# Na+-binding modes involved in thrombin's allosteric response as revealed by molecular dynamics simulations, correlation networks and Markov modeling†

**Jiajie Xiao[a,b]**, **Freddie R. Salsbury Jr[a]**

[a]Department of Physics, Wake Forest University, Winston Salem, NC, USA.

[b]Department of Computer Science, Wake Forest University, Winston Salem, NC, USA

## Abstract

The monovalent sodium ion ($Na^+$) is a critical modulator of thrombin. However, the mechanism of thrombin's activation by $Na^+$ has been widely debated for more than twenty years. Details of the linkage between thrombin and $Na^+$ remain vague due to limited temporal and spatial resolution in experiments. In this work, we combine microsecond scale atomic-detailed molecular dynamics simulations with correlation network analyses and hidden Markov modeling to probe the detailed thermodynamic and kinetic picture of $Na^+$-binding events and their resulting allosteric responses in thrombin. We reveal that ASP189 and ALA190 comprise a stable $Na^+$-binding site (referred as "inner" $Na^+$-binding site) along with the previously known one (referred as "outer" $Na^+$-binding site). The corresponding newly identified $Na^+$-binding mode introduces significant allosteric responses in thrombin's regulatory regions by stabilizing selected torsion angles of residues responsive to $Na^+$-binding. Our Markov model indicates that the bound $Na^+$ prefers to transfer between the two $Na^+$-binding sites when an unbinding event takes place. These results suggest a testable hypothesis of a substrate-driven $Na^+$ migration ($G \sim 1.7$ kcal mol$^{-1}$) from the "inner" $Na^+$-binding site to the "outer" one during thrombin's catalytic activities. The binding of a $Na^+$ ion at the "inner" $Na^+$-binding site should be inferred as a prerequisite for thrombin's efficient recognition to the substrate, which opens a new angle for our understanding of $Na^+$-binding's allosteric activation on thrombin and sheds light on detailed processes in thrombin's activation.

## 1 Introduction

Thrombin is a serine protease that plays a critical role in cancer development[1–4] and blood coagulation.[5,6] Many experimental[7–11] and computational works[12–14] have revealed that the molecular properties and enzymatic activities of thrombin are mediated by monovalent sodium ($Na^+$) ions. In particular, experiments show that thrombin presents higher catalytic efficiency in coagulant reactions it is involved in when sodium ions are available in the

Conflicts of interest
There are no conflicts to declare.

aqueous solution.[7–9,11] However, it has been an over 20 year debate about the mechanism of thrombin's activation by $Na^+$ ions.[10,12,14–19]

X-ray crystallographic studies indicate that a sodium ion can be bound between thrombin's 220s and 180s loops and coordinated by 220s loop residues ARG221‡ and LYS224 and 180s loop residue TYR184A.[20,21] The 220s loop that contributes two binding residues is thereby often referred to as a "sodium binding loop" or "sodium loop" in literature (e.g. see Lechtenberg et al.[15] and Davie and Kulman[22]), although different articles have used different ranges of residues to denote this $Na^+$-binding loop (e.g. see Davie and Kulman,[22] Fuglestad et al.,[23] and Huntington[24]). Moreover, mutagenesis experiments also show that mutations of other residues, including ASP189, GLU217, ASP222, TYR225, TYR215, ASP221, GLY223 etc., result in a dramatic deficiency of $Na^+$-binding and/or perturbation of thrombin's function.[24–26] Because of these apparent associations between sodium-binding and sodium's regulations on thrombin's activities, the binding of a $Na^+$ ion to the sodium binding loop has been widely interpreted as a critical allosteric effector that facilitates the functional switch between the anticoagulant "slow" and coagulant "fast" forms of thrombin. [10,11,15–18] The activation reaction scheme of thrombin has been described as described as E* ↔ Na–E ↔ E-substrate, where E* and E denote the "slow" and "fast" thrombin respectively.[27]

As a result, in the literature, the $Na^+$-bound and $Na^+$-unbound thrombin were respectively referred to as the high activity "fast" and low activity "slow" forms. Based on measurements of $Na^+$-binding's affinity, Dang et al. estimated that thrombin respectively adopts a "fast" and "slow" form conformation for 60% and 40% of time in vivo in 1995.[18] In several later studies, the blockage of the substrate-binding pocket S1, S2, and/or the active site cleft were treated as the structural characters that distinguish the "slow" form against the "fast" one. [19,28,29] However, different solved structures of the putative "slow" thrombin don't reach an agreement on what conformational features of the "slow" thrombin should have.[16,24,28,29] In fact, while substrate-like binding ligands (such as PPACK inhibitor and its analogs) were commonly used in most crystallographic studies on thrombin, all solved crystal structures of wild-type thrombin share an overlapped backbone structure regardless of the presence of a bound $Na^+$ ion.[24] The structural differences seen in these crystal structures of the putative "slow" and "fast" forms are ascribed to small conformational changes in the inherently flexible regions,[16,24] suggesting that it is improper to interpret the mechanistic differences between the "fast" and "slow" thrombin merely based on their putative static crystal structures.

Nuclear magnetic resonance (NMR) spectroscopies reveal that thrombin's surface loops establish more dynamic motions in the absence of $Na^+$ ions and the binding of $Na^+$ can alter the conformational distribution of the active state.[16] These observations raise another explanation of thrombin's activation by $Na^+$. Huntington thereby prosed a dynamic model that thrombin should be considered as a "plastic" enzyme instead of the allosteric one in 2012.[16]

---

‡All residue numbers and insertion codes were based on the chymotrysin numbering scheme adopted in PDB 4DIH. Residue ARG221 is also sometimes denoted as ARG221A in literature, where A indicates the insertion code as being used in the PDB files.

Leveraging the development of hardware and software parallelizations, molecular dynamics (MD) simulations on thrombin have presented experimentally consistent but more detailed pictures of thrombin's conformational ensembles and dynamics.[12,30,31] Particularly, our previous simulations on thrombin illustrate a $Na^+$-mediated generalized allostery,[32] *i.e.* $Na^+$-binding subtlely perturbs the conformational distribution of the catalytic pocket instead of induced new conformations.[12] Our observations are not only consistent with Huntington's dynamic model[16] but also explain why all putative "fast" and "slow" thrombin should present similar structures.[12] Interestingly, through MD simulations, Kurisaki and Nagaoka found that the bound $Na^+$ ion can release from the sodium binding site and enter the S1 pocket.[19] Their analyses on $Na^+$-binding's impact on thrombin's S1 pocket suggest that $Na^+$-binding is irrelevant to the S1 pocket formation.[19] As an open S1 pocket was treated as a distinct structural feature in the active thrombin by some studies,[28,29] Kurisaki and Nagaoka pointed out a possibility that $Na^+$-binding is irrelevant to thrombin's functional switch.[19] Moreover, Kurisaki *et al.* further proposed that $Na^+$-binding is a negative effector in thrombin–substrate stereospecific complex formation based on their calculations on the interactions between substrate's primary binding residue (ArgP1) and thrombin's functional sites.[33]

