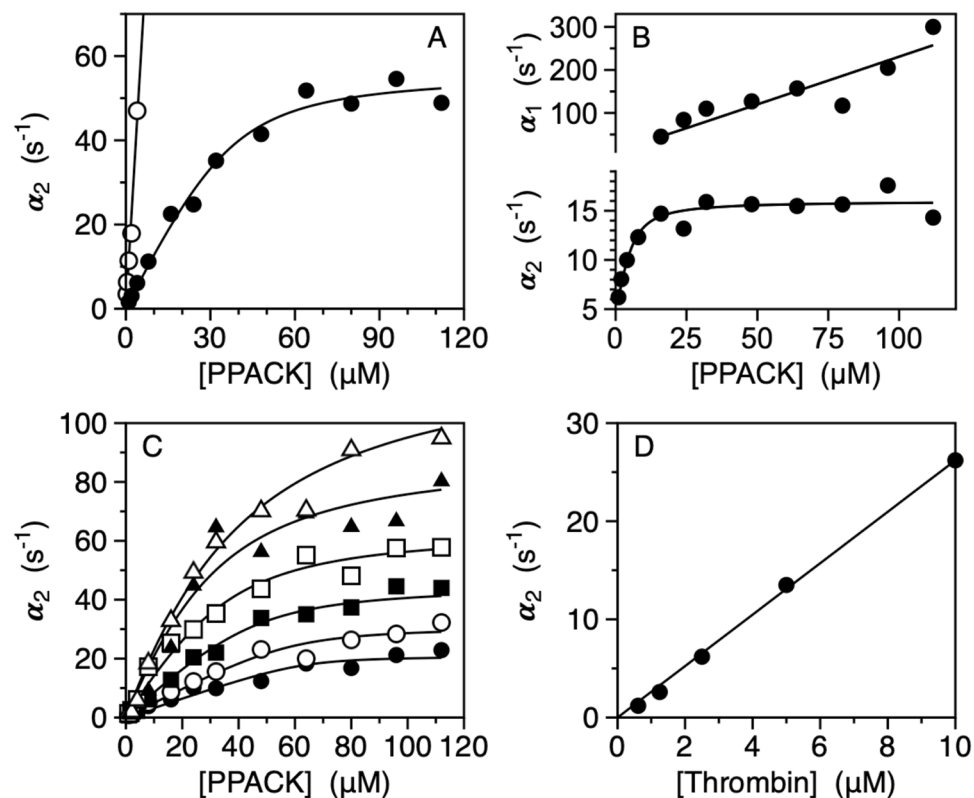


Figure 1. (A) Rapid kinetics of PPACK binding to free thrombin (closed circles) showing a single, saturable relaxation that increases hyperbolically with the ligand concentration. The continuous line was drawn according to Eq. 2 in the text with best-fit parameter values listed in Table 1. In the presence of Na^+ , the kinetic profile changes into a straight line (open circles) that obeys the lock-and-key expression^{16,49} $\alpha = k_{\text{off}} + k_{\text{on}}[L]$, with best-fit parameter values $k_{\text{off}} = 0 \text{ s}^{-1}$ and $k_{\text{on}} = 11 \pm 1 \mu\text{M}^{-1}\text{s}^{-1}$. (B) Rapid kinetics of PPACK binding to the thrombin mutant S195A showing the two relaxations predicted by the mechanism of conformational selection in Eq. 1. Continuous lines were drawn according to Eq. 2 in the text with best-fit parameter values listed in Table 1. (C) Rapid kinetics of PPACK binding to free thrombin over the temperature range 5–30 °C. The continuous lines were drawn according to Eq. 2 in the text with each rate constant expressed as in Eq. 3 and represent a global fit of the entire data set with best-fit parameter values: $k_{\text{on}} = 1.1 \pm 0.1 \mu\text{M}^{-1}\text{s}^{-1}$, $k_{12} = 45 \pm 3 \text{ s}^{-1}$, $k_{21} = 6.4 \pm 0.4 \text{ s}^{-1}$, $E_{\text{on}} = 20 \pm 3 \text{ kcal/mol}$, $E_{12} = 12 \pm 1 \text{ kcal/mol}$, $E_{21} = 37 \pm 5 \text{ kcal/mol}$. The values of the rate constants are at the reference temperature $T_0 = 288.15 \text{ K}$ (15 °C) for the sake of comparison with the data in Figs 1A and 2. Experimental conditions are: 400 mM ChCl, 50 mM Tris, 0.1% PEG8000, pH 8.0 at 5 °C (closed circles), 10 °C (open circles), 15 °C (closed squares), 20 °C (open squares), 25 °C (closed triangles), 30 °C (open triangles). (D) Rapid kinetics of PPACK binding to thrombin in the presence of excess enzyme. The single, saturable hyperbolic increase observed with excess ligand (A) becomes a simple linear relaxation in the presence of excess macromolecule if the interaction takes place according to the mechanism of conformational selection. The slope of the straight line ($2.1 \pm 0.1 \mu\text{M}^{-1}\text{s}^{-1}$) is a measure of the k_{on} times the fraction of thrombin in the E form. The value is in good agreement with the best-fit values obtained from analysis of the data in panels A (Table 1) and C. Experimental conditions are: 400 mM ChCl, 50 mM Tris, 0.1% PEG8000, pH 8.0, at 15 °C. Data in the presence of Na^+ (A) were obtained by replacing ChCl with NaCl in the buffer. Experimental errors for all data shown in Fig. 1A–D are 5% or less.

	$k_{12} (\text{s}^{-1})$	$k_{21} (\text{s}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$	$k_{\text{on}} (\mu\text{M}^{-1}\text{s}^{-1})$	$\tau (\text{ms})$	E*:E
wt	56 ± 3	8.8 ± 0.3	0 ± 0	1.5 ± 0.1	15 ± 1	1:6
S195A	16 ± 1	3.8 ± 0.8	4.6 ± 0.2	2.1 ± 0.1	51 ± 3	1:4
W215A	51 ± 3	110 ± 10	0 ± 0	1.2 ± 0.1	6.2 ± 0.3	2:1
E217A	—	—	0 ± 0	0.070 ± 0.002	—	—
E192A	11 ± 1	19 ± 1	0 ± 0	1.0 ± 0.1	33 ± 2	2:1
W215A/E217A	—	—	0 ± 0	0.00019 ± 0.00001	—	—

Table 1. Kinetic rate constants for PPACK binding to thrombin wild-type and mutants. Experimental conditions are: 400 mM ChCl, 50 mM Tris, 0.1% PEG8000, pH 8.0, at 15 °C.

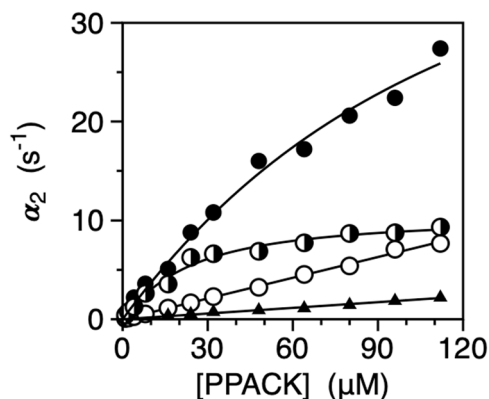


Figure 2. Rapid kinetics of PPACK binding to the thrombin mutants W215A (closed circles), E192A (mixed circles), E217A (open circles) and W215A/E217A (closed triangles). Binding to W215A and E192A obeys a saturable relaxation that increase hyperbolically with PPACK, as seen for wild-type (Fig. 1A). Binding to E217A and W217A/E217A obeys a simple straight line consistent with a lock-and-key mechanism of interaction. Values for W215A/E217A were increased 100-fold to enable comparison in the plot. Continuous lines were drawn according to Eq. 2 in the text (W215A and E192A) or with the lock-and-key expression^{16,49} $\alpha = k_{off} + k_{on}[L]$ (E217A and W215A/E217A) with best-fit parameter values listed in Table 1. Experimental conditions are: 400 mM ChCl, 50 mM Tris, 0.1% PEG8000, pH 8.0, at 15 °C. Experimental errors are 5% or less.

due to the higher activation energy associated with the $E \rightarrow E^*$ transition (37 ± 5 kcal/mol) compared to that of the $E^* \rightarrow E$ transition (12 ± 1 kcal/mol).

A necessary test of the validity of conformational selection requires measurements of PPACK binding under conditions where thrombin is in large excess relative to the inhibitor¹⁵. With excess macromolecule, the profile of α_2 remains hyperbolic if binding obeys induced fit, but changes to a simple straight line in the case of conformational selection^{15,39–42}. This is because a conformational transition preceding the binding step is not perturbed significantly when the macromolecule is in excess, in which case binding takes place only in the E form as a simple lock-and-key interaction⁴². Binding of PPACK under conditions where thrombin is in large excess obeys a straight line (Fig. 1D), as recently observed for FPR binding to the thrombin mutant S195A in excess concentrations¹⁵. Further support to the mechanism of conformational selection comes from measurements of PPACK binding in the presence of Na^+ , that is known to rigidify the fold^{43,44} and to stabilize the E form¹⁵. Under these conditions, PPACK binding obeys a single relaxation that increases linearly with ligand concentration (Fig. 1A), as expected for a simple lock-and-key rigid body association with a negligible rate of dissociation. The effect of Na^+ is of physiological relevance because it counters the progressive shift of E to E^* as temperature increases (Fig. 1C).

Probing the structural determinants of the E^* -E equilibrium. Having established the mechanism for PPACK binding in terms of conformational selection, we investigated the structural determinants of the E^* -E equilibrium using the same strategy based upon rapid kinetics measurements. The structural database suggests that closure of the active site in the E^* form is caused by the side chain of W215 coming in contact with W60d in the 60-loop and by a shift in the backbone of 215–217 that collapses the aperture leading to the primary specificity pocket^{5,6}. A good measure of this aperture is the $\text{C}\alpha$ - $\text{C}\alpha$ distance between the highly conserved residues G193 in the oxyanion hole and G216 in the 215–217 segment that features a bimodal distribution with peaks at 8.3 Å for structures in the E^* form and 12.2 Å for structures in the E form^{5,6}. The aperture in the E^* form is not wide enough for a ligand like PPACK to access the primary specificity pocket^{5,6}. Additional steric hindrance in the E^* form may come from the side chain of E192 guarding the east wall of the active site and often collapsing against the catalytic S195^{33,35}, or the side chain of E217 as shown recently for plasma kallikrein³¹. When a residue occludes the active site in the E^* form through its side chain, replacement with Ala is expected to change the mechanism of binding from conformational selection to a lock-and-key interaction producing a simple linear relaxation^{16,45,46}. Removal of the steric hindrance in the E^* form should equalize the environments of E^* and E regarding accessibility of the primary specificity pocket, thereby eliminating the main functional difference between the two forms. Furthermore, the kinetic signatures of such Ala substitution should be the same as those of the E form, unless additional structural perturbations are at play such as stabilization of a new conformation intermediate between E^* and E. Alternatively, the Ala replacement may result in a perturbed kinetic profile that remains hyperbolic as for the wild-type, but with a shift of the upper asymptote caused by a change in k_{12} and/or a change in the initial slope caused by a change in k_{on} .

Role of W215 and structure of the W215A mutant. Figure 2 shows the kinetic profile for the W215A mutant of thrombin where the role of the indole side chain is tested with an Ala replacement. Rapid kinetics of PPACK binding produce a single relaxation that increases hyperbolically with the concentration of PPACK as seen for wild-type (Fig. 1A). The profile is consistent with a pre-equilibrium between E^* and E that is shifted in favor of the E^* form, with a value of $k_{12} = 51 \pm 3 \text{ s}^{-1}$ for the $E^* \rightarrow E$ transition that is comparable to that of wild-type (Table 1). The exact value of the E^* :E distribution requires knowledge of k_{21} that cannot be resolved

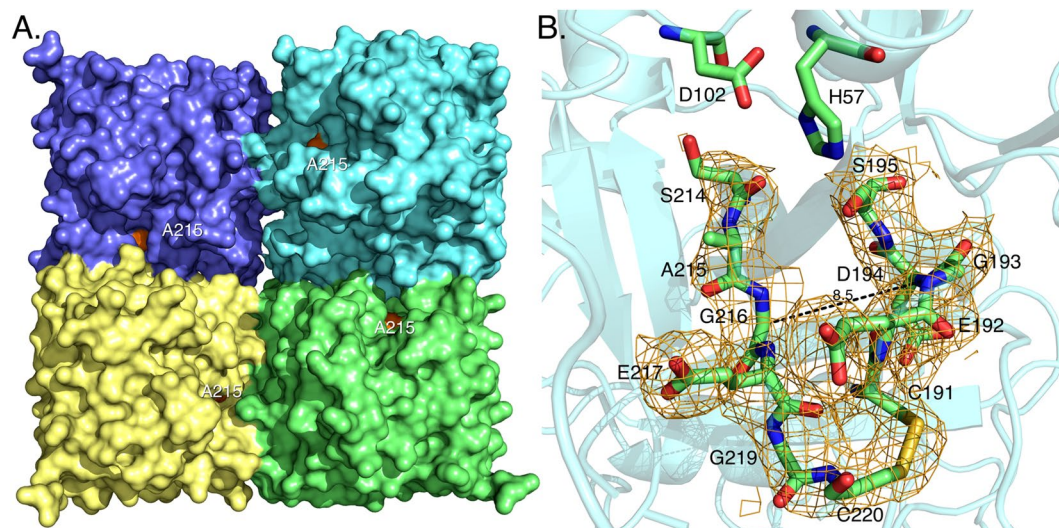


Figure 3. X-ray crystal structure of the thrombin mutant W215A. **(A)** Surface representation showing the architecture of the fold and the site of mutation (orange) on the west wall of the active site. **(B)** Details of the conformation of the active site region with relevant side chains and density map contoured at 1.2σ . Access to the primary specificity pocket is constrained by a collapse of the 215–217 segment that brings the $C\alpha$ atoms of G193 and G216 within 8.2 Å, as typically observed in the E^* form^{5,6}. Details of the structure are given in Table 2.

unequivocally from α_2 only. Again, searching for a minimum around the range of values measured for wild-type gives an estimate of $k_{21} = 110 \pm 10 \text{ s}^{-1}$ for the reverse reaction $E \rightarrow E^*$ that is > 10-fold faster than that of wild-type (Table 1). The time scale of the E^* -E interconversion is $\tau = (k_{12} + k_{21})^{-1} = 6.2 \text{ ms}$ and only slightly faster than that of wild-type. The estimated $E^*:E$ ratio for W215A is 2:1 and reversed compared to the 1:6 ratio measured for wild-type.

Support to the conclusion that the W215A substitution retains the E^* -E equilibrium but in a perturbed fashion that stabilizes the E^* form comes from the X-ray crystal structure of the mutant (Fig. 3). Although the structure could only be solved at low resolution (Table 2), it reveals a collapsed conformation of the 215–217 segment with the $C\alpha$ - $C\alpha$ distance between G193 in the oxyanion hole and G216 in the adjacent 215–217 segment that shrinks to 8.5 Å. Interestingly, the perturbed environment of the active site and stabilization of a conformation similar to E^* is not linked to disruption of the oxyanion hole. The backbone N atoms of G193 and S195 point in the same direction to organize a pocket for stabilization of the developing partial charge in the transition state during catalysis^{1–3}. Removal of the side chain of W215 does not equalize the environment of the active site between E^* and E, which would have generated a linear dependence of α_2 on ligand concentration. We conclude that the role of residue W215 is to keep the active site open and to slow down the $E \rightarrow E^*$ conversion by establishing an interaction with the benzene ring of F227. Disruption of this important hydrophobic interaction accelerates closure of the active site to the E^* form and results in reduced catalytic activity^{25,26}. These conclusions are in agreement with recent measurements of FPR binding to the S195A/W215A mutant of thrombin²¹.

Role of E192. Residue E192 is an uncompensated negative charge guarding access to the primary specificity pocket on the segment defining the east wall of the active site^{27,34}. Replacement of E192 has been reported to change the substrate specificity of thrombin^{47,48} and may perturb the E^* -E transition by altering the electrostatic balance around the active site³⁵. Rapid kinetics of PPACK binding to the E192A mutant produce a single relaxation that increases hyperbolically with the concentration of PPACK (Fig. 2) as seen for wild-type (Fig. 1A). The value of $k_{12} = 11 \pm 1 \text{ s}^{-1}$ for the $E^* \rightarrow E$ transition is significantly slower than that of wild-type, but the value of $k_{21} = 19 \pm 1 \text{ s}^{-1}$ for the reverse reaction $E \rightarrow E^*$ is 2-fold faster. The time scale of the E^* -E interconversion is $\tau = (k_{12} + k_{21})^{-1} = 33 \text{ ms}$ and similar to that of wild-type. The resulting $E^*:E$ ratio reverses from 1:6 in the wild-type to 2:1 in the mutant, again due to stabilization of the E^* form. Removal of the side chain of E192 does not equalize the environment of the active site between E^* and E, as seen for the W215A substitution, but causes the active site to open slower and to close slightly faster than in the wild-type. The effect is consistent with an electrostatic clash with a neighboring negatively charged side chain.

Role of E217. A likely candidate for this electrostatic clash is E217, a highly flexible residue that moves inside the active site in kallikrein³¹. Interestingly, rapid kinetics of PPACK binding to the E217A mutant produce a profile that obeys a single linear relaxation (Fig. 2) as expected for a rigid body lock-and-key mechanism of binding that does not involve conformational transitions^{16,49}. The profile is drastically different from that of wild-type and indicates that removal of the side chain of E217 freezes the conformation into a state intermediate between E^* and E to which PPACK binds with a $k_{on} = 0.070 \pm 0.002 \mu\text{M}^{-1}\text{s}^{-1}$ (Table 1). The value is > 20-fold slower than that of wild-type and not consistent with the interaction with a conformation like E. We conclude that the side chain of E217 is a major structural determinant of the E^* -E equilibrium and functions as a lever that triggers opening and closing of the access to the primary specificity pocket through an electrostatic crosstalk with E192. The more

PDB entry	6P9U
Buffer/salt	10 mM ZnSO ₄ , 100 mM MES, pH 6.5
PEG	550 MME (25%)
Data collection:	
Wavelength (Å)	1.5418
Space group	P2
Unit cell dimensions (Å)	a = 136.3, b = 44.2, c = 136.2, β = 90.04
Molecules/asymmetric unit	4
Resolution range (Å)	40–3.3
Observations	92698
Unique observations	22632
Completeness (%)	90.4 (90.6)
R _{sym} (%)	17.1 (69.5)
I/σ(I)	5.8 (2.0)
Refinement:	
Resolution (Å)	40–3.3
R _{cryst} , R _{free}	0.23, 0.31
Reflections (working/test)	21544/1087
Protein atoms	8942
Solvent molecules	7
Zn ⁺⁺ ions	7
Rmsd bond lengths ^a (Å)	0.015
Rmsd angles ^a (°)	1.8
Rmsd ΔB (Å ²) (mm/ms/ss) ^b	4.39/3.57/4.13
 protein (Å ²)	68.5
 Zn ⁺⁺ ions (Å ²)	67.2
 Solvent (Å ²)	54.5
Ramachandran plot:	
Most favored(%)	98.2
Generously allowed (%)	1.2
Disallowed (%)	0.6

Table 2. Crystallographic data for the thrombin mutant W215A. ^aRoot-mean-squared deviation (Rmsd) from ideal bond lengths and angles and Rmsd in B-factors of bonded atoms. ^bmm, main chain-main chain; ms, main chain-side chain; ss, side chain-side chain.

pronounced perturbation of the kinetic profile observed with the E217A substitution compared to the E192A mutation is explained by E217 being part of the critical 215–217 segment that moves during the E*-E transition. Without the driving force provided by E217 in its electrostatic clash with E192, the 215–217 segment freezes into an intermediate state that compromises but does not abolish access to the primary specificity pocket.

A similar scenario is observed for the double mutant W215A/E217A where two critical residues of the 215–217 segment are replaced by Ala⁵⁰. The mutant was originally engineered to enhance the anticoagulant properties of the individual substitutions of W215²⁵ and E217^{23,24}. W215A/E217A has progressed through pre-clinical^{51–58} and Phase 1 (NCT03453060) studies and will soon start Phase 2 testing (NCT039638950). Rapid kinetics of PPACK binding to the W217A/E217A double mutant produce a profile similar to that observed for the E217A mutant (Fig. 2). The combined removal of the side chains of W215 and E217 freezes the E*-E equilibrium into an intermediate state to which PPACK binds with a drastically reduced value of $k_{on} = 190 \pm 10 \text{ M}^{-1}\text{s}^{-1}$, which is >350-fold slower than that of E217A and almost 8,000-fold slower than that of wild-type (Table 1). Several structures of the W215A/E217A support this conclusion and reveal a collapsed conformation of the 215–217 segment^{32,59} that greatly compromises binding to the primary specificity pocket.

Discussion

A number of high resolution crystal structures^{5,6} and rapid kinetics studies^{15,19} support the conclusion that the trypsin fold recognizes ligand at the active site according to the mechanism of conformational selection. A pre-existing equilibrium between E and E* forms controls activity in the protease and defines a key property of the fold as it transitions from zymogen to mature enzyme. The work recently carried out for thrombin and its precursors prothrombin, prethrombin-2 and meizothrombin illustrates this scenario and offers a template for the analysis of other biological systems¹⁹. The zymogen exists predominantly in the E* form and transitions gradually to the E form as the fold matures¹⁹. A similar transition from E* to E likely takes place for poorly active proteases upon interaction with specific cofactors, as observed in clotting factor VIIa^{60,61} or complement factor D⁶². Proteases can be engineered into allosteric switches that transition from E* to E on demand^{63,64}, as demonstrated by mutants of clotting factor Xa that bypass the intrinsic pathway of coagulation and ameliorate

hemophilia conditions⁶⁵, or mutants of thrombin^{23–26} that are effective in the treatment of thrombotic complications and stroke. Finally, a mutation that stabilizes E* may compromise activity without direct interference with active site residues or catalysis, thereby offering context to better understand the molecular origin of pathologic phenotypes⁶⁶. The relative distribution of E* and E offers a molecular framework to interpret many aspects of protease function^{66,67}, as well as key properties of the zymogen. The ability to autoactivate observed in proprotein convertases^{68–70}, plasma hyaluronan-binding protein⁷¹, factor VII⁷², matriptases^{73,74}, prethrombin-2 and protein C^{75,76} should be cast in the context of the E*-E equilibrium, and so should the activating effect of streptokinase on plasminogen⁷⁷, staphylocoagulase on prothrombin⁷⁸, and of ad hoc peptides on hepatocyte growth factor^{79,80}. In all of these systems, a relevant question to be asked is what are the structural determinants of the E*-E equilibrium and how the equilibrium can be perturbed to inhibit or activate function for translational applications. The results presented here provide valuable new insights.

We hypothesize that three critical residues decorating access to the active site control the E*-E equilibrium of thrombin. Their roles differ and target specific aspects of the equilibrium. The side chain of W215 keeps the active site open by interacting with F227 and slows down the transition from E to the E* form. The side chain of E192 keeps the active site open by accelerating the transition to the E form through electrostatic repulsion of the neighbor side chain of E217. Residues W215 and E192 altogether stabilize the E form by influencing the two rates governing the E*-E equilibrium, with W215 influencing k_{21} through hydrophobic coupling with F227 and E192 influencing k_{12} by electrostatic repulsion of E217 anchored to the 215–217 segment. The side chain of E217 is under the influence of both W215 and E192 and functions as a lever promoting the E*-E equilibrium. Removal of this side chain freezes the equilibrium into an intermediate state and locks the 215–217 segment into a partially collapsed conformation that greatly reduces but does not abrogate access to the primary specificity pocket. Additional components may influence the E*-E equilibrium and future mutagenesis studies will provide details. However, residues W215, E217 and E192 likely represent the end points of a molecular mechanism that opens and closes access to the primary specificity pocket. These are the three key players that control the E*-E equilibrium in thrombin according to the model proposed here.

Whether the mechanism proposed for thrombin applies to other members of the trypsin family of proteases and zymogens remains to be established by future studies. Residue 215 is highly conserved as Trp in the trypsin family (93% of the cases), but residues 217 and 192 show great variability. Specifically, residue 217 is most commonly Glu (41%), followed by Ser (29%) and Tyr (8%). Residue 192 is most commonly Gln (52%), followed by Glu (28%) and Lys (9%). Trypsin and chymotrypsin both carry W215, but residues 217 and 192 are Tyr and Gln in trypsin or Ser and Met in chymotrypsin. Preliminary analysis of PPACK binding to rat trypsin shows a phase that is too fast to resolve by stopped flow, underscoring a mechanism of recognition that differs from that of thrombin. The presence of E217 and E192 supports a mechanism for the E*-E transition as reported in the present study, but in other cases the E*-E equilibrium may be controlled by different mechanisms or may be frozen in an intermediate conformation as seen for the E217A mutant of thrombin.

A final comment should be made regarding the effect of Na⁺ on the E*-E equilibrium (Fig. 1A). Temperature studies (Fig. 1C) resolve the kinetic rate constants and activation energies associated with the scheme in Eq. 1 and predict an E*:E ratio that changes from 1:6 at 15 °C to 3:1 at 37 °C. Hence, 75% of free thrombin exists in the closed E* form at physiological temperature and would not be able to effectively cleave the procoagulant substrate fibrinogen and the prothrombotic substrate PAR1. Nearly full activity under physiological conditions is ensured by the binding of Na⁺ and conversion of E* to the active E form (Fig. 1A), an important effect that directly opposes the effect of temperature on the E*-E equilibrium. Future studies on cognate trypsin-like proteases that bind Na⁺^{81,82} will reveal if their mechanism of ligand recognition obeys conformational selection with kinetic features similar to those reported here for thrombin and whether Na⁺ plays a similar physiologically important role.

Methods

Reagents. Thrombin wild-type and its mutants W215A, E217 and E192 were expressed as prethrombin-2, purified and activated as previously described^{10,30}. The irreversible inhibitor H-D-Phe-Pro-Arg-CH₂Cl (PPACK) was purchased from Haematological Technologies. PPACK is a relevant probe of the active site as documented by detailed structural information^{27,34}.

Stopped-flow experiments. Rapid kinetic experiments of PPACK binding were conducted on an Applied Photophysics SX20 stopped-flow spectrometer using an excitation of 295 nm and a cutoff filter at 320 nm. The dead time of the mixing cell for this instrument is 0.5–1 ms. Final concentrations of 150–250 nM thrombin wild-type or mutants were used in buffer containing 50 mM Tris, 0.1% PEG8000, 400 mM ChCl, pH 8.0, at 15 °C. The solution containing the protein was mixed 1:1 with 60 μL solutions of PPACK in the same buffer. Baselines were measured by mixing the protein into buffer in the absence of ligand. Each kinetic trace was taken as the average of at least ten determinations and fit to single or double exponentials based on the analysis of residuals using software supplied by Applied Photophysics. Values of the relaxations for single and double exponential fits were derived from at least three independent titrations.

X-ray studies. Crystallization for the human thrombin mutant W215A was achieved at 20 °C by the vapor diffusion technique, using the Art Robbins Instruments PhoenixTM liquid handling robot with 20 mg/ml protein 0.3 μl mixed with an equal volume reservoir solution. Optimization of crystal growth was achieved by the hanging drop vapor diffusion method mixing 3 μl of protein with equal volumes of reservoir solution (Table 1). Crystals were grown in 1 week at 20 °C and frozen in the solution of 10 mM ZnSO₄, 100 mM MES, pH 6.5 and 40% PEG 550 MME. X-ray diffraction data were collected at 100 °K with a home source (Rigaku 1.2 kW MMX007 generator with VHF optics) Rigaku Raxis IV⁺⁺ detector and were indexed, integrated and scaled with the HKL2000 software package⁸³. Structure was solved by molecular replacement using PHASER from the CCP4 suite⁸⁴ and

the structure of slow form of thrombin bound with PPACK (PDB entry 1SHH) as starting model. Refinement and electron density generation were performed with REFMAC5 from the CCP4 suite. 5% of the reflections were randomly selected as a test set for cross-validation. Model building and analysis were carried out using COOT⁸⁵. Twinned crystals were detected and the refinement was performed by twin laws. Ramachandran plot was calculated using PROCHECK⁸⁶. Statistics for data collection and refinement are summarized in Table 2. Atomic coordinates and structure factors have been deposited in the PDB (accession code: 6P9U). A structure of the meizothrombin desF1 mutant S195A bound to PPACK was solved at 2.4 Å resolution (Fig. S1, Table S1) and deposited in the PDB (accession code: 6PX5). The structure offers direct evidence that PPACK acylates H57 even in the presence of the S195A mutation (Fig. S1). However, the irreversible reaction is too slow to resolve within the short time (1–3 ms) of stopped flow measurements that detect binding of PPACK as being reversible and with a finite value of k_{off} (Fig. 1B).

Mechanism of binding. A detailed discussion of ligand binding mechanisms studied by rapid kinetics is given elsewhere^{16,21,45}, and is summarized below for the special case of irreversible binding. The relevant kinetic scheme for ligand binding to the active site of a trypsin-like protease or zymogen is the conformational selection mechanism



E^* and E depict the pre-existing conformations with active site accessible (E) or inaccessible (E^*) to ligand binding that interconvert with first-order rate constants k_{12} and k_{21} . The ratio k_{21}/k_{12} gives the $E^*:E$ partitioning between the two forms (Table 1). The ligand, L , selectively binds to E with a second-order rate of association k_{on} and dissociates with a first-order rate k_{off} . Under conditions where L is in large excess over the macromolecule, the reaction scheme in Eq. 1 gives two independent rates of relaxation to equilibrium according to the expression

$$2\alpha_{1,2} = k_{12} + k_{21} + k_{on}[L] + k_{off} \pm \sqrt{(k_{on}[L] + k_{off} - k_{12} - k_{21})^2 + 4k_{21}k_{on}[L]} \quad (2)$$

The fast relaxation, α_1 (+sign in Eq. 2), reflects the binding event and eventually grows linearly with $[L]$. The slow relaxation, α_2 (– sign in Eq. 2) always saturates for high $[L]$ and reflects the conformational transition associated with binding. Depending on the sign of the expression $k_{off} - k_{12}$, the value of this relaxation hyperbolically decreases ($k_{off} > k_{12}$) or increases ($k_{off} < k_{12}$) with $[L]$, and remains constant when $k_{off} = k_{12}$ ^{16,21,45}. Such property is exclusive of conformational selection and makes this mechanism far more general than induced fit, for which the slow relaxation can only increase hyperbolically with $[L]$. Indeed, the whole kinetic repertoire of induced fit is recapitulated by conformational selection as a mathematical special case²¹, which makes it necessary to distinguish between the two mechanisms every time experimental measurements of α_2 produce a hyperbolic increase with $[L]$. The need becomes especially obvious when dealing with irreversible inhibitors such as PPACK where $k_{off} = 0$. PPACK possesses structural determinants for high affinity binding^{27,34}. Its P1 Arg residue makes a strong ionic interaction with D189 at the bottom of the primary specificity pocket, Pro at the P2 position fits snugly against the hydrophobic surface of the 60-loop and Phe in the D enantiomer at the P3 position makes a strong edge-to-face interaction with W215 defining the western wall of the active site. Replacement of any residue of PPACK or thrombin residues interacting with PPACK significantly compromises binding^{26,87–89}.

When only the slow relaxation α_2 is accessible experimentally, it becomes difficult to resolve all four independent parameters in Eq. 2. The complication is circumvented by measurements carried out as a function of temperature, as originally shown for the analysis of steady state kinetics³⁸. Each rate constant in Eq. 2 can be expressed in terms of its temperature dependence according to the Arrhenius equation

$$k = k_0 \exp \left\{ -\frac{E}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \right\} \quad (3)$$

where k_0 is the value of k at the reference temperature T_0 , E is the activation energy and R the gas constant. When data are collected over a wide enough temperature range, the contribution of the various terms in Eq. 2 change because the values of activation energies differ for processes that involve ligand binding, dissociation and conformational transitions. A global fit of the data resolves all individual rate constants and their associated activation energies, as shown by the results in Fig. 1C.

Data Availability

Recombinant reagents and data presented in this study are available from the corresponding author upon reasonable request.

References

- Hedstrom, L. Serine protease mechanism and specificity. *Chem Rev* **102**, 4501–24 (2002).
- Perona, J. J. & Craik, C. S. Structural basis of substrate specificity in the serine proteases. *Protein Sci* **4**, 337–60 (1995).
- Page, M. J. & Di Cera, E. Serine peptidases: classification, structure and function. *Cell Mol Life Sci* **65**, 1220–1236 (2008).
- Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157–62 (1967).
- Gohara, D. W. & Di Cera, E. Allosteric in trypsin-like proteases suggests new therapeutic strategies. *Trends Biotechnol* **29**, 577–585 (2011).
- Pozzi, N., Vogt, A. D., Gohara, D. W. & Di Cera, E. Conformational selection in trypsin-like proteases. *Curr Opin Struct Biol* **22**, 421–431 (2012).

7. Rohr, K. B. *et al.* X-ray structures of free and leupeptin-complexed human alpha-tryptase mutants: indication for an alpha->beta-tryptase transition. *J Mol Biol* **357**, 195–209 (2006).
8. Niu, W. *et al.* Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry* **50**, 6301–6307 (2011).
9. Wang, D., Bode, W. & Huber, R. Bovine chymotrypsinogen A X-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. *J Mol Biol* **185**, 595–624 (1985).
10. Pozzi, N. *et al.* Crystal structures of prothrombin-2 reveal alternative conformations under identical solution conditions and the mechanism of zymogen activation. *Biochemistry* **50**, 10195–10202 (2011).
11. Bah, A., Garvey, L. C., Ge, J. & Di Cera, E. Rapid kinetics of Na⁺ binding to thrombin. *J Biol Chem* **281**, 40049–40056 (2006).
12. Fersht, A. R. & Requena, Y. Equilibrium and rate constants for the interconversion of two conformations of α -chymotrypsin. The existence of a catalytically inactive conformation at neutral pH. *J Mol Biol* **60**, 279–90 (1971).
13. Lai, M. T., Di Cera, E. & Shafer, J. A. Kinetic pathway for the slow to fast transition of thrombin. Evidence of linked ligand binding at structurally distinct domains. *J Biol Chem* **272**, 30275–82 (1997).
14. Vogt, A. D., Bah, A. & Di Cera, E. Evidence of the E*-E equilibrium from rapid kinetics of Na(+) binding to activated protein C and factor Xa. *J Phys Chem B* **114**, 16125–30 (2010).
15. Vogt, A. D., Chakraborty, P. & Di Cera, E. Kinetic Dissection of the Pre-existing Conformational Equilibrium in the Trypsin Fold. *J Biol Chem* **290**, 22435–45 (2015).
16. Vogt, A. D. & Di Cera, E. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* **51**, 5894–5902 (2012).
17. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci USA* **44**, 98–104 (1958).
18. Eigen, M. Determination of general and specific ionic interactions in solution. *Discuss Faraday Soc* **24**, 25–36 (1957).
19. Chakraborty, P., Acquasaliente, L., Pelc, L. A. & Di Cera, E. Interplay between conformational selection and zymogen activation. *Sci Rep* **8**, 4080 (2018).
20. Huber, R. & Bode, W. Structural basis of the activation and action of trypsin. *Acc Chem Res* **11**, 114–122 (1978).
21. Chakraborty, P. & Di Cera, E. Induced fit is a special case of conformational selection. *Biochemistry* **56**, 2853–2859 (2017).
22. Di Cera, E. Thrombin. *Mol Aspects Med* **29**, 203–254 (2008).
23. Gibbs, C. S. *et al.* Conversion of thrombin into an anticoagulant by protein engineering. *Nature* **378**, 413–6 (1995).
24. Tsiang, M. *et al.* Protein engineering thrombin for optimal specificity and potency of anticoagulant activity *in vivo*. *Biochemistry* **35**, 16449–57 (1996).
25. Arosio, D., Ayala, Y. M. & Di Cera, E. Mutation of W215 compromises thrombin cleavage of fibrinogen, but not of PAR-1 or protein C. *Biochemistry* **39**, 8095–101 (2000).
26. Marino, F., Pelc, L. A., Vogt, A., Gandhi, P. S. & Di Cera, E. Engineering thrombin for selective specificity toward protein C and PAR1. *J Biol Chem* **285**, 19145–19152 (2010).
27. Pineda, A. O. *et al.* Molecular dissection of Na⁺ binding to thrombin. *J Biol Chem* **279**, 31842–31853 (2004).
28. Pozzi, N., Chen, Z. & Di Cera, E. How the Linker Connecting the Two Kringles Influences Activation and Conformational Plasticity of Prothrombin. *J Biol Chem* **291**, 6071–6082 (2016).
29. Pozzi, N. *et al.* Crystal structure of prothrombin reveals conformational flexibility and mechanism of activation. *J Biol Chem* **288**, 22734–22744 (2013).
30. Pozzi, N., Chen, Z., Pelc, L. A., Shropshire, D. B. & Di Cera, E. The linker connecting the two kringles plays a key role in prothrombin activation. *Proc Natl Acad Sci USA* **111**, 7630–7635 (2014).
31. X, M. *et al.* Crystal structure of plasma kallikrein reveals the unusual flexibility of the S1 pocket triggered by Glu217. *FEBS Lett* **592**, 2658–2667 (2018).
32. Gandhi, P. S., Page, M. J., Chen, Z., Bush-Pelc, L. A. & Di Cera, E. Mechanism of the anticoagulant activity of the thrombin mutant W215A/E217A. *J Biol Chem* **284**, 24098–24105 (2009).
33. Carter, W. J., Myles, T., Gibbs, C. S., Leung, L. L. & Huntington, J. A. Crystal structure of anticoagulant thrombin variant E217K provides insights into thrombin allostery. *J Biol Chem* **279**, 26387–26394 (2004).
34. Bode, W., Turk, D. & Karshikov, A. The refined 1.9-Å X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci* **1**, 426–71 (1992).
35. Pozzi, N. *et al.* Loop Electrostatics Asymmetry Modulates the Preexisting Conformational Equilibrium in Thrombin. *Biochemistry* **55**, 3984–94 (2016).
36. Perona, J. J., Hedstrom, L., Rutter, W. J. & Fletterick, R. J. Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* **34**, 1489–99 (1995).
37. Pelc, L. A. *et al.* Why Ser and not Thr brokers catalysis in the trypsin fold. *Biochemistry* **54**, 1457–1464 (2015).
38. Ayala, Y. M. & Di Cera, E. A simple method for the determination of individual rate constants for substrate hydrolysis by serine proteases. *Protein Sci* **9**, 1589–93 (2000).
39. Paul, F. & Weikl, T. R. How to Distinguish Conformational Selection and Induced Fit Based on Chemical Relaxation Rates. *PLoS Comput Biol* **12**, e1005067 (2016).
40. Galletto, R., Jezewska, M. J. & Bujalowski, W. Kinetics of Allosteric Conformational Transition of a Macromolecule Prior to Ligand Binding: Analysis of Stopped-Flow Kinetic Experiments. *Cell Biochem Biophys* **42**, 121–144 (2005).
41. Gianni, S., Dogan, J. & Jemth, P. Distinguishing induced fit from conformational selection. *Biophys Chem* **189**, 33–9 (2014).
42. Halford, S. E. Escherichia coli alkaline phosphatase. *An analysis of transient kinetics*. *Biochem J* **125**, 319–27 (1971).
43. Fuglestad, B. *et al.* The dynamic structure of thrombin in solution. *Biophys J* **103**, 79–88 (2012).
44. Lechtenberg, B. C., Johnson, D. J., Freund, S. M. & Huntington, J. A. NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc Natl Acad Sci USA* **107**, 14087–14092 (2010).
45. Vogt, A. D. & Di Cera, E. Conformational Selection Is a Dominant Mechanism of Ligand Binding. *Biochemistry* **52**, 5723–5729 (2013).
46. Vogt, A. D., Pozzi, N., Chen, Z. & Di Cera, E. Essential role of conformational selection in ligand binding. *Biophys Chem* **186**, 13–21 (2014).
47. Le Bonniec, B. F. & Esmon, C. T. Glu-192-Gln substitution in thrombin mimics the catalytic switch induced by thrombomodulin. *Proc Natl Acad Sci USA* **88**, 7371–5 (1991).
48. van de Loch, A. *et al.* The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *Embo J* **16**, 2977–84 (1997).
49. Tummino, P. J. & Copeland, R. A. Residence time of receptor-ligand complexes and its effect on biological function. *Biochemistry* **47**, 5481–92 (2008).
50. Cantwell, A. M. & Di Cera, E. Rational design of a potent anticoagulant thrombin. *J Biol Chem* **275**, 39827–30 (2000).
51. Berny, M. A. *et al.* Thrombin mutant W215A/E217A acts as a platelet GpIb antagonist. *Arterioscler Thromb Vasc Biol* **18**, 329–334 (2008).
52. Berny-Lang, M. A. *et al.* Thrombin mutant W215A/E217A treatment improves neurological outcome and reduces cerebral infarct size in a mouse model of ischemic stroke. *Stroke* **42**, 1736–1741 (2011).
53. Gruber, A., Cantwell, A. M., Di Cera, E. & Hanson, S. R. The thrombin mutant W215A/E217A shows safe and potent anticoagulant and antithrombotic effects *in vivo*. *J Biol Chem* **277**, 27581–4 (2002).

54. Gruber, A. *et al.* Limited generation of activated protein C during infusion of the protein C activator thrombin analog W215A/E217A in primates. *J Thromb Haemost* **4**, 392–7 (2006).
55. Gruber, A. *et al.* Relative antithrombotic and antihemostatic effects of protein C activator versus low molecular weight heparin in primates. *Blood* **109**, 3733–3740 (2007).
56. Tanaka, K. A. *et al.* Interaction between thrombin mutant W215A/E217A and direct thrombin inhibitor. *Blood Coagul Fibrinolysis* **19**, 465–8 (2008).
57. Verboon, N. G. *et al.* Thrombin mutant W215A/E217A treatment improves neurological outcome and attenuates central nervous system damage in experimental autoimmune encephalomyelitis. *Metab Brain Dis* (2014).
58. Wood, D. C. *et al.* WEDGE: an anticoagulant thrombin mutant produced by autoactivation. *J Thromb Haemost* **13**, 111–4 (2015).
59. Pineda, A. O. *et al.* The anticoagulant thrombin mutant W215A/E217A has a collapsed primary specificity pocket. *J Biol Chem* **279**, 39824–8 (2004).
60. Edgington, T. S., Mackman, N., Brand, K. & Ruf, W. The structural biology of expression and function of tissue factor. *Thromb Haemost* **66**, 67–79 (1991).
61. Banner, D. W. *et al.* The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **380**, 41–6 (1996).
62. Forneris, F. *et al.* Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science* **330**, 1816–20 (2010).
63. Craik, C. S., Page, M. J. & Madison, E. L. Proteases as therapeutics. *Biochem J* **435**, 1–16 (2011).
64. Di Cera, E. Thrombin as an anticoagulant. *Prog Mol Biol Transl Sci* **99**, 145–84 (2011).
65. Bunce, M. W., Toso, R. & Camire, R. M. Zymogen-like factor Xa variants restore thrombin generation and effectively bypass the intrinsic pathway *in vitro*. *Blood* **117**, 290–8 (2011).
66. Lancellotti, S., Basso, M. & De Cristofaro, R. Congenital prothrombin deficiency: an update. *Semin Thromb Hemost* **39**, 596–606 (2013).
67. Bertina, R. M. Factor V Leiden and other coagulation factor mutations affecting thrombotic risk. *Clin Chem* **43**, 1678–83 (1997).
68. Gawlik, K. *et al.* Autocatalytic activation of the furin zymogen requires removal of the emerging enzyme's N-terminus from the active site. *PLoS One* **4**, e5031 (2009).
69. Arstenstein, A. W. & Opal, S. M. Proprotein convertases in health and disease. *N Engl J Med* **365**, 2507–18 (2011).
70. Piper, D. E. *et al.* The crystal structure of PCSK9: a regulator of plasma LDL-cholesterol. *Structure* **15**, 545–52 (2007).
71. Yamamoto, E., Kitano, Y. & Hasumi, K. Elucidation of crucial structures for a catechol-based inhibitor of plasma hyaluronan-binding protein (factor VII activating protease) autoactivation. *Biosci Biotechnol Biochem* **75**, 2070–2 (2011).
72. Sichler, K. *et al.* Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J Mol Biol* **322**, 591–603 (2002).
73. Whitcomb, D. C. *et al.* Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* **14**, 141–5 (1996).
74. Stirnberg, M. *et al.* Proteolytic processing of the serine protease matriptase-2: identification of the cleavage sites required for its autocatalytic release from the cell surface. *Biochem J* **430**, 87–95 (2010).
75. Pozzi, N. *et al.* Autoactivation of thrombin precursors. *J Biol Chem* **288**, 11601–11610 (2013).
76. Pozzi, N., Barranco-Medina, S., Chen, Z. & Di Cera, E. Exposure of R169 controls protein C activation and autoactivation. *Blood* **120**, 664–670 (2012).
77. Wakeham, N. *et al.* Effects of deletion of streptokinase residues 48–59 on plasminogen activation. *Protein Eng* **15**, 753–61 (2002).
78. Friedrich, R. *et al.* Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation. *Nature* **425**, 535–9 (2003).
79. Landgraf, K. E. *et al.* Allosteric peptide activators of pro-hepatocyte growth factor stimulate Met signaling. *J Biol Chem* **285**, 40362–72 (2010).
80. Landgraf, K. E. *et al.* An allosteric switch for pro-HGF/Met signaling using zymogen activator peptides. *Nat Chem Biol* **10**, 567–73 (2014).
81. Dang, Q. D. & Di Cera, E. Residue 225 determines the Na(+)-induced allosteric regulation of catalytic activity in serine proteases. *Proc Natl Acad Sci USA* **93**, 10653–6 (1996).
82. Krem, M. M. & Di Cera, E. Molecular markers of serine protease evolution. *Embo J* **20**, 3036–45 (2001).
83. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected by oscillation methods. *Methods Enzymol.* **276**, 307–326 (1997).
84. Dodson, E. J., Winn, M. & Ralph, A. Collaborative Computational Project, number 4: providing programs for protein crystallography. *Methods Enzymol* **277**, 620–33 (1997).
85. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–32 (2004).
86. Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345–64 (1992).
87. Vindigni, A., Dang, Q. D. & Di Cera, E. Site-specific dissection of substrate recognition by thrombin. *Nat Biotechnol* **15**, 891–5 (1997).
88. Krem, M. M. & Di Cera, E. Dissecting substrate recognition by thrombin using the inactive mutant S195A. *Biophys Chem* **100**, 315–23 (2003).
89. Butenas, S., DiLorenzo, M. E. & Mann, K. G. Ultrasensitive fluorogenic substrates for serine proteases. *Thromb Haemost* **78**, 1193–201 (1997).

Acknowledgements

We are grateful to Dr. Pradipta Chakraborty for her contributions during early stages of this investigation and to Dr. David Gohara for the analysis of conservation of residues 215, 217 and 192. This work was supported in part by the National Institutes of Health Research Grants HL049413, HL139554 and HL147821.

Author Contributions

Z.C., S.K.K., L.A.P. and E.D.C. designed the research and analyzed the data; Z.C., N.E.G., S.K.K. and L.A.P. performed the research; E.D.C. wrote the manuscript. All Authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-48839-1>.

Competing Interests: E.D.C. has a financial interest in Verseon Corporation.

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



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

© The Author(s) 2019