Published in final edited form as:

Phys Chem Chem Phys. 2017 September 20; 19(36): 24522–24533. doi:10.1039/c7cp03671j.

Mechanistic insights into thrombin's switch between “slow” and “fast” forms[†]

Jiajie Xiao^{a,b}, Ryan L. Melvin^{a,c}, and Freddie R. Salsbury Jr.^{*,a}

^aDepartment of Physics, Wake Forest University, Winston Salem, NC, USA

^bDepartment of Computer Science, Wake Forest University, Winston Salem, NC, USA

^cDepartment of Mathematics and Statistics, Wake Forest University, Winston Salem, NC, USA

Abstract

Thrombin is a multifunctional enzyme that plays an important role in blood coagulation, cell growth, and metastasis. Depending upon the binding of sodium ions, thrombin presents significantly different enzymatic activities. In the environment with sodium ions, thrombin is highly active in cleaving the coagulated substrates and this is referred to as the “fast” form; in the environment without sodium ions, thrombin turns catalytically less active and is in the “slow” form. Although many experimental studies over the last two decades have attempted to reveal the structural and kinetic differences between these two forms, it remains vague and disputed how the functional switch between the “fast” and “slow” forms is mediated by Na⁺ cations. In this work, we employ microsecond-scale all-atom molecular dynamics simulations to investigate the differences in the structural ensembles in sodium-bound/unbound and potassium-bound/unbound thrombin. Our calculations indicate that the regulatory regions, including the 60s, γ loops, and exosite I and II, are primarily affected by both the bound and unbound cations. Conformational free energy surfaces, estimated from principal component analysis, further reveal the existence of multiple conformational states. The binding of a cation introduces changes in the distribution of these states. Through comparisons with potassium-binding, the binding of sodium ions appears to shift the population toward conformational states that might be catalytically favorable. Our study of thrombin in the presence of sodium/potassium ions suggests Na⁺-mediated generalized allostery is the mechanism of thrombin's functional switch between the “fast” and “slow” forms.

[†]Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp03671j
salsbufr@wfu.edu; Fax: +1 (336) 758 6142; Tel: +1 (336) 758 4975.

Note added after first publication

This article replaces the version published on 29 August 2017 to correct editorial errors introduced following acceptance of the paper. On page 10, left column, 8 lines from the bottom, the text “To clarify why no significantly different structures of thrombin in “slow” and “slow” forms have been solved yet,...” has been corrected to “To clarify why no significantly different structures of thrombin in “fast” and “slow” forms have been solved yet,...”.

On page 10, right column, line 12, the sentence “This stabilized configuration is likely a favorable factor for catalytic efficiency and a critical kinetic signature of the “slow” form thrombin.” has been corrected to “This stabilized configuration is likely a favorable factor for catalytic efficiency and a critical kinetic signature of the “fast” form thrombin.”

On page 10, right column, Conclusions section, the last sentence of the first paragraph has been corrected from “This stabilization may cause the functional switch between the “slow” and “slow” forms of thrombin.” to “This stabilization may cause the functional switch between the “slow” and “fast” forms of thrombin.”

Conflicts of interest

There are no conflicts to declare.

1 Introduction

Thrombin is an attractive drug target for antithrombotic therapy and chemotherapeutic development due to its critical role in blood coagulation,^{1,2} as well as in cancer cell growth and metastasis.³ As the final protease in the blood coagulation cascade, thrombin regulates both coagulant and anticoagulant processes.¹ In the coagulation cascade, activated from its zymogen prothrombin, thrombin facilitates the formation of blood clots by interacting with its procoagulant substrates. For example, thrombin converts soluble fibrinogen into insoluble fibrin and activates factor XIII and protease-activated receptors (PARs), which serve to generate the fibrin meshwork and platelet plug at the site of the wound.⁴ Thrombin can also activate factor V and VIII to stimulate its own conversion from its precursor prothrombin.⁵ In the anticoagulant process, thrombin acts as an anticoagulant through binding to the endothelial cell receptor thrombomodulin (TM) and activating protein C zymogen. The interaction between thrombin and thrombomodulin not only suppresses thrombin's cleavage of the fibrinogen and PARs, but also increases thrombin's specificity to the anticoagulant protein C. The product from thrombin's activation, activated protein C, then proteolytically inactivates the activated factor V and VIII and limits the further generation of thrombin. The balance between bleeding and thrombosis requires an efficient switch between the pro- and anticoagulant actions of thrombin.

In addition to the relatively widely known functions in blood coagulation cascade, many recent studies have shown that thrombin has a critical role in the development of tumors and introducing invasion and metastasis in a variety of cancers.^{3,6–8} For instance, thrombin facilitates the growth of colon cancer cells by mediating the proteolysis of PAR1 and fibrinogen.⁶ As a result, less favorable interactions between thrombin and these procoagulant substrates (*i.e.* PAR-1 and fibrinogen) may lead to an increased tumor necrosis and decreased tumor vascular density. Therefore, understanding the molecular properties of thrombin – multifunctional allosteric enzyme – can aid in developing further treatments against cancers as well.

Along with several associations between thrombosis and abnormal levels of sodium in the blood,^{9–12} many studies have suggested that the paradoxically opposing functions of thrombin are regulated by the univalent cation sodium (Na^+).^{13–16} Orthner and Kosow first showed in 1980 that the monovalent cation Na^+ optimally enhances thrombin's hydrolysis rate of various synthetic and physiological substrates, such as peptidyl *p*-nitroanilide substrates and fibrinogen.¹³ An accelerating PARs cleavage by thrombin^{15,16} and a tighter TM affinity to thrombin¹⁴ has also been seen in the presence of Na^+ . Although different influences of Na^+ on protein C's hydrolysis by thrombin were reported,^{14,17} thrombin generally cleaves substrates at a faster rate when Na^+ is present.

To explain the different activities of thrombin under sodium-present and sodium-absent environments, Wells and Di Cera proposed a two-state model in which thrombin can be allosterically activated by the binding of Na^+ and thrombin has two distinct conformations – high activity “fast” form (Na^+ -bound) and a low activity “slow” form (Na^+ -unbound).¹⁸ Dang *et al.* further demonstrated that the significantly populated “fast” and “slow” forms (~3 : 2 ratio) are preferred by coagulant and anticoagulant activities respectively.¹⁷ The Na^+ -

binding site was later identified as being about 15 Å away from the catalytic triad of thrombin^{19,20} (Fig. 1), partially supporting the hypothesis that the bound-Na⁺ is an allosteric effector of thrombin. However, structural studies in the past decades suggest that the “slow” form of thrombin surprisingly has no significant conformational change compared to the “fast” form.^{20–22}

Huntington argued that the structure of the thrombin in the “slow” form is barely achieved due to the experimental limitations such as crystal contact effects and ligand’s stabilization of the “fast” form.²² Nuclear magnetic resonance (NMR) spectra indicate thrombin is somewhat “plastic” with respect to the dynamics of surface loops.²³ Huntington thus hypothesized that, in contrast to the “fast” form, the “slow” form thrombin needs more time to sample more states including the protease-like and zymogen-like states.²⁴ These experimental results and explanations trigger our interest in investigating which model – the static two-state model or the dynamic multi-state model – reveals the mechanisms of thrombin’s functional switch.

Computational modeling has emerged as an insightful and inspiring way to understand the patterns behind various phenomena,^{25–29} as well as an aid in drug discovery and design.^{27,30–32} In particular, molecular dynamics (MD) simulations have proven useful in providing a structural, dynamical and kinetic picture of biomolecules at a high temporal and spatial resolution that current experimental techniques may not always achieve.^{32–36} In a MD simulation, starting from the experimentally determined static structures of the biomolecules, each atom’s coordinates in the simulated system are updated according to classical mechanics. The parameters of the empirical force fields involved in the calculations have been tuned and optimized for many decades.^{37,38} Therefore, MD studies on thrombin may offer a closer look at thrombin’s structural ensemble and dynamical behaviors.

The application of MD simulations on thrombin can be traced back as early as the 1990s. Ayala and Di Cera inferred a potential induce-fit in thrombin–hirudin association by showing the isolated thrombin and hirudin have larger root mean square deviations (RMSD) to their crystal structures than their complex form in their 5 ps MD simulations.³⁹ More recently, the MD study by de Amorim *et al.* in 2010, indicates that non-liganded thrombin – including Na⁺-unbound thrombin – has a smaller global solvent-accessible surface area than the liganded thrombin through the statistics in the last 5 ns of their MD simulations up to 10 ns.⁴⁰ Furthermore, Fuglestad *et al.* combined NMR experiments and both 20 ns time scaled conventional and accelerated MD calculations to reveal the surface loops of the PPACK-inhibited thrombin exhibit slow time scale dynamics in addition to the fast dynamic motion at the ps–ns timescale,⁴¹ suggesting longer time scale MD simulations are required to study the influence of the ligand and ion binding to thrombin. Our previous microsecond-scaled simulations have illustrated that thrombin exhibits a generalized allosteric response to the binding of thrombin aptamer.²⁹ This not only reveals the mechanisms of thrombin’s molecular recognition to the aptamer binding but also helps understand thrombin’s response to the potential nucleic acid treatment at the molecular level.

In this work, we performed microsecond-scaled all-atom MD simulations on unliganded thrombin in Na⁺-present and Na⁺-free environments, aiming to identify the differences in

thrombin's structural ensembles in the Na⁺-present/absent and Na⁺-bound/unbound situations and shed light on the mechanisms of the dual functional switch of thrombin.

2 Materials and methods

2.1 Study design and simulated systems

The switch between the “fast” and “slow” thrombin has been hypothesized to be allosterically regulated by the binding of Na⁺ to thrombin, even though all thrombin structures solved so far are fairly similar.²² To elucidate the mechanism of the functional switch of thrombin, we proposed to simulate thrombin in the solvent environments with the presence and absence of Na⁺. During our microsecond-scaled MD simulations, the relaxation in the solution should alleviate the potential experimental drawbacks due to the stabilization of the “fast” form from crystal contacts and ligation locks. Comparisons between the ensembles of thrombin in Na⁺-free and Na⁺-present environments in the simulations offer a comprehensive – thermodynamic – examination of the conformational changes due to ion effects, which may clarify the structural and dynamical differences in the “fast” and “slow” forms.

In the Na⁺-present and Na⁺-free simulations, we chose the initial structures of thrombin from PDB 4DIH and 4DII⁴² respectively. The high-resolution – 1.8 Å for PDB 4DIH and 2.05 Å for PDB 4DII – structures of thrombin in these two PDBs were crystallized and solved from buffers with 0.125 M Na⁺ and K⁺ respectively. Despite the argument that the fast form thrombin could be stabilized by the binding ligands,⁴³ all available thrombin structures – including unliganded thrombin as well as the Na⁺-free ones – are actually very similar. Therefore, under relatively long time scale (up to microseconds), without a loss of generality, we started our simulations of unliganded thrombin in Na⁺-free and Na⁺-present environments by removing the aptamer in those PDBs.

Although there is a reported affinity for K⁺ ions to thrombin and an enhancement of thrombin activity under higher concentration of K⁺,¹³ it has been shown that the K⁺ has an order of magnitude weaker affinity for thrombin ($K_A = 0.28$ M) than Na⁺ ($K_A = 0.035$ M)¹³ and that K⁺ and other alkali metal ions will not affect thrombin's structure as Na⁺ does.⁴⁴ Accordingly, the “fast” and “slow” forms may be largely characterized through simulations with sodium chloride and potassium chloride buffers respectively. On the other hand, it has been hypothesized that the “fast” form is corresponding to thrombin with Na⁺-binding and the “slow” form is related to the Na⁺-unbound thrombin.¹⁸ Nevertheless, comparisons among the structural ensembles of the Na⁺-bound/unbound and K⁺-bound/unbound thrombin will help understand the ion effects on thrombin and provide insight into thrombin's conformational and functional switch.

2.2 MD simulations

Using the GPU-enabled ACEMD simulation package⁴⁵ and Titan GPUs in Metrocubo workstations produced by Acellera, we performed five one-microsecond-long all-atom MD simulations for each system, *i.e.* a total of 5 microseconds for thrombin in NaCl buffer and 5 microseconds for thrombin in KCl buffer respectively. As discussed above, the initial

structures of the Na⁺-environment and K⁺-environment simulations were based on the thrombin coordinates in PDB 4DIH and 4DII⁴² respectively. The initial coordinates of the missing residues in these two PDBs (19 and 18 out of 295 residues respectively for PDB 4DIH and 4DII) were modeled *via* the comparative modeling tool Modeller,⁴⁶ and other missing hydrogen atoms were added through VMD's psfgen package,⁴⁷ using default parameters. Although progress has been made in computationally estimating the of pK^a shifts and protonation states,^{48–54} accurate assignment of the protonation states is still challenging in general.^{48,55–57} To balance the reliability and conformational sampling, and from visual assessments of possible hydrogen bonds on the X-ray structure, all ionizable residues were considered in their default protonation state at physiological pH. In particular, all histidine residues were assigned delta protonation (meaning a proton H⁺ was added onto delta nitrogen), since under the physiological conditions the delta protonation form has been suggested more stable than the epsilon protonation one.^{58–60}

For both systems, an explicit TIP3P⁶¹ water box with a 10 Å padding in *x*, *y*, and *z* directions was added around thrombin. Then sodium chloride and potassium chloride were respectively added to neutralize the systems and the Na⁺ and K⁺ concentration was set to a typical experimental value of 0.125 M, resulting in a total number of 22 Na⁺ and 21 K⁺ in each system. The ions were randomly placed around thrombin and a minimum distance of 5 Å was set as a requirement to avoid a too close initial placement to the water molecules and other ions. All simulations were performed with the CHARMM22/CMAP force field^{62,63} under *NPT* ensemble (constant particle number, temperature and pressure). The pressure was maintained at 1 atm using Berendsen pressure control⁶⁴ with a 400 fs relaxation time. The temperature was held at 300 K using Langevin thermostat⁶⁵ with a damping coefficient of 0.1. For van der Waals and electrostatic forces, we applied a 9 Å cutoff and 7.5 Å switching distance. The particle mesh Ewald implementation^{45,66} with 72 evenly spaced grid points in all three directions was used when calculating the long-range electrostatics. As a hydrogen mass repartitioning scheme⁶⁷ was used, the time step was set to 4 fs during the simulations. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm.⁶⁸ Conformations sampled in the simulations were saved in the MD trajectory file in every 2500 steps for the balance of temporal resolution and cost in storage. Therefore, we have a extensive sampling of 100 000 conformations across one-microsecond time scale for each run (500 000 conformations in total for each buffer condition).

2.3 Processing and analysis

Prior to analysis, we aligned all frames in MD trajectories to the first frame of the simulations in the potassium environment *via* rigid body rotations and translations to minimize the root-mean-square distance (RMSD) of thrombin alpha carbon positions. The alignment was carried out with the 'align' class in MD Analysis,^{69–72} a Python library for manipulating MD data. As the equilibrium can reach within ~20 nanoseconds (less than 2% of simulation time) for all simulation runs according to thrombin's root-mean-squared distances to the corresponding initial structure in each type of ion environments, acquiring statistical information based on the whole microsecond-scaled MD trajectories, including the equilibrium and production runs, will not introduce considerable errors. For the convenience of data processing, we performed a variety of structural analyses on the whole trajectory for

each system (5 microseconds – comprised of 500 000 frames – in total for each system). Note that the residue sequence numbers start from negative integers in the PDBs 4DIH and 4DII and different residues can have the same residue number with extra codes for insertion of residues. For the convenience of recalling residues in the following context, we followed the same sequential residue numbering scheme adopted by previous studies^{41,73} to renumber residue indexes of the light (36 residues) and heavy chains (259 residues) of thrombin from 1 to 295. As a result, the catalytic triad of His57, Asp102, and Ser195 in the chymotrypsin numbering scheme have residue indexes (denoted as ‘resid’) 79, 135, and 241 in the sequential residue numbering scheme.

2.3.1 Ion-binding detection—We followed the same approach previously described by Zhang *et al.*⁷⁴ to quantify ion binding to thrombin, which considers how close and long the nearest ion can remain interacting with the sodium binding loop (*i.e.* 220s loop, resid 263 to 272). For each frame, we computed pairwise distances between the sodium binding loop atoms and selected type of ions. The time series of the minimum distance d between the 220s loop atoms and ions was then used as inputs for probability and temporal distribution calculations. Histograms of such d illustrate the distribution of how adjacent an ion is located on the sodium binding loop. The temporal distributions of d , which were estimated by counting how long the ion stays within a range of distance d in average, depict how stable the ion can interact with the sodium binding loop.

The stably bound ions should establish peaks in both probability and temporal distribution of d . Therefore, the ion-bound and ion-unbound states were defined based on the distance d . Consistent with relevant literature,⁷⁴ a frame of cation-binding (also called cation-on) state was identified if the value of d in that frame was not greater than 3.8 Å. Such approach should be able to identify the exterior ion binding (*i.e.* the ion bounds on the exterior surface of the 220s loop) and the interior one (*i.e.* the ion is trapped in the 220s loop). To compare the populations of these two types of binding, we also computed the mean distance between the 220s loop atoms and the nearest Na⁺/K⁺ ion. Since such mean distance quantifies how far the nearest metal ion is relative to the whole sodium binding loop, the much more stable interior binding can be separated out from the exterior one based on the mean distance’s probability and temporal distributions. In the following context, we conducted our analyses using the ion-bound and ion-unbound states defined based on the minimum distance measurements, which includes both exterior and interior bindings onto the sodium binding loop.

2.3.2 Identification of responses to ion-binding—As a first line to highlight the regions showing different response upon ion-binding/unbinding, the atomic fluctuations of thrombin were evaluated based upon the alpha carbon atoms. We computed their root-mean-square fluctuations (RMSF) across all conformations in the related states based on

$$\text{RMSF}^{\alpha} = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(\overleftarrow{r}_t^{\alpha} - \langle \overleftarrow{r}_t^{\alpha} \rangle \right)^2}, \quad (1)$$

where $\overleftarrow{r}_t^\alpha$ is the position vector of atom α in frame t and $\langle \overleftarrow{r}_t^\alpha \rangle$ denotes the average position vector of atom α over all frames.

To identify the similarities and differences in structural ensembles, we concatenated all trajectories (ten MD trajectories in total) together and clustered the conformations we sampled in both Na⁺ or K⁺ environments. To save the computation cost, here we actually clustered the conformations in every tenth frames (*i.e.* 100 000 out of 1 000 000 frames in total were sampled for clustering).

According to the potentially affected regions suggested by the RMSF comparison, the clustering analysis was performed on the regulatory regions of thrombin, which explicitly includes the exosite I (resid 57 98 104 106 109 142 143) and II (resid 125 134 281 284 288), 60s (resid 182 to 190), 220s (resid 263 to 272), and γ (resid 82 to 91) loops. All heavy atoms (*i.e.* not hydrogen) were used in the clustering analysis to take into account of the side chain fluctuations. The clustering algorithm we utilized here was the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN),⁷⁵ which has been adapted and employed as a non-parametric approach to group high dimensional MD data by our research group.^{76,77} This clustering algorithm has been shown the effectiveness of identifying global conformational changes⁷⁶ and thereby was employed here as well to evaluate ion-binding's influence on the structural ensembles.

The solvent accessible surface areas (SASA) of the residues forming catalytic triad and sub-pockets (S1–6) and the whole protein molecule were evaluated to investigate the molecular response to Na⁺/K⁺-binding. The solvent accessible surface areas (SASA) of the residues forming catalytic sub-pockets were computed *via* the Shrake and Rupley algorithm⁷⁸ – as implemented by the ‘shrake_rupley’ function in MDTraj⁷⁹ – to estimate the surface areas through counting the number of the sphere points representing the surface of each atom. The residues defining the catalytic sub-pockets were listed in detail in Table S1 (ESI[†]). To save computing time, the SASAs were measured in every tenth frames as what we sampled in the clustering analysis. Welch's *t*-tests⁸⁰ (not assuming equal variance) were carried out to test for statistically significant differences in the distributions of SASA. The Welch's *t*-tests were performed *via* the ‘ttest2’ function in Matlab. The significance level was set as 1% in the Welch's *t*-tests.

To illustrate thrombin's conformational space in terms of free energy surfaces, we employed principal component analysis (PCA) to project the high dimensional structural information into a reduced conformational space. By diagonalizing the covariance of the heavy atoms in the regulatory regions defined in the clustering analysis, we obtained a few components capturing the majority of dynamic variance of the regulatory regions of thrombin in both Na⁺ and K⁺ environments. By projecting the coordinates of heavy atoms in the regulatory regions in the aligned trajectories onto the eigenvectors with two largest eigenvalues, the different structures are expected to have distinguished projections in the reduced

[†]Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp03671j

conformational space. Through binning the projections and converting the frequency P of each bin into free energy G via

$$\Delta G = -kT \ln \left(\frac{P}{P_0} \right), \quad (2)$$

where kT is the product of Boltzmann's constant k and temperature T and the minimum frequency P_0 among all bins was subtracted in order to obtain the relative free energy change

G , we constructed the conformational free energy surfaces of the regulatory regions. Furthermore, using the same approach, we also constructed the conformational free energy surfaces of the residues involved in the catalytic pocket, including catalytic sub-pockets S1 to S6 (Table S1, ESI[†]) and catalytic triad (resid 79 135 241). The PCA decomposition was carried out via the 'pca' module in PyEmma⁸¹ Python package.

The visualization of the structural ensembles in the conformational free energy wells was performed via the molecular graphics program VMD 1.9.2.⁴⁷ With a consideration of the variance of the representative structure,⁸² we visualized the structural ensembles by showing the representative structure in solid and the other conformations in the same structural ensemble as shadows. In particular, we picked the closest structure to the average conformations within the most populated bin in the selected well as the representative structure. For the representative structures, we also displayed the side chains of the residues of interest. For the shadows, which were randomly selected 50 structures in the same bin to indicate the variance, we only presented the backbone using the new cartoon representation for simplicity of the image.

3 Results

3.1 Ion association and dissociation can be observed for Na⁺ and K⁺ rather than Cl⁻ in the simulations

We first monitored the potential ion association and dissociation throughout MD trajectories, as the presence of Na⁺ is the key apparent factor that regulates the activity of thrombin. The calculations on probability distribution of the minimum distance between the sodium binding loop or sodium loop and its nearest ions show that only sodium and potassium cations present significant peaks corresponding to ion binding (Fig. 2a). This indicates that both sodium and potassium cations are able to recognize and remain around the 220s loop, whereas the negatively charged chlorine ion is barely located near the 220s loop. Na⁺ ions establish two peaks (at 2.2 Å and 3.4 Å), suggesting two different strengths of Na⁺ association on the sodium binding loop. For K⁺, there is only one major peak (at 2.7 Å) representing K⁺ binding. The higher peaks for the Na⁺ suggest Na⁺ is more likely to stay near the 220s loop residues than K⁺, implying a stronger affinity of Na⁺ than K⁺. This is consistent with the relationships between the association and dissociation constants of Na⁺ and K⁺, which are well-determined through various experimental approaches.^{13,18}

Furthermore, as shown in Fig. 2b, the temporal distributions under the bin width of 0.8 Å show that Na⁺ has a peak of 883 ps residence time within distance range (1.7 Å d < 2.5 Å) and another subtle peak of 28 ps within (3.4 Å d < 4.2 Å), indicating the second binding peak (at d = 3.39 Å on Fig. 2a) is kinetically more transient than the tighter Na⁺ binding (at d = 2.2 Å on Fig. 2a). K⁺ present a residence time peak of 42 ps within (2.6 Å d < 3.4 Å) (Fig. 2b). Although these values may vary depending on the bin width – *e.g.* doubling the bin width results in that Na⁺ and K⁺ present a residence time of 1131 ps and 713 ps within (1.8 Å d < 3.4 Å) (Fig. S1, ESI[†]), the temporal distributions of Na⁺ and K⁺ demonstrate the binding of a Na⁺ ion is thermodynamically and kinetically more stable than a K⁺ ion.

As quantitatively assessed by Table 1, the sodium loop of thrombin has a bound Na⁺ over 69.18% simulation time, which is a multiple of 1.6 more frequent than K⁺-binding. Using the mean distance as the inputs for probability and temporal distribution calculations, one can see that K⁺ ions have a higher percentage of exterior binding than Na⁺ ions do. Note that ratio of the Na⁺-bound and Na⁺-unbound thrombin under the physiological condition is presumably 3 : 2 according to the experimental determination of equilibrium constant K_d .¹⁷ Such close fractions of Na⁺-bound and Na⁺-bound thrombin suggest the validity of our simulations to reflect thrombin's *in vivo* behaviors and properties.

3.2 Na⁺-Binding stabilizes the regulatory regions on thrombin

To highlight the regions with significant conformational changes due to ion effects, atomic fluctuation was evaluated by calculating the root-mean-square fluctuations (RMSF) on the alpha-carbons. As shown in the RMSF plots of the Na⁺-bound/unbound and K⁺-bound/unbound thrombin (Fig. 3), the global ion environment and the local ion binding status have primary influences on specific functional sites.

Loops on the surface of thrombin generally exhibit greater fluctuations in the potassium environment than in the sodium one. In particular, no matter if the K⁺ ions are bound or unbound, the γ loop (resid 182 to 190) displays a significantly larger fluctuation than it does in Na⁺ simulations. Similarly, residues 147 to 149 that are adjacent to the exosite I residues on the same surface loop also present about 0.8 Å more fluctuations in K⁺ simulations.

On the other hand, both of the Na⁺ and K⁺ bindings to the 220s loop (sodium binding site) seem to result in stabilizing effects on the 60s (resid 83 to 89), 180s (resid 226 to 235), and 220s loop. Furthermore, the Na⁺ binding causes a considerable stabilization on the 70s loop (resid 103 to 111) that contains the exosite I residues and the alpha helix (resid 15 to 21) in the light chain. The alpha helix and turn region (resid 160 to 169) spatially proximate to the exosite II show particularly a higher mobility upon Na⁺-unbinding, while the 60s loop (resid 83 to 89) are much more flexible upon K⁺-unbinding.

As seen in Fig. 3, the global flexibilities of the catalytic triad do not appear directly affected by the bound and unbound ions. Instead, mainly the regulatory loops such as the 60s and γ loops, as well as the regions that are spatially adjacent to the exosite I and II, present different atomic fluctuations upon different ion conditions. Therefore, clustering analysis was performed on these regulatory regions to identify structural ensembles corresponding to Na⁺-bound/unbound and K⁺-bound/unbound states. Using a non-parametric clustering tool

we have recently developed based on HDSCAN algorithm,^{75,76,82} the conformations of the thrombin in Na⁺-bound/unbound and K⁺-bound/unbound states are classified into different groups due to their structural similarity.

From the cluster label of the structure of thrombin at each snapshot (Fig. 4), it is clearly seen that thrombin has much more diverse structures in the environment lacking sodium ions. Quantitatively, thrombin establishes 5 non-singleton conformational clusters of the regulatory regions in the presence of Na⁺, whereas 91 clusters could be observed in the K⁺ environment. This indicates the regulatory regions, including the 60s and γ loops and exosite I and II, are selectively stabilized by the Na⁺ ions.

Furthermore, our clustering result suggests, instead of the two-state model proposed by Wells and Di Cera in 1992,¹⁸ more than two conformational states of thrombin may exist. Based on the visualizations of the structural ensembles with a consideration of the variance of the representative structure,⁸² the multi-state model proposed by Huntington is more supported by our simulations since the 60s and γ loops have obviously presented varied poses above and beyond the catalytic triad (Fig. S3, ESI[†]).

3.3 Ion environments impact solvent accessible surface of the catalytic sub-pockets and the whole thrombin molecule

Regarding different conditions of ion binding/unbinding, the distributions of the solvent accessible surface area (SASA) of the catalytic sub-pockets S1–6, catalytic triad, and the entire thrombin molecule were plotted in Fig. 5a–h respectively.

Depending on the ion types, the unbound cations present different impacts on the SASA of the catalytic pockets (S1–6), catalytic triad and even the whole protein molecule (Fig. 5). The substrate pockets S1–6 present wider distributions in the environment with K⁺ ions, implying a more stable catalytic cleft in the Na⁺ environment. While most of these SASA distributions have only one central peak, the S2, S4, S5, S6 and catalytic triad present a second peak in the presence of K⁺ ions. Such double peaks in the K⁺ simulations suggest there are two significantly distinct conformations of involved residues. In the presence of K⁺ ions, the SASA distributions of the S2 pocket and the whole protein both have a shift to larger SASA values, indicating higher accessibility to the molecules in the solvent.

Moreover, as shown in Fig. 5, the binding of cations appear to have a visible influence on the S3–6 pockets. For instance, based on the positions of related central peaks, S3 and S4 pockets tend to become less exposed to the solvent after the binding of Na⁺ or K⁺. The S5 and S6 pockets appear relatively more stable when a Na⁺ or K⁺ ion is bound.

With at least 10 000 samples for SASA in each ion binding/unbinding condition, Welch's *t*-tests⁸⁰ were carried out for to determine the statistical significance of differences in the distributions of SASA, as shown in Fig. 5. Only 4 such pairs failed the significance test at a 1% level, with many having a much lower *p*-value as shown in Table S2 (ESI[†]).

According to Table S2 (ESI[†]), regardless of comparing the effects from the ion binding/unbinding or the ion types, S1, S3, S4, catalytic triad, and the whole protein establish statistically different SASA distributions. Only 4 ion types or ion bindings do not yield a

statistically significant difference in the SASA distributions of sub-pockets. These are S1 (NaOn vs. NaOff), S2 (NaOn vs. NaOff), S3 (NaOn vs. NaOff), S3 (KOn vs. KOff), using the labeling given in Table S2 (ESI[†]).

Our temporal distribution analysis on these SASA time series further suggests the influence of ion environments on SASA is not just thermodynamic but also kinetic. For example, the whole catalytic pocket can have two relatively long-lived meta-stable states in the environment of Na⁺ (Fig. S2(a), ESI[†]). One of these two states results in a relatively stable open catalytic pocket in the presence of Na⁺ molecules, while catalytic pocket fluctuate fast among exposed and collapsed states in the K⁺ environment. In particular, the S1 sub-pocket tends to remain mid-range open in the presence of Na⁺ ions (Fig. S2(b), ESI[†]), which is likely favorable for catalytic substrate binding. Moreover, the entire protein prefers to retain a slightly smaller SASA in the presence of Na⁺ than K⁺ (Fig. S2(c), ESI[†]).

3.4 Ion-binding introduces population shifts in the conformational free energy surfaces of thrombin

Consistent with the clustering results, the conformational free energy surfaces of regulatory regions estimated from principal component analysis (PCA) (Fig. 6) show that the regulatory regions establish a narrower conformational space in the Na⁺ environment than the K⁺ one. Although the sampled conformational spaces of thrombin in Na⁺ and K⁺ simulations mainly overlap, there are ineligible areas that are barely accessible to both simulated environments. Again, this suggests Na⁺ and K⁺ have different global ion effects on the structural ensemble of the regulatory regions.

As a more sensitive analysis than the HDBSCAN clustering, the PCA highlights the dominant differences in the structural ensembles. Comparisons between Fig. 6a and b indicate that the binding of Na⁺ introduces a new free energy well (labeled as 2 on the figure). Upon the binding of Na⁺, the well 3 also shifts slightly to well 1. Similarly, the wells 6 to 9 are destabilized by the binding of K⁺. After K⁺-binding, conformations in well 5 become the most preferential and the size of the conformational space shrinks (Fig. 6c and d). As the new wells introduced by the binding of ions are mostly within the conformational space before ion-binding, the binding of a Na⁺/K⁺ ion merely perturbs conformational distributions instead of causing a significant induced-fit, such that the corresponding conformational well is beyond original conformational space. The visualization of the structures corresponding to different wells (Fig. 6e) reveals that primarily the γ loop exhibit the multiple poses in response to the types of ions and the status of ion binding.

Conformational free energy surfaces of the heavy atoms involved in the catalytic pocket, including the catalytic sub-pockets S1 to S6 (Table S1, ESI[†]) and catalytic triad (resid 79 135 241), were further constructed to examine possible subtle differences around the catalytic site under different solvent conditions. Similarly, the conformational free energy surface of the catalytic pocket is broader in the K⁺ environment than in the Na⁺ one. As seen in Fig. 7, the wells labeled by 4 and 6 are only observed in the simulations with K⁺ ions. The structural examination on the conformations in the wells 4 and 6 indicates that the 60s loop has a chance to flip up in the environment with potassium cations in the buffer, leaving the catalytic triad highly exposed (Fig. 7e).

Except for wells 4 and 6, the rest of the area in the conformational free energy surfaces basically overlaps regardless of global ion environments or local ion binding statuses. However, when a Na^+ or K^+ cation is bound, the population of conformations will redistribute toward multiple kinetically connected wells along the second principal component axis's direction (Fig. 7). Instead of the most populated well 2 for the Na^+ -unbound thrombin and well 5 for the K^+ -unbound thrombin, the well 1 around ($\text{PC1} = -0.5$, $\text{PC2} = -1$) and the well 3 around ($\text{PC1} = 0.5$, $\text{PC2} = -1.75$) become more populated after the binding of Na^+ and K^+ respectively. The structural examination on the structures projected into the wells 1, 2 and 3 illustrates that the principal component 2 mainly captures significant conformational variances in the S1 sub-pocket residues. When the cation is bound to the 220s loop, it is more likely to avoid a blockage over the catalytic triad by the side chains of S1 sub-pocket residues. Such blockage has been referred to as a stabilized self-inhibited conformation and seen in the crystal structure of the thrombin mutant D102N.⁸³

4 Discussion

Given the extensive sampling from 10 μs of simulations of the conformations of Na^+ -bound/unbound and K^+ -bound/unbound thrombin, the above results predict that the global ion environment and local ion binding conditions both have distinctive influences on thrombin's conformational ensembles of the surface functional loops. The regulatory regions including exosite I and II, 60s and γ loops on the surface of thrombin are more likely to remain in some long-lived conformational states when Na^+ is available in the environment. Comparisons among the visualization of several most populated clusters (Fig. S3, ESI[†]) particularly indicate that the highly dynamic γ loop can exhibit two similar long-lived poses in the Na^+ environment, while the γ loop only presents many fast transitive poses in the K^+ simulations. This also suggests thrombin may have different kinetics under different environments.

The conformational free energy surfaces estimated based on the principal component analysis (PCA) provides more sensitive measurements of the conformational ensembles of thrombin. Instead of the induced fit to the binding ion, our results demonstrate the ion-binding perturbs the conformational populations of thrombin's catalytic pocket. As shown in Fig. 7a and b, Na^+ -binding destabilizes well 2 and stabilizes the less populated wells 1 and 3; the binding of a K^+ cation destabilizes wells 5 and 6 and stabilizes wells 3 and 1. Due to the experimental observations that Na^+ ions significantly maximize thrombin's enzymatic activities rather than K^+ ions,¹³ the conformations within well 1 are likely to be more favorable for the catalytic functions of thrombin. Therefore, the reason why the monovalent Na^+ cation has more dramatic effects on thrombin's catalytic activities than the K^+ ion does may be that (1) the binding of Na^+ ion stabilizes the catalytically active conformations in well 1, while K^+ -binding also presents some stabilization of the catalytically less favorable forms in wells 3 and 4 and/or (2) the binding of a Na^+ ion is more selected over a K^+ one due to Na^+ 's relatively smaller ionic radius and the optimal shape of the 180s loops ensured by its length.⁸⁴

It should be noticed that the structural differences between wells 1, 2, and 3 are mainly the side chain orientations (Fig. 7e). Free energy barriers between the majority of wells are within or around the thermal fluctuation energy of kT (~ 0.6 kcal mol⁻¹ at room temperature). A small perturbation can easily cause a conformational change in the orientations of the side chains. Considering the crystal packing effects that can provide up to 6 kcal mol⁻¹ of free energy contribution to protein stability,⁸⁵ the hypothesis that the “fast form” of thrombin could be stabilized during crystallization²² is partially supported by our estimated free energy surfaces.

Moreover, after Na⁺/K⁺-binding, there is only a single deep free energy well corresponding to the backbone conformation of the catalytic pocket residues (Fig. S4, ESI[†]). These observations on the free energy surfaces demonstrate that, Na⁺/K⁺-binding modulates the side chain orientation of the catalytic pocket residues among multiple conformational states, while primarily a single backbone conformation is stabilized by the ion binding. As Na⁺-binding mainly narrows the conformational distribution of the backbones of the catalytic pocket residues, it explains why no significant backbone changes induced by Na⁺ ions has been seen in this region among solved crystal structures.

Through the simulations, we see that the surrounding cations – bound and unbound – have a complex regulatory effect on thrombin’s conformational ensembles. Instead of the catalytic triad, the surface loops of thrombin, such as the 60s, 70s, and the γ loops, present different mobilities and stabilities with respect to the unbound Na⁺ and K⁺ molecules. These regulatory regions appear structurally more rigid in the presence of Na⁺ ions, which may be important for thrombin’s catalytic functions. Such different global ion effects on the flexibility of thrombin might be associated with the different density distributions of unbound monovalent cations that were recently investigated by Kurisaki *et al.*⁸⁶ These authors also show the binding of Na⁺ is ineffective in forming the primary substrate pocket S1⁸⁷ and infer that, instead of the bound Na⁺ molecule, the unbound ones plays important roles in the allosteric activation of thrombin.

At first blush, our results seem to agree as the S1 pocket appears unaffected by the bound cations when visually comparing its ion⁺-bound and ion⁺-unbound ensembles’ solvent accessible surface areas (SASAs) (Fig. 5a). However, the statistical tests demonstrate both the bound and unbound cations result in significantly different SASA distributions of S1 at the 1% significance level (Tables S2 and S3, ESI[†]). This discrepancy is likely due to the difference in simulations timescales. Our 50-fold longer simulations appear to have uncovered a slightly (but statistically significant) different mean in SASA that is only apparent in microsecond-timescale simulations and not seen in shorter computational runs previously discussed in literature.⁸⁷ Moreover, we see that the bound and unbound cations can exhibit statistically (Tables S2 and S3, ESI[†]) and physically (Fig. 5d–g) significant effects on the solvent accessible surface areas in other sub-pockets of thrombin. Our SASA analysis indicates the necessity of a more comprehensive examination of the catalytic sub-pockets in order to understand Na⁺’s activation of thrombin.

As shown in Fig. 6 and 7, the locally bound Na⁺/K⁺ ion subtly shift the conformational distributions of thrombin’s catalytic pocket and regulatory regions. Multiple previously

existing conformational states become more thermodynamically accessible after the binding of a Na^+/K^+ ion. As a result, it should become easier for the thrombin to “find” a favorable conformation to recognize and interact with the substrate. Note that the main differences between the induced-fit and generalized allostery (also called conformational selection) – two major models for molecular recognitions – are whether the conformations selected by the ligand–substrate complex can naturally occur in the *apo*-form (*i.e.* ligand-unbound).^{88–91} In the induced-fit model, new conformational free energy wells are introduced upon ligand’s binding.^{89,91} In the generalized allostery model, the holo-form (*i.e.* ligand-bound) structures can be selected from the apo-form and the population of conformations redistributes after the binding of ligands, resulting in subtle shifts of wells in the free energy surfaces.^{29,88–91} Therefore, our simulations suggest generalized allostery is the mechanism of thrombin’s Na^+ -mediated functional switch between the “fast” and “slow” forms in catalytic events. As opposed to the two-state model¹⁸ that is based on the catalytic activities, Huntington’s (2012) multiple-state model²² is more supported by our explanation of generalized allostery in the structural ensemble viewpoint.

To clarify why no significantly different structures of thrombin in “fast” and “slow” forms have been solved yet, in this study, we have shown the unbound and bound cation’s impacts on thrombin’s mechanical and thermodynamic and revealed the generalized allostery of thrombin. Considering the correlation between the kinetic behavior and biological function, which has been seen in other systems,^{92,93} future work should focus on the how the kinetic properties are affected by the bound and unbound cations. For example, although we do not see the overall flexibilities of the catalytic pocket are significantly affected by Na^+/K^+ , our calculations on the temporal distributions of SASAs have suggested that catalytic pocket have different kinetic responses to Na^+/K^+ environments. The catalytic pocket doesn’t present any states with relatively long residence time in the absence of Na^+ (Fig. S2(c), $\text{ESI}^\dagger$). Particularly, the most exposed state of the catalytic pocket is transient in the K^+ environment (Fig. S2(c), $\text{ESI}^\dagger$). This result implies that the catalytic pocket should undergo a faster fluctuation in the absence of Na^+ . Along with the relative stable middle-range S1 substrate pocket (Fig. S2(a), $\text{ESI}^\dagger$), our simulations indicate that Na^+ preserves a “well-formed” shape of the catalytic pocket. This stabilized configuration is likely a favorable factor for catalytic efficiency and a critical kinetic signature of the “fast” form thrombin.

5 Conclusion

Through detailed comparisons of thrombin’s structural ensembles in the microsecond-scale all-atom MD simulations in sodium chloride and potassium chloride buffers, we reveal that thrombin is an ion-mediated enzyme in terms of its mechanical and thermodynamic properties. *Via* both the bound and unbound cations in the solvent, thrombin has relatively complex regulation of its distribution of conformational states. Both Na^+ - and K^+ -bindings narrow the conformational space of the regulatory regions and the catalytic pocket residues and cause population shifts toward multiple conformation states. Na^+ cations selectively stabilize the regulatory regions, particularly the highly dynamic γ loop. The binding of Na^+ also selectively stabilizes some catalytic pocket conformations, including the potentially catalytically favorable and unfavorable ones. This stabilization may cause the functional switch between the “slow” and “fast” forms of thrombin.

Generalized allostery is likely the mechanism of thrombin's delicate ion-mediation. Our finding provides mechanistic insights into how thrombin is activated by the Na⁺ ions, *via* stabilizing thrombin's regulatory surface loop and shifting the conformational distribution of the side chains around the catalytic pocket. Our understanding of generalized allostery on thrombin further suggests potential in the design and development of allosteric drugs that target thrombin.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Crystallography & Computational Biosciences Shared Resource services were supported by the Wake Forest Baptist Comprehensive Cancer Center's NCI Cancer Center Support Grant P30CA012197. Some computations were performed on the Wake Forest University DEAC Cluster, a centrally managed resource with support provided in part by the University. This work was partially supported by National Institute of General Medical Sciences grant T32-GM095440, supporting RLM. JX acknowledges a fellowship from the Center for Molecular Signaling at Wake Forest University. FRS also acknowledges a Reynolds Research leave from Wake Forest University.

References

1. Crawley JTB, Zanardelli S, Chion CKNK, Lane DA. *J Thromb Haemostasis*. 2007; 5:95–101. [PubMed: 17635715]
2. Di Cera E. *Mol Aspects Med*. 2008; 29:203–254. [PubMed: 18329094]
3. Kobrinisky, B., Karpatkin, S. *Thrombin*. Vol. ch. 9. Springer; New York: 2009. p. 161-172.
4. Cesarman-Maus G, Hajjar KA. *Br J Haematol*. 2005; 129:307–321. [PubMed: 15842654]
5. Davie E, Kulman J. *Semin Thromb Hemostasis*. 2006; 32:003–015.
6. Adams GN, Rosenfeldt L, Frederick M, Miller W, Waltz D, Kombrinck K, McElhinney KE, Flick MJ, Monia BP, Revenko AS, Palumbo JS. *Cancer Res*. 2015; 75:4235–4243. [PubMed: 26238780]
7. Radjabi AR, Sawada K, Jagadeeswaran S, Eichbichler A, Kenny HA, Montag A, Bruno K, Lengyel E. *J Biol Chem*. 2008; 283:2822–2834. [PubMed: 18048360]
8. Nierodzik ML, Karpatkin S. *Cancer Cell*. 2006; 10:355–362. [PubMed: 17097558]
9. Goldman JA, Eckerling B. *Am J Obstet Gynecol*. 1972; 112:1132–1133. [PubMed: 5017644]
10. Adrogué HJ, Madias NE. *N Engl J Med*. 2000; 342:1581–1589. [PubMed: 10824078]
11. Adrogué HJ, Madias NE. *N Engl J Med*. 2000; 342:1493–1499. [PubMed: 10816188]
12. Rubin D, Christian C. *Pediatric Emergency Care*. 2001; 17:313–314.
13. Orthner CL, Kosow DP. *Arch Biochem Biophys*. 1980; 202:63–75. [PubMed: 7396537]
14. De Cristofaro R, Picozzi M, Morosetti R, Landolfi R. *J Mol Biol*. 1996; 258:190–200. [PubMed: 8613987]
15. Dang QD, Guinto ER, Cera ED. *Nat Biotechnol*. 1997; 15:146–149. [PubMed: 9035139]
16. Bush LA, Nelson RW, Di Cera E. *J Biol Chem*. 2006; 281:7183–7188. [PubMed: 16428384]
17. Dang OD, Vindigni A, Di Cera E. *Proc Natl Acad Sci U S A*. 1995; 92:5977–5981. [PubMed: 7597064]
18. Wells CM, Di Cera E. *Biochemistry*. 1992; 31:11721–11730. [PubMed: 1445907]
19. Di Cera E, Guinto ER, Vindigni A, Dang QD, Ayala YM, Wuyi M, Tulinsky A. *J Biol Chem*. 1995; 270:22089–22092. [PubMed: 7673182]
20. Zhang E, Tulinsky A. *Biophys Chem*. 1997; 63:185–200. [PubMed: 9108691]
21. Pineda AO, Carrell CJ, Bush LA, Prasad S, Caccia S, Chen ZW, Mathews FS, Di Cera E. *J Biol Chem*. 2004; 279:31842–31853. [PubMed: 15152000]
22. Huntington JA. *Biochim Biophys Acta, Proteins Proteomics*. 2012; 1824:246–252.

23. Lechtenberg BC, Johnson DJD, Freund SMV, Huntington JA. *Proc Natl Acad Sci U S A*. 2010; 107:14087–14092. [PubMed: 20660315]
24. Lechtenberg BC, Freund SMV, Huntington JA. *Biol Chem*. 2012; 393:889–898. [PubMed: 22944689]
25. Dror RO, Dirks RM, Grossman J, Xu H, Shaw DE. *Annu Rev Biophys*. 2012; 41:429–452. [PubMed: 22577825]
26. Papaleo E. *Front Mol Biosci*. 2015; 2:1–6. [PubMed: 25988169]
27. Salsbury FR Jr. *Curr Opin Pharmacol*. 2010; 10:738–744. [PubMed: 20971684]
28. Melvin RL, Thompson WG, Godwin RC, Gmeiner WH, Salsbury FR. *Front Phys*. 2017; 5:10.
29. Xiao J, Salsbury FR. *J Biomol Struct Dyn*. 2016; 1102:1–16.
30. Vasilyeva A, Clodfelter JE, Rector B, Hollis T, Scarpinato KD, Salsbury FR. *DNA Repair*. 2009; 8:103–113. [PubMed: 18955167]
31. Vasilyeva A, Clodfelter JE, Gorczynski MJ, Gerardi AR, King SB, Salsbury F, Scarpinato KD. *J Nucleic Acids*. 2010:1–13.
32. Godwin, RC., Melvin, R., Salsbury, FR. *Methods in Pharmacology and Toxicology*. Zhang, W., editor. Springer; New York: 2015. p. 1-30.
33. Karplus M, McCammon JA. *Nat Struct Biol*. 2002; 9:646–652. [PubMed: 12198485]
34. Piana S, Klepeis JL, Shaw DE. *Curr Opin Struct Biol*. 2014; 24:98–105. [PubMed: 24463371]
35. Godwin RC, Melvin RL, Gmeiner WH, Salsbury FR. *Biochemistry*. 2017; 56:623–633. [PubMed: 28035815]
36. Melvin RL, Gmeiner WH, Salsbury FR. *J Phys Chem B*. 2016; 120:10269–10279. [PubMed: 27606431]
37. Monticelli, L., Tieleman, DP. *Biomolecular Simulations: Methods and Protocols*. Vol. 924. Humana Press; 2013. p. 197-213.
38. Lopes, PEM., Guvench, O., MacKerell, AD. *Molecular Modelling of Proteins*. Vol. ch 3. Springer; New York: 2015. p. 47-71.1215
39. Ayala YM, Vindigni A, Nayal M, Spolar RS, Record TM Jr, Di Cera E. *J Mol Biol*. 1995; 253:787–798. [PubMed: 7473752]
40. de Amorim HLN, Netz PA, Guimarães JA. *J Mol Model*. 2010; 16:725–735. [PubMed: 19816721]
41. Fuglestad B, Gasper PM, Tonelli M, McCammon JA, Markwick PR, Komives EA. *Biophys J*. 2012; 103:79–88. [PubMed: 22828334]
42. Russo Krauss I, Merlino A, Randazzo A, Novellino E, Mazzarella FSica L. *Nucleic Acids Res*. 2012; 40:8119–8128. [PubMed: 22669903]
43. Huntington JA. *Biol Chem*. 2008; 389:1025–1035. [PubMed: 18979627]
44. De Filippis V, De Dea E, Lucatello F, Frasson R. *Biochem J*. 2005; 390:485–492. [PubMed: 15971999]
45. Harvey MJ, De Fabritiis G. *J Chem Theory Comput*. 2009; 5:2371–2377. [PubMed: 26616618]
46. ali A, Blundell TL. *J Mol Biol*. 1993; 234:779–815. [PubMed: 8254673]
47. Humphrey W, Dalke A, Schulten K. *J Mol Graphics*. 1996; 14:33–38.
48. Awoonor-Williams E, Rowley CN. *J Chem Theory Comput*. 2016; 12:4662–4673. [PubMed: 27541839]
49. Anandakrishnan R, Aguilar B, Onufriev AV. *Nucleic Acids Res*. 2012; 40:W537–W541. [PubMed: 22570416]
50. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. *J Chem Theory Comput*. 2011; 7:525–537. [PubMed: 26596171]
51. Antosiewicz J, McCammon J, Gilson MK. *J Mol Biol*. 1994; 238:415–436. [PubMed: 8176733]
52. Bashford D, Gerwert K. *J Mol Biol*. 1992; 224:473–486. [PubMed: 1313886]
53. Lee MS, Salsbury FR, Brooks CL. *Proteins: Struct, Funct, Genet*. 2004; 56:738–752. [PubMed: 15281127]
54. Khandogin J, Brooks CL. *Biophys J*. 2005; 89:141–157. [PubMed: 15863480]
55. Kim MO, Nichols SE, Wang Y, McCammon JA. *J Comput-Aided Mol Des*. 2013; 27:235–246. [PubMed: 23579613]

56. Stanton CL, Houk KN. *J Chem Theory Comput.* 2008; 4:951–966. [PubMed: 26621236]
57. Vila JA, Arnautova YA, Vorobjev Y, Scheraga HA. *Proc Natl Acad Sci U S A.* 2011; 108:5602–5607. [PubMed: 21422292]
58. Hansen N, Van Gunsteren WF. *J Chem Theory Comput.* 2014; 10:2632–2647. [PubMed: 26586503]
59. Huang Z, Lin Z, Song C. *J Phys Chem A.* 2007; 111:4340–4352. [PubMed: 17474721]
60. Li S, Hong M. *J Am Chem Soc.* 2011; 133:1534–1544. [PubMed: 21207964]
61. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79:926–935.
62. Mackerell AD, Feig M, Brooks CL. *J Comput Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
63. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B.* 1998; 102:3586–3616. [PubMed: 24889800]
64. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. *J Chem Phys.* 1984; 81:3684–3690.
65. Lemons DS. *Am J Phys.* 1997; 65:1079.
66. Darden T, York D, Pedersen L. *J Chem Phys.* 1993; 98:10089–10092.
67. Feenstra KA, Hess B, Berendsen HJC. *J Comput Chem.* 1999; 20:786–798.
68. van Gunsteren W, Berendsen H. *Mol Phys.* 1977; 34:1311–1327.
69. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. *J Comput Chem.* 2011; 32:2319–2327. [PubMed: 21500218]
70. Gowers, RJ., Linke, M., Barnoud, J., Reddy, TJE., Melo, MN., Seyler, SL., Doma ski, J., Dotson, DL., Buchoux, S., Kenney, IM., Beckstein, O. *Proceedings of the 15th Python in Science Conference*; 2016; p. 98–105.
71. Liu P, Agrafiotis DK, Theobald DL. *J Comput Chem.* 2009; 31:1561–1563.
72. Theobald DL. *Acta Crystallogr, Sect A: Found Crystallogr.* 2005; 61:478–480.
73. Fuglestad B, Gasper PM, McCammon JA, Markwick PRL, Komives EA. *J Phys Chem B.* 2013; 117:12857–12863. [PubMed: 23621631]
74. Zhang Z, Ohl M, Diallo SO, Jalarvo NH, Hong K, Han Y, Smith GS, Do C. *Phys Rev Lett.* 2015; 115:1–6.
75. Campello, RJGB., Moulavi, D., Sander, J. *Advances in Knowledge Discovery and Data Mining.* 2013. p. 160–172.
76. Melvin RL, Godwin RC, Xiao J, Thompson WG, Berenhaut KS, Salsbury FR. *J Chem Theory Comput.* 2016; 12:6130–6146. [PubMed: 27802394]
77. Melvin RL, Salsbury FR. *Visual Statistics.* 2016; doi: 10.6084/m9.figshare.3398266.v1
78. Shrake A, Rupley J. *J Mol Biol.* 1973; 79:351–371. [PubMed: 4760134]
79. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. *Biophys J.* 2015; 109:1528–1532. [PubMed: 26488642]
80. Welch BL. *Biometrika.* 1947; 34:28–35. [PubMed: 20287819]
81. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH, Noé F. *J Chem Theory Comput.* 2015; 11:5525–5542. [PubMed: 26574340]
82. Melvin RL, Salsbury FR. *J Mol Graphics Modell.* 2016; 67:44–53.
83. Gandhi PS, Chen Z, Mathews FS, Di Cera E. *Proc Natl Acad Sci U S A.* 2008; 105:1832–1837. [PubMed: 18250335]
84. Prasad S, Wright KJ, Banerjee Roy D, Bush LA, Cantwell AM, Di Cera E. *Proc Natl Acad Sci U S A.* 2003; 100:13785–13790. [PubMed: 14612565]
85. Drenth J, Haas C. *J Cryst Growth.* 1992; 122:107–109.
86. Kurisaki I, Takayanagi M, Nagaoka M. *J Phys Chem B.* 2015; 119:3635–3642. [PubMed: 25654267]

87. Kurisaki I, Nagaoka M. *J Phys Chem B*. 2016; 120:11873–11879. [PubMed: 27781431]
88. Gunasekaran K, Ma B, Nussinov R. *Proteins: Struct, Funct, Genet*. 2004; 57:433–443. [PubMed: 15382234]
89. Motlagh HN, Wrabl JO, Li J, Hilser VJ. *Nature*. 2014; 508:331–339. [PubMed: 24740064]
90. Nussinov R, Tsai CJ. *Curr Opin Struct Biol*. 2015; 30:17–24. [PubMed: 25500675]
91. Csermely P, Palotai R, Nussinov R. *Trends Biochem Sci*. 2010; 35:539–546. [PubMed: 20541943]
92. Zhou HX, Wlodek ST, McCammon JA. *Proc Natl Acad Sci U S A*. 1998; 95:9280–9283. [PubMed: 9689071]
93. Zhang Z, Zheng Y, Petukh M, Pegg A, Ikeguchi Y, Alexov E. *PLoS Comput Biol*. 2013; 9:1–10.

Fig. 1. Thrombin–Na⁺ complex structure. The thrombin is shown in green. The 60s (in pink), 220s (in blue), and γ (in gray) loops and exosite I (in orange) and II (in yellow) are highlighted as the known regulatory regions. The side chains of the catalytic triad are displayed in red. The bound Na⁺ is shown as the yellow bead.

Fig. 2.
Probability and temporal distributions of minimum distance between 220s loop and ions.

Fig. 3.

Root-mean-square fluctuation of alpha carbons. In each ion-binding/unbinding status, regions with significant differences in C α atom RMSFs are primarily located in the surface loops and known regulatory sites.

Fig. 4.

Sorted clustering time series of thrombin's regulatory regions across the concatenated MD trajectories. The clustering time series were sorted according to the ion-binding statuses. Frames corresponding to K^+ -bound/unbound and Na^+ -bound/unbound were separated by red vertical lines. The makers with the largest cluster ID denote the noise, *i.e.* singleton conformational clusters.

Fig. 5. Distributions of solvent accessible surface area (SASA). Regarding conformations with a Na^+ -binding/unbinding and K^+ -binding/unbinding, the SASA distribution of the catalytic sub-pocket S1–6 residues are plotted respectively in panels (a–f). The SASA distribution of the catalytic triad and the SASA of the whole protein are also plotted in (g and h).

Fig. 6. Conformational free energy surfaces of the regulatory regions. Regarding conformations with a Na^+ -binding/unbinding and K^+ -binding/unbinding, four free energy surfaces are plotted respectively in panels (a–d) to compare the ion effects. Structural ensembles corresponding to the labeled wells on the free energy surfaces are visualized in panel (e). The representative structures of the whole protein are displaced *via* the NewCartoon representation in transparent green. The representative structure of the side chains in the regulatory regions are indicated by the Licorice representation in red. Gray shadows in NewCartoon representation display the variances in the regulatory regions. The side chains of the catalytic triad residues are shown in blue. Significant conformational differences in the regulatory regions are highlighted by the arrows.

Fig. 7.

Conformational free energy surfaces of the catalytic pocket. Regarding conformations with a Na^+ -binding/unbinding and K^+ -binding/unbinding, four free energy surfaces are plotted respectively in panels (a–d) to compare the ion effects. Structural ensembles corresponding to the labeled wells on the free energy surfaces are visualized in panel (e). The visualization strategy here is the same as in Fig. 6. As indicated by the arrows, PC1 captures much of the variance in 60s loop, and PC2 mainly captures different shapes in the S1 sub-pocket.

Table 1

Statistics of frames with ion-binding/unbinding

	Bound (frames/percentage)	Unbound (frames/percentage)
Na ⁺	345 917/69.18% (306 521/61.30%)	154 083/30.82% (193 479/38.70%)
K ⁺	213 812/42.76% (128 791/25.76%)	286 188/58.24% (371 209/74.24%)

Note: the numbers inside the parentheses represent the numbers or percentages of interior ion binding instead of surface binding (also called “exterior binding” in the main context), which are defined based on the mean distance between the sodium binding loop and its nearest Na⁺/K⁺ ion.