Despite decades of efforts for understanding how thrombin is activated by $Na^+$, there appears to be still divergence in $Na^+$'s role in thrombin's functional activities. A detailed study on $Na^+$'s binding/unbinding processes and its influence on thrombin is necessary to clarify how sodium binding site and the thereby residues are linked to thrombin's functions. Here, we employ MD simulations to characterize detailed interactions – in both spatial and temporal space – between $Na^+$ ions and residues of apo-thrombin (*i.e.*, unliganded thrombin). Our present study not only bridges knowledge from previous structural and mutagenesis experiments but also reveals the existence of an new stable $Na^+$-binding site. Our correlation calculations between the torsion angles and $Na^+$ occupancy to each residue display the new $Na^+$-binding mode introduces dramatical allosteric responses in thrombin's regulatory regions. Finally, the hidden Markov modeling quantifies the $Na^+$-binding/unbinding processes and leads to a testable hypothesis on a substrate-driven migration of the bound $Na^+$ during thrombin's recognition to the substrate.

## 2 Materials and methods

### 2.1 Simulation configuration

Apo-thrombin in the 125 mM sodium chloride buffer was simulated in five runs of all-atom molecular dynamics simulations. All simulations were run using CHARMM22/CMAP force field[34,35] under the isothermal–isobaric *(NPT)* ensemble. In particular, simulations were run using Berendsen pressure control[36] with a 400 fs relaxation time to maintain 1 atm pressure (*P*). Langevin thermostat[37] with a damping coefficient of 0.1 was applied to set the temperature (*T*) as 300 K.

The initial coordinates of thrombin was determined based on the protein structure in the PDB 4DIH[38] and the missing residue fill-in protocol in the comparative modeling tool Modeller.[39] To balance the reliability and conformational sampling, and through visual assessments of possible hydrogen bonds on the X-ray structure, missing hydrogen atoms

were added to set all ioniz-able residues to their default protonation state at physiological pH. Particularly, all histidine residues were assigned delta protonation. More details about initial protein structure setup and corresponding validity discussion can be seen in the Materials and Method section in our previous work.[12]

Thrombin was solvated in an explicit TIP3P[40] water box with nm-padding in each of $x$, $y$, and $z$ direction. Leveraging the GPU parallelization enabled by ACEMD simulation package[41] and Metrocubo workstations produced by Acellera, the simulations we performed in this study can reach the microsecond time scale. Such long time scales ensure different Na$^+$-binding and unbinding events can naturally take place up to thousands of times in the MD simulations (Table S1, ESI†). Therefore, without loss of generality, all ions, including 22 Na$^+$ cations, were randomly placed in the 1 nm-padding (each of $x$, $y$, and $z$ direction) solvent box initially. As seen in our previous work,[12] the same set of simulations presented an experimentally consistent sodium occupancy in the thrombin's sodium cavity, suggesting the system has reached thermodynamic equilibrium with regard to Na$^+$-binding to thrombin.

### 2.2 Characterizations of Na$^+$-binding activities

#### 2.2.1 Na$^+$-occupancy and residence time.—To quantify how frequent and how long Na$^+$ ions interact with each thrombin residue, the occupancy and residence time of all possible residue-Na$^+$ contacts were computed. A contact $C_{i,j}^t$ between the $i$th residue and the $j$th Na$^+$ was defined if the $j$th Na$^+$ is within 0.38 nm away from the $i$th thrombin residue at time frame $t$. Such distance was used as previous work has shown that it is a proper threshold separating the unbound and bound Na$^+$ ion from both thermodynamic and kinetic aspects.[12] Therefore, the instantaneous number of bound Na$^+$ ions on the $i$th thrombin residue $n_i^t$ is

$$n_i^t = \sum_j C_{i,j}^t,$$

(1)

where $C_{i,j}^t$ is 1 if the minimum distance between heavy atoms in thrombin's $i$th residue and the $j$th Na$^+$ is within 0.38 nm at frame $t$ and 0 otherwise. The time-average Na$^+$-occupancy to the $i$th thrombin residue in MD trajectories with total frames of $T$ is

$$\langle n_i \rangle = \frac{1}{T} \sum_t n_i^t.$$

(2)

Meanwhile, given the binary time series of all residue-Na$^+$ contacts (i.e. a set of time series of $C_{i,j}^t$), we can estimate the residence time of any Na$^+$ ion to a specific residue $i$ based on the mean first passage time from the Na$^+$-on state (i.e. $C_{i,j}^t = 1$) to Na$^+$-off one (i.e. $C_{i,j}^t = 0$). In particular, the mean first passage time for a residue was directly calculated by counting how many frames the Na$^+$-on state remains on average over all Na$^+$ ions in all simulation

---

frames and it was assumed that any $Na^+$ ion would not be in contact with thrombin residues after a simulation (*i.e.* $C_{i,j}^{T+1} = 0$ for any $i$ and $j$, given $T$ is the length of a simulation). The error due to this mandatory assumption should be small given we have ample enough sampling in the microsecond scale simulations.

**2.2.2   Cooperativity in $Na^+$ binding.**—As multiple residues can accommodate the bound $Na^+$ and non-$Na^+$-binding site residues appeared also linked to the binding of $Na^+$, [20,21,25,26] it is important to quantitatively assess the binding cooperativity of $Na^+$ with respect to different residues. To differentiate positive and negative couplings, we calculated Pearson correlation coefficients between the numbers of bound $Na^+$ ions of pairs of thrombin residues using

$$\rho_{n_i, n_j} = \frac{1}{\sigma_{n_i} \sigma_{n_j}} \sum_t \left( n_i^t - \langle n_i^t \rangle \right) \left( n_j^t - \langle n_j^t \rangle \right),$$

(3)

where the $\sigma_{n_i} \, and \, \sigma_{n_j}$ are standard deviations of the time series of the numbers of bound $Na^+$ ions on the $i$th and $j$th residues of thrombin (*i.e.* $n_i^t \, and \, n_j^t$ from eqn (1)) respectively. Two residues will present high positive – up to 1 – correlation if they both interact with the bound $Na^+$. On the other hand, a negative correlation of −1 will be seen if the $Na^+$-binding events never take place simultaneously on these residues. In other words, when a $Na^+$-binding event occurs on one of these two residues, there will be a $Na^+$-unbinding event happening at the same time on the other residue. The transportation of a $Na^+$ between a pair of residues is a primary kinetic process that results in a negative correlation coefficient we see here. Random short-lived surface binding/unbinding activities among different residues should act as noise that can either lead to a near zero Pearson correlation coefficient for the linearly uncoupled residue pairs or reduce the magnitude of the Pearson correlation coefficient for the pairs responsible for coorperative binding.

Furthermore, we constructed a network graph to visualize the resultant correlation matrix whose elements were calculated according to eqn (3). In the network graph, the nodes represent thrombin's residues and the thicknesses of edges are proportional to the magnitude of the correlation coefficients between two residues whose corresponding nodes were connected. Such representation illustrates the detailed kinetic linkages among residues with respect to the $Na^+$-binding activities. To highlight the most significant linkages, 0.3 and −0.15 were used as the thresholds in the constructions of networks for positive and negative coorperative binding respectively. For the network representing positive coorperative binding, a pair of residues were connected by a line if $\rho_{n_i n_j} > 0.3$; for the network representing the negative coorperative binding, only pairs with $\rho_{n_i, n_j} < -0.15$ were connected as the negative couplings were all very weak. In general, positive couplings are stronger and more often than the negative ones since residues within the same secondary structure tend to present the similar dynamic behaviors.

### 2.2.3 Changes in backbone torsion angles upon Na+-binding.—To identify

what residues are responsive to $Na^+$-binding/unbinding, we measured the correlation (denoted as $r_{n_i, \alpha_j}$) between the number of bound $Na^+$ ions on the $i$th residue (denoted as $n_i$) and the $j$th residue's backbone torsion angles $(\alpha_j)$. Since the torsion angles – $\Phi$ and $\psi$ – are circular variables, which do not have meaningful arithmetic means, we quantified such correlation using linear-circular correlation coefficients $r_{n_i, \alpha_j}$ as follows,[42]

$$r_{n_i, \alpha_j} = \sqrt{\frac{\rho_{n_i, \cos(\alpha_j)}^2 + \rho_{n_i, \sin(\alpha_j)}^2 - 2 \cdot \rho_{n_i, \cos(\alpha_j)} \cdot \rho_{n_i, \sin(\alpha_j)} \cdot \rho_{\cos(\alpha_j), \sin(\alpha_j)}}{1 - \rho_{\cos(\alpha_j), \sin(\alpha_j)}^2}}, \quad (4)$$

where $\rho$ denotes the Pearson correlation function as already shown by eqn (3) except for changing related variables according to the corresponding subscripts of $\rho$. The value of $r_{n_i, \alpha_j}$ ranges from 0 to 1, where the value of 1 indicates linearly dependence and the value of 0 indicates the opposite.

Weighted directed graphs were also constructed to visualize the networks of the backbone torsion angles' responses upon changes in the number of bound $Na^+$ for each residue. Thrombin residues were denoted by the nodes. The thicknesses of edges are proportional to the corresponding correlation and a cutoff threshold of 0.3 was set here to only show couplings which have at least an intermediate strength.[43,44] As the correlation matrix here is asymmetric $\left(i.e.\ r_{n_i, \alpha_j} \neq r_{n_i, \alpha_i} \text{ if } i \neq j\right)$, arrows were added for the edges and the directions of arrows were from the residue being counted the number of bound $Na^+$ to the one being measured the backbone torsion angles. In other words, the direction of the edge, which was weighted by $r_{n_i, \alpha_j}$, is from residue $i$ to residue $j$. The backbone torsion angles were obtained *via* the coordinate featurizer module in Python package PyEmma.[45]

### 2.2.4 Hidden Markov modeling.—Markov modeling has proven useful for

understanding conformational states of biomolecules and their transitions.[46] Given the large-scale atomic molecular simulations, we seek a Markov model to obtain an insightful kinetic model that quantitatively describes the association and dissociation of $Na^+$ relative to the major $Na^+$ sodium binding residues.

As we are interested in the slow binding/unbinding processes rather than the fast thermal fluctuations, a time-lag independent component analysis. (Also called time-structure independent component analysis in literature. In the following context, we used TICA to stand for this decomposition method)[47,48] was performed before conducting clustering to obtain discretized trajectories. Choosing a time lag of 100 ps and using the distance matrix between the major $Na^+$ sodium binding residues (*i.e.*, ASP189, ALA190, ARG221 and LYS224) and their nearest $Na^+$ ion as the input feature, we obtained dominant TICA components that retain 95% kinetic variance.[49] Afterwards, the original feature space was

efficiently discretized into 100 microstates by running *k*-means clustering on the TICA projections along these dominant TICA components.

As demonstrated by other studies,[50–52] when being described by its cluster discretization of the state space, the system doesn't necessarily hold Markovian dynamics anymore. Therefore, we constructed a hidden Markov model (HMM) instead of a direct observed Markov model from clustering algorithms. The HMM is also a type of Markov model but its construction doesn't rely on a potentially improper assumption of Markovianity of the discrete dynamics.[50] Moreover, because of the Bayesian statistics employed during the hidden Markov modeling, the resulting HMMs are more robust even for a bad discretization when comparing to the direct observed Markov models.[49] By optimally grouping fast mixing microstates into metastable macrostates using the reversible Bayesian HMM estimator,[50] the HMMs provide coarse-grained kinetic models describing the complete $Na^+$ dissociation and association processes to thrombin. The number of macrostates was set to 3 based on the following three factors: (1) structural investigations of frames in the three major wells in the free energy surfaces constructed along TICA projections clearly demonstrate outer $Na^+$-binding, inner $Na^+$-binding, and $Na^+$-unbound modes (Fig. 4(a)); (2) there are three separable timescales of processes (Fig. S2, ESI†); and (3) trial and error using different values of macrostates suggests that setting 3 macrostates during Markov modeling results in the highest quality of Markov models (Fig. S3 and S4, ESI†). An HMM with a 1 ns Markov lag time was picked for further analyses as the relaxation timescales appeared independent of the lag time at this point (Fig. S2, ESI†) and this model successfully passed the Chapman–Kolmogorov test.[53] The generation and validation of the HMMs as well as the quantities representing the model (such as transition matrix and mean first passage time between states) were carried out and computed *via* the Python package PyEmma.[45]

## 3   Results

### 3.1   Multiple $Na^+$ binding sites exist

Although all simulations here were not started from a thrombin–$Na^+$ complex with a bound $Na^+$ at the known sodium binding site, the $Na^+$ cations can recognize the "sodium binding loop" and bind to it in the microsecond MD simulations. As seen in Fig. 1, we illustrated the density map of $Na^+$ ions. Consistent with several studies in literature,[12,14] here we also see that $Na^+$ ions establish both surface and interior bindings to the molecule of thrombin. The surface bindings of $Na^+$ appear associated with the side chains of residues in the 60s, 220s and $\gamma$ loops and chain termini, while the interior bindings are located between the 220s and 180s loops. Setting different number density values as the isovalue for the isosurface construction, the isosurfaces corresponding to the surface bindings largely vanish. These results indicate that the interior binding modes are highly populated and relatively stable, as our previous work showed that the system has reached thermodynamic equilibrium with respect to stable $Na^+$-binding to the sodium binding loop in the same set of simulations.[12]

Calculations of each residue's $Na^+$ ion occupancy reveal that ASP189, ALA190, GLU247, ARG17, ASP15, and GLY219 stand out as the top six residues with the most frequent contacts with $Na^+$ (Table S1, ESI†). While the highly flexible chain terminus residues GLU247, ARG17, and ASP15 are more likely to encounter the $Na^+$ ions in the solution by

chance rather than stably interact with them, only ASP189, ALA190, and GLY219 are in contact with a Na$^+$ ion for more than 20% of time (56.3%, 54.9%, and 26.3% respectively). ASP189, ALA190, and GLY219 are involved in thrombin's substrate pocket S1, although these three residues are also spatially adjacent (within ~0.4 nm) to the known Na$^+$ binding site (*i.e.,* ARG221, LYS224, and TYR184A). The ARG221 and LYS224 in the 220s loop are bound to a Na$^+$ ion for ~4% of time, while the nearby ASP221A has a higher Na$^+$ occupancy of ~9%. Moreover, the 180s loop residue TYR184A doesn't present frequent interactions with Na$^+$ ions (less than 1% of time). This result is consistent with the observation that TYR184A does not directly coordinate the bound Na$^+$ *via* its backbone oxygen atom like LYS224 and ARG221 do.[21,25]

Despite the fact that the previously known Na$^+$ binding site residues (*i.e.,* ARG221, LYS224, and TYR184A) are not very thermodynamically favorable for Na$^+$-binding comparing with several other residues, ARG221 and LYS224 stand out due to their kinetic aspect regarding Na$^+$-binding. These two residues establish the third and fourth longest Na$^+$-binding residence time among all thrombin residues. As seen in Table S1 (ESI†), ASP189, ALA190 have $31.2 \pm 0.3$ ns and $20.5 \pm 0.2$ ns Na$^+$ residence time respectively, which is about 1–3.7 fold longer than ARG221 and LYS224 do. Full statistics of Na$^+$-binding to other thrombin residues indicate ASP189, ALA190, ARG221 and LYS224 are most kinetically favorable residues for Na$^+$'s attachment.

Therefore, along with the known Na$^+$ binding site (ARG221 and LYS224), there should exist another one – composed of ASP189 and ALA190 – in the deeper interior region near thrombin's primary substrate binding pocket S1. This newly identified binding site is both thermodynamically and kinetically stable for the binding of a Na$^+$ ion. In the following and based on the relative positions of the residues ASP189, ALA190, ARG221 and LYS224, we refer to the deeper Na$^+$-binding site as "inner" Na$^+$-binding site and the known Na$^+$-binding site as "outer" Na$^+$-binding site. While it is well-known that the bound Na$^+$ ion at the "outer" Na$^+$-binding site interacts with four water molecules and the backbone oxygen atoms of ARG221 and LYS224,[21] the Na$^+$ at the "inner" Na$^+$-binding site appears to be coordinated to the side chain oxygen atoms of ASP189 and the backbone oxygen atom of ALA190 and to be hydrated with three water molecules (Fig. S1, ESI†). In addition, other nearby fluctuating residues such as GLY219 and water molecules can also occasionally provide energetic stabilization of the "inner"- Na$^+$-binding mode.

### 3.2 "Inner" and "outer" Na$^+$-binding modes are competitive

Given multiple binding modes of a Na$^+$ ion, the correlation network of each residue's Na$^+$-occupancy status sheds insights into linkages between each pair of thrombin residues with respective to Na$^+$-binding/unbinding. Our calculation reveals that the giant component – as shown in Fig. 2(a) – is composed of eight residues (~2.7% of total thrombin residues), suggesting the binding/unbinding events on most residues are transient and non-cooperative. Negative coupling with a magnitude greater than 0.15 only occurs in the kinetically slow binding/unbinding processes involved residues ASP189, ALA190, ARG221 and LYS224.

As seen in Fig. 2, the contacts of Na$^+$-LYS224 and Na$^+$-ARG221 are positively coupled with a high strength, indicating the bound Na$^+$ is coordinated by both LYS224 and ARG221 in

the "outer" Na$^+$-binding mode. Similarly, ASP189 and ALA190 coordinate a Na$^+$ in the "inner" Na$^+$-binding mode. As the Na$^+$ occupancy of GLY219 presents some positive correlation with ones of ASP189 and ALA190, residues GLY219 also directly contributes to the inner Na$^+$-binding. On the other hand, the Na$^+$ occupancies of ASP189 and ALA190 exhibit negative coupling with the one of LYS224 and ARG221, suggesting the attachment of a Na$^+$ ion at one of these interior Na$^+$-binding sites prevents the binding of Na$^+$ at the another one.

Furthermore, it is interesting to observe that the distant light chain residue SER14I is negatively coupled with ASP189 and ALA190 (Fig. 2). Meanwhile, SER14I is also positively coupled with the heavy chain residue GLN131 (Fig. 2). These phenomena suggest a linkage between thrombin's light and heavy chains, which can be seen given various aspects in the literature.[54–62]

### 3.3 Na$^+$ binding induces allosteric changes *via* stabilizing selected backbone torsion angles

To understand thrombin residues' allosteric linkages regarding the bound Na$^+$, we examined the coupling between the Na$^+$- bindings and residue torsion angles. For most residues, the numbers of locally bound Na$^+$ do not correlate with the fluctuations of other residues' backbone torsion angles (*i.e.* $\psi$ and $\Phi$). However, there are a group of residues presenting considerable coupling – correlation coefficient greater than 0.3 – between their $\psi$ and $\Phi$ angles and other residues' Na$^+$-occupancy. As listed in Fig. 3(a and b), when a Na$^+$ is bound to ASP189, ALA190, ARG221, LYS224, GLY219, TYR225 or ASP14 (indicated by red node labels in the network graphs and red beads in the structure plots except for GLY219 as it is also responsive to the Na$^+$-binding to other residues), specific backbone torsions of other residues (indicated by the blue node labels and blue beads in the structure plots) are affected. Consistent with residence time analysis, ASP189, ALA190, ARG221 and LYS224 keep distinguishing themselves as major Na$^+$-binding residues by presenting considerable correlations with other residues' $\psi$ and $\Phi$ angles. In particular, as indicated by the arrows from ASP189/ALA190 and ARG221/LYS224 in Fig. 3(a and b), the "inner" Na$^+$-binding mode induces significantly distinct allosteric responses in other residues' $\psi$ and $\Phi$ angles than the "outer" Na$^+$-binding mode does. The regions responsive to the Na$^+$-binding at the "inner" binding site covers the most known regulatory regions of thrombin, which includes the 180s, 220s and $\gamma$ loops and exosite I/II.

The most common effect introduced by Na$^+$-binding is to stabilize the torsion angle fluctuations of residues in thrombin's functional regions. For example, as suggested by the heat-map in Fig. 3(c), the binding of a Na$^+$ ion on residue ALA190 stabilizes a particular $\Phi$ angle ($\sim -\pi/2$) of residue GLY219. Such effect is known as the generalized allosteric effect[32,63–65] because the selected $\Phi$ angle can naturally occur before the Na$^+$-binding event. Another type of effect is "induced-fit"[63,65]-like. The bound Na$^+$ results in a unique new conformation such that another residue only presents that torsion angle after Na$^+$-binding. We could observe the "induced-fit"-like response just once in our analysis. As seen in Fig. 3(d), the conformations with a negative $\psi$ angle of residue ASN78 does not occur until a Na$^+$ is in contact with residue ASP189.

Along with the construction of the Markov model described in the Method section as well as the next subsection, clustering analysis on the residues that allosterically respond to $Na^+$-binding (see nodes in Fig. 3) further demonstrates that the inner $Na^+$-binding selectively stabilizes a conformational cluster of backbone torsion angles of these residues (Fig. S5, ESI†). The most obvious influence of $Na^+$-binding we saw here is the perturbation of the conformational distribution, although the conformational changes appear to be subtle. This observation is also consistent with our previous work,[12] which explains why it is too difficult to solve the distinct enough X-ray structures of the "slow" and "fast" thrombin and suggests the mechanism of this functional switch should be a $Na^+$-mediated generalized allostery instead of induced-fit.

### 3.4    "Inner" and "outer" Na⁺-binding modes are preferentially transferable into each other

As two distinct $Na^+$-binding modes in thrombin were identified, we proceeded to characterize the $Na^+$ association and dissociation processes *via* the construction of a three-state Markov model. Through the time-lag independent component analysis (TICA) using lag time of 100 ps, 95% kinetics of the nearest $Na^+$ to the major $Na^+$-binding residues ASP189, ALA190, ARG221, and LYS224 were captured by the three slowest TICA components. As shown in Fig. 4(a), the free energy surface estimated along the slowest two kinetic processes also illustrates three major wells, rationalizing the choice of three metastable states in the construction of a coarse-grained kinetic model.

Hidden Markov modeling has proved more tolerant of trajectory discretization errors than the directed observed Markov modeling.[50] After clustering the dominant TICA projections, using a Markov lag time of 1 ns (Fig. S2, ESI†), we optimally grouped fast mixing microstate into metastable macrostates (Fig. S3, ESI†). As a result, we built a hidden Markov model (HMM) that approximates the complete $Na^+$ dissociation and association processes to thrombin. Such coarse-grained model is statistically robust as it passed a Chapman–Kolmogorov test[53] in 95% confidence interval (Fig. S4, ESI†). Moreover, as illustrated in Fig. 4c, structurally, the metastable sets in the resulting HMM as well as different wells in the free energy surface were respectively identified as the $Na^+$-unbound, "inner" $Na^+$-bound, and "outer" $Na^+$-bound states, suggesting the rationality of the model physically.

Consistent with $Na^+$-occupancy results, the metastable state with a $Na^+$ ion near the "inner" binding site appears energetically more favorable ( $G \sim$ –1.7 kcal mol$^{-1}$) than the one with a $Na^+$ ion near the "outer" binding site. As illustrated by size of nodes in the transition network of the HMM (in Fig. 4(b)), the "outer" $Na^+$-bound state is the least populated one. Meanwhile, "inner" $Na^+$-bound, "outer" $Na^+$-bound and $Na^+$-unbound states are completely connected and transferable, indicating that there are two $Na^+$-binding/unbinding pathways for each $Na^+$-binding site.

The labels above the arrows in Fig. 4(b) demonstrate the mean first passage time from one metastable state to another. While it takes 26–57 ns for a $Na^+$ ion to bind to the "inner" $Na^+$-binding site, it takes 145–265 ns for that $Na^+$ ion to release to the "outer" $Na^+$-binding site or the external $Na^+$ reservoir. On the other hand, the "outer" $Na^+$-binding site presents the opposite kinetic behavior such that it is faster for a $Na^+$ to dissociate from the "outer" $Na^+$-

binding site than to attach to this site. These values demonstrate that, kinetically, it is easier for a Na$^+$ ion to bind to the "inner" Na$^+$-binding site than the "outer" one. Moreover, Fig. 4(b) also illustrates it takes shorter time for a bound Na$^+$ at either Na$^+$-binding sites to migrate to another one (26 ns and 145 ns for 0 $\leftrightarrow$ 2) than to unbind and diffuse back to the Na$^+$ reservoir (265 ns for 2 $\rightarrow$ 1 and 173 ns for 0 $\rightarrow$ 1). Therefore, it is kinetically more favorable for a bound Na$^+$ ion migrate between two Na$^+$-binding sites than to release to the Na$^+$ reservoir directly.

## 4 Discussion

Experimental techniques such as crystallography and NMR spectroscopy offer essential information to understand structures of biomolecules and corresponding mechanisms of their role in biological processes. At the same time, computational modeling has proved useful to shed insights into atomic-detailed dynamic perspectives of the system and fill the gaps between our understanding and observations. One example of this is when MD simulations have been employed to refine the metal ion binding site found by solution NMR experiments[66] as well as to reveal new ion binding sites in proteins in a *de novo* way.[67,68] Although the crystal structure of thrombin with a Na$^+$ at the "inner" Na$^+$-binding site has not been seen yet, several previous MD simulations using different MD force fields – including CHARMM,[34,35] Amber[69] as well as parameter adjustments[70,71] for corresponding water models[40,72] – have also shown the large density of a Na$^+$ in vicinity of ASP189[14] and suggested a second binding site in thrombin for Na$^+$.[62] Moreover, recent NMR experiments[61] have confirmed the allosteric pathways revealed by our previous MD studies. [12,31,62] Such an excellent agreement between experiments and computations indirectly validates the simulation protocol in the present study, given our previous work employed the same MD protocol as we did here.

Consistent with mutagenesis experiments,[24–26] our calculations also identified a list of residues presenting strong linkages with the binding of Na$^+$ (Fig. 2). The correlation network of residues' Na$^+$-occupancy helps understand these mutagenesis experiments by providing insights into the mechanism of the direct linkage among residues regarding two major Na$^+$-binding events – "outer" and "inner" Na$^+$-bindings. The sodium binding loop is a blurred name representing what region accommodates the bound Na$^+$ ion. More precisely, we observed that LYS224, ARG221, ASP189, ALA190, GLY219, CYC220, SER14I and GLN131 are directly linked regarding Na$^+$-binding. Particularly, as it has been revealed by mutagenesis experiments,[24–26] mutating residue ASP189 will extremely impair Na$^+$-binding. Such weakening Na$^+$-binding seems a direct consequence of disrupting the "inner" Na$^+$-binding site and the whole correlation network related to Na$^+$-binding. Meanwhile, there are other residues affecting Na$^+$-binding (see the review article by Huntington in 2008[24]) not showing in the giant component of the Na$^+$-occupancy correlation network. Such a result doesn't indicate these residues are irrelevant to Na$^+$-binding. Instead, our analysis clarifies that these residues should indirectly contribute to Na-binding. Most of these residues are spatially closed to the residues in the correlation networks shown in Fig. 2 and 3, implying that the mutations of these residues may affect Na$^+$-binding by perturbing the structural ensembles or cause hostile local environment changes for Na$^+$-binding kinetics.

While mutagenesis studies have already revealed the importance of ASP189 and other sodium binding loop residues in modulating $Na^+$-binding and thrombin's activation,[24–26] it was also suggested by Kurisaki *et al.* that $Na^+$-binding in the vicinity of residue ASP189 may present a negative effect on thrombin's recognition to substrate.[33] Note that ASP189 is involved in the primary substrate binding pocket S1 and it is responsible for the specific interactions with the positively charged P1 residue (usually ARG/LYS) in the substrate. We agree that the stable binding of $Na^+$ on ASP189 we demonstrate here may potentially impede substrate's recognition of the S1 subpocket due to the repulsive force between the bound $Na^+$ and the positively charged P1 residue. On the other hand, the repulsive force between the substrate P1 residue and the bound $Na^+$ may lead to a ion transition like the ion knock-on process in $K^+$ channel examined by Sumikama and Oiki in 2016.[73] As indicated by our HMM (Fig. 4(b)), a bound $Na^+$ prefers transferring between the "inner" and "outer" $Na^+$-binding site rather than going back to the external $Na^+$ reservoir directly. As a result, during the binding of thrombin's substrate, the substrate's P1 residue may induce a migration of the bound $Na^+$ from the "inner" $Na^+$-binding site to the "outer" one. The free energy difference between the these two $Na^+$-binding states (~1.7 kcal mol$^{-1}$) as well as the free energy barrier (~2.8 kcal mol$^{-1}$) suggests the ion migration from the "inner" $Na^+$-binding site to the "outer" one is not likely to occur under thermal fluctuations (~0.6 kcal mol$^{-1}$). This substrate or ligand induced ion migration is likely one of the reasons crystal structures of thrombin display the "outer" $Na^+$-binding mode, especially considering there is some substrate-like binding ligands (such as PPACK inhibitor and its analogs) at the active site in most crystal structures of thrombin. One can commonly observe these ligands interacting with ASP189 through the polar interaction, which likely have forced the $Na^+$ ion away from the "inner" $Na^+$-binding site. For the small selection of the thrombin structures, which are available without a ligand interacting (see PDB 2UUF[74] and 3U69[75]), they may not be germane due to other ligands bound at either the allosteric regulation site (for PDB 2UUF) or the S3 and S4 subpockets adjacent to S1 (for PDB 3U69). Unlike the conformations in "inner" $Na^+$-binding state, the subtle rearrangement of the side chains of residues around ASP189 and ALA190 partially occludes the "inner" $Na^+$-binding site. Instead, the water-binding in the S1 subpocket, as seen in these crystal structures, suggests the possibility that $Na^+$-binding at the "inner" binding site binding site competes with the water-binding. Detwetting of the same region has been previously discussed by Kurisaki *et al.* as a rate-limiting process in the interaction of the S1 pocket.[76] In summary, based on the evidence from our calculation and experimental works, it is highly likely there is a substrate-driven ion migration during thrombin–substrate interaction. High-resolution crystallographic studies on inhibitor- and ligand-free thrombin are needed to obtain more samples of the thrombin–Na complex, which may provide more direct evidence of the predicted inner-$Na^+$-binding site prior to thrombin's recognition to the substrate. On the other hand, as experimental approaches may be limited to the unphysiologically relevant conditions during the structural determination, computational methods such as MD simulations may provide further insights. Long timescale (hundred of microseconds to milliseconds) MD simulations that simulate the thrombin–substrate recognition process in the presence of $Na^+$ ions are needed to test if the inner $Na^+$-binding is indeed a prerequisite for thrombin's recognition to the substrate. If the proposed hypothesis is true, in the simulations, the substrate should appear to recognize the active site of thrombin more efficiently *(i.e. $k_{on}$ is larger)* after a Na

ion binds to the "inner" Na$^+$-binding site and allosterically stabilizes the enzyme. It is also expected to see that, during the association process of the substrate, the bound Na$^+$ ion should release from the "inner" Na$^+$-binding site and move to the "outer" Na$^+$-binding site.

Na$^+$-binding has been revealed as an allosteric effector of thrombin by many experimental and computational works.[7–9,11–14] In particular, our previous study indicates the binding of Na$^+$ doesn't perturb the backbones of the catalytic pocket residues but their side chains' orientation.[12] Here the side chains' allosteric responses upon Na$^+$-binding to each residue are quantified by the correlation coefficients between $\psi/\Phi$ angles and Na$^+$'s occupancies. Fig. 3 displays the Na$^+$-bindings at the "inner" and "outer" Na$^+$-binding sites introduce extremely different allosteric responses. The "outer" Na$^+$-binding mode only appears to stabilize the $\Psi$ angle of the nearby residue CYS220. In contrast, the "inner" Na$^+$-binding mode results in a much more significant influence on adjacent and distant residues. These affected residues are located in the 180s, 220s and $\gamma$ loop, exosite I/II, catalytic pocket and light chain, which are all functionally related to Na$^+$-binding and catalytic efficiency.[22,77,78] Except for residue ASN78, whose $\Psi$ angle can adopt a completely new value after a Na$^+$ binds to residue ASP189 (Fig. 3), the effect induced by the Na$^+$-binding at the "inner" Na$^+$-binding site is to select and stabilize a priorly occurring $\Psi/\Phi$ torsion angle of the other responsive residues. In other words, the "inner" Na$^+$-binding mode stabilizes a selected conformation (Fig. S5, ESI†). Therefore, these findings demonstrate the "inner" Na$^+$-binding induced generalized allostery[32,63–65] is the main mechanism of Na$^+$'s mediation on thrombin, which also agrees with previous discussions on thrombin's conformational ensembles.[12,79]

With respect to the debate about whether Na$^+$-binding is critical for thrombin's activation, while our previous work also demonstrated the importance of the unbound Na$^+$ ions,[12] this work emphasizes that Na$^+$-binding is a necessary biological event for thrombin's activation. Our quantitative and detailed results appear to align with various experimental studies.[7–11,25,26,61] On the other hand, Kurisaki *et al.'s* delicate calculations also pointed out that the existence of a bound Na$^+$ ion at the "outer" Na$^+$-binding site should result in about 4 kcal mol$^{-1}$ energy disadvantage in the interaction between the substrate residue ArgP1 and thrombin's S1, 220s loop residues, implying the bound Na$^+$ may play a negative role for thrombin–substrate stereospecific complex formation.[33] However, these calculations were based on MD simulations starting from the thrombin-substrate complex in the presence or absence of the bound Na$^+$ ion. Such an initial setup was only able to capture the late formation stage of the thrombin-substrate complex, particularly given that the simulation length of 500 ns may not be long enough to well study the complete association/dissociation process of a substrate to thrombin. Along with our previous work,[12] our simulations illustrate that the binding of a sodium ion perturbs thrombin's conformational distribution. As the equilibrium constant for conformational transition can significantly affect molecular recognition,[80] the Na$^+$-binding-induced conformational distribution shift should be profound for thrombin–substrate interactions. The Na$^+$ binding event at the "inner" sodium binding site stabilizes specific residues' torsion angles (Fig. S5, ESI†). The resultant conformation is likely favorable for interact with the substrate, given the apparent experimentally demonstrated association between Na$^+$-binding and thrombin's functional activities. The relatively small free energy barriers among each Na$^+$ binding/unbinding states may actually

reflect how Na-binding subtly regulates the thrombin's activation. If these free energy differences are too large, the bound Na ion may provide too large resistance for positively charged P1 residue of the substrate to recognize the active pocket, which is against the experimental observations that Na ions are favorable for efficient interactions between thrombin and its substrate. Therefore, despite the repulsive force between the bound Na$^+$ and substrate P1 residues, Na$^+$ may not necessarily be a negative effector for the entire thrombin–substrate formation process. The bound Na$^+$ ion seems to play a positive role in the early stage of thrombin–substrate formation in terms of stabilizing the specific conformation of thrombin. In this sense, the inner Na$^+$-binding may be more like a promoter, which promotes thrombin's catalytic activities by shifting thrombin's conformational ensembles from "slow" form (denoted as E*) the towards the "fast" form (denoted as E) in the traditional activation reaction scheme of enzyme.[27] To further validate this hypothesis, Hundreds of microsecond to millisecond scale MD simulations are be needed in future work to examine the role of Na$^+$ ions in the entire thrombin-substrate recognition processes.

## 5  Conclusions

Combining extensive all-atom MD simulations and various analyses, we have characterized all possible residue-Na$^+$ contacts and demonstrated the existence of a new Na$^+$-binding site in apo-thrombin. This Na$^+$-binding site is located beside the previously known Na$^+$-binding site (referred to as "outer" Na$^+$-binding site in the context) and largely accommodated by ASP189 and ALA190 of the S1 subpocket. Compared with the "outer" Na$^+$-binding site, this "inner" Na$^+$-binding site is thermodynamically and kinetically more favorable for the binding of a Na$^+$ ion.

Instead of the binding activity at the "couter" Na$^+$-binding site, it is actually the "inner" binding mode that induces significant allosteric responses in thrombin's 180s, 220s and γ loop, exosite I/II, catalytic pocket and light chain. Such a Na$^+$ binding stabilizes selected torsion angles of the residues in these regulatory regions. Our findings do not only provide another evidence of the generalized allostery as the mechanism of Na$^+$'s regulation on thrombin, but also depict a more detailed picture of the allosteric pathways. The allosterically linked residues' dihedral angles are stabilized by the Na$^+$-binding at the "inner" Na$^+$-binding site. The stabilized conformation with selected dihedral angles is likely favorable for the "fast" thrombin.

Hidden Markov modeling offers a coarse-grained kinetic model that quantifies the complete association and dissociation processes of Na$^+$ relative to the "outer" and "inner" Na$^+$-binding sites. We observed that a bound Na$^+$ at either of these two sites prefers transferring to another Na$^+$-binding site than going to the solution directly. This property explains why we saw a competitive binding between the two Na$^+$-binding modes in our correlation analysis. Moreover, our kinetic model proposes a testable hypothesis on the underlying process of thrombin–substrate recognition. The bound Na$^+$ at the "inner" Na$^+$-binding site first stabilizes a particular conformation that is favorable for thrombin's catalytic efficiency. Secondly, the substrate recognizes the well-formed stable catalytic pocket of thrombin and starts binding. During this substrate-binding process, the bound Na$^+$ ion at the "inner" Na$^+$-binding site is being repelled by the positively charged P1 residue of the substrate and then

transfers to the "outer" Na$^+$-binding site ($G \sim 1.7$ kcal mol$^{-1}$). This hypothesis provides a likely explanation why crystal structures of thrombin do not display the "inner" Na$^+$-binding mode yet, given the fact that the ASP189 is mostly occupied by a bound chemical compound *(e.g.* PPACK inhibitor) in these crystal structures. Further studies, such as inhibitor-free structural determination or large scaled inhibitor-bound molecular dynamics of thrombin–substrate complex in NaCl buffer, could test this hypothesis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

## References

1. Adams GN, Rosenfeldt L, Frederick M, Miller W, Waltz D, Kombrinck K, McElhinney KE, Flick MJ, Monia P, Revenko AS and Palumbo JS, Cancer Res, 2015, 75, 4235–4243. [PubMed: 26238780]

2. Radjabi AR, Sawada K, Jagadeeswaran S, Eichbichler A, Kenny HA, Montag A, Bruno K and Lengyel E, J. Biol. Chem, 2008, 283, 2822–2834. [PubMed: 18048360]

3. Nierodzik ML and Karpatkin S, Cancer Cell, 2006, 10, 355–362. [PubMed: 17097558]

4. Kobrinsky B and Karpatkin S, in Thrombin, ed. Maragoudakis ME and Tsopanoglou NE, Springer, New York, NY, 2009, vol. 19, ch. 9, pp. 161–172.

5. Crawley JTB, Zanardelli S, Chion CKNK and Lane DA, J. Thromb. Haemostasis, 2007, 5, 95–101. [PubMed: 17635715]

6. Di Cera E, Mol. Aspects Med, 2008, 29, 203–254. [PubMed: 18329094]

7. Orthner CL and Kosow DP, Arch. Biochem. Biophys, 1980, 202, 63–75. [PubMed: 7396537]

8. De Cristofaro R, Picozzi M, Morosetti R and Landolfi R, J. Mol. Biol, 1996, 258, 190–200. [PubMed: 8613987]

9. Dang QD, Guinto ER and Cera ED, Nat. Biotechnol, 1997, 15, 146–149. [PubMed: 9035139]

10. De Filippis V, De Dea E, Lucatello F and Frasson R, Biochem. J, 2005, 390, 485–492. [PubMed: 15971999]

11. Bush LA, Nelson RW and Di Cera E, J. Biol. Chem, 2006, 281, 7183–7188. [PubMed: 16428384]

12. Xiao J, Melvin RL and Salsbury FR, Phys. Chem. Chem. Phys, 2017, 19, 24522–24533. [PubMed: 28849814]

13. de Amorim HLN, Netz PA and Guimarães JA, J. Mol. Model, 2010, 16, 725–735. [PubMed: 19816721]

14. Kurisaki I, Takayanagi M and Nagaoka M, J. Phys. Chem. B, 119, 3635–3642. [PubMed: 25654267]

15. Lechtenberg BC, Freund SMV and Huntington JA, Biol. Chem, 2012, 393, 889–898. [PubMed: 22944689]

16. Huntington JA, Biochim. Biophys. Acta, Proteins Proteomics, 2012, 1824, 246–252.

17. Wells CM and Di Cera E, Biochemistry, 1992, 31, 11721–11730. [PubMed: 1445907]

18. Dang OD, Vindigni A and Di Cera E, Proc. Natl. Acad. Sci. U. S. A, 1995, 92, 5977–5981. [PubMed: 7597064]

19. Kurisaki I and Nagaoka M, J. Phys. Chem. B, 2016, 120, 11873–11879. [PubMed: 27781431]

20. Di Cera E, Guinto ER, Vindigni A, Dang QD, Ayala YM, Wuyi M and Tulinsky A, J. Biol. Chem, 1995, 270, 22089–22092. [PubMed: 7673182]

21. Zhang E and Tulinsky A, Biophys. Chem, 1997, 63,185–200. [PubMed: 9108691]

22. Davie E and Kulman J, Semin. Thromb. Hemostasis, 2006, 32, 003–015.

23. Fuglestad B, Gasper PM, McCammon JA, Markwick PRL and Komives EA, J. Phys. Chem. B, 2013, 117, 12857–12863. [PubMed: 23621631]

24. Huntington JA, Biol. Chem, 2008, 389, 1025–1035. [PubMed: 18979627]

25. Pineda AO, Carrell CJ, Bush LA, Prasad S, Caccia S, Chen ZW, Mathews FS and Di Cera E, J. Biol. Chem, 2004, 279, 31842–31853. [PubMed: 15152000]

26. Page MJ and Di Cera E, Physiol. Rev, 2006, 86, 1049–1092. [PubMed: 17015484]

27. Bah A, Garvey LC, Ge J and Di Cera E, J. Biol. Chem, 2006, 281, 40049–40056. [PubMed: 17074754]

28. Johnson DJ, Adams TE, Li W and Huntington JA, Biochem. J, 2005, 392, 21–28. [PubMed: 16201969]

29. Pineda AO, Savvides SN, Waksman G and Cera ED, J. Biol. Chem, 2002, 277, 40177–40180. [PubMed: 12205081]

30. Fuglestad B, Gasper PM, Tonelli M, McCammon JA, Markwick PR and Komives EA, Biophys. J, 2012, 103, 79–88. [PubMed: 22828334]

31. Xiao J and Salsbury FR, J. Biomol. Struct. Dyn, 2016,1102, 1–16.

32. Gunasekaran K, Ma B and Nussinov R, Proteins: Struct., Funct., Genet, 2004, 57, 433–443. [PubMed: 15382234]

33. Kurisaki I, Takayanagi M and Nagaoka M, J. Phys. Chem. B, 2016, 120, 4540–4547. [PubMed: 27164318]

34. Mackerell AD, Feig M and Brooks CL, J. Comput. Chem, 2004, 25, 1400–1415. [PubMed: 15185334]

35. 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 and Karplus M, J. Phys. Chem. B, 1998, 102, 3586–3616. [PubMed: 24889800]

36. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A and Haak JR, J. Chem. Phys, 1984, 81, 3684–3690.

37. Lemons DS, Am. J. Phys, 1997, 65, 1079.

38. Russo Krauss I, Merlino A, Randazzo A, Novellino E, Mazzarella L and Sica F, Nucleic Acids Res, 2012, 40, 8119–8128. [PubMed: 22669903]

39. Šali A and Blundell TL, J. Mol. Biol, 1993, 234, 779–815. [PubMed: 8254673]

40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW and Klein ML, J. Chem. Phys, 1983, 79, 926–935.

41. Harvey MJ and De Fabritiis G, J. Chem. Theory Comput, 2009, 5, 2371–2377. [PubMed: 26616618]

42. Mardia KV, Biometrika, 1976, 63, 403.

43. Fenwick RB, Orellana L, Esteban-Martín S, Orozco M and Salvatella X, Nat. Commun, 2014, 5, 4070. [PubMed: 24915882]

44. Mukaka MM, Malawi Med. J, 2012, 24, 69–71. [PubMed: 23638278]

45. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H and Noé F, J. Chem. Theory Comput, 2015, 11, 5525–5542. [PubMed: 26574340]

46. Husic BE and Pande VS, J. Am. Chem. Soc, 2018, 140, 2386–2396. [PubMed: 29323881]

47. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G and Noé F, J. Chem. Phys, 2013, 139, 015102.

48. Schwantes CR and Pande VS, J. Chem. Theory Comput, 2013, 9, 2000–2009. [PubMed: 23750122]

49. Noé F and Clementi C, J. Chem. Theory Comput, 2015, 11, 5002–5011. [PubMed: 26574285]

50. Noé F, Wu H, Prinz JH and Plattner N, J. Chem. Phys, 2013, 139, 184114.

51. Suárez E, Pratt AJ, Chong LT and Zuckerman DM, Protein Sci, 2016, 25, 67–78. [PubMed: 26131764]

52. Suárez E, Adelman JL and Zuckerman DM, J. Chem. Theory Comput, 2016, 12, 3473–3481. [PubMed: 27340835]

53. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schtte C and Noé F, J. Chem. Phys, 2011, 134, 1–23.

54. Akhavan S, Rocha E, Zeinali S and Mannucci PM, Br. J. Haematol, 1999, 105, 667–669. [PubMed: 10354128]

55. Akhavan S, Mannucci PM, Lak M, Mancuso G, Mazzucconi MG, Rocino A, Jenkins PV and Perkins SJ, Thromb. Haemostasis, 2000, 84, 989–997. [PubMed: 11154146]

56. Lefkowitz JB, Haver T, Clarke S, Jacobson L, Weller A, Nuss R, Manco-Johnson M and Hathaway WE, Br. J. Haematol, 2000, 108, 182–187. [PubMed: 10651742]

57. Sun WY, Burkart MC, Holahan JR and Degen SJ, Blood, 2000, 95, 711–714. [PubMed: 10627484]

58. Papaconstantinou ME, Bah A and Di Cera E, Cell. Mol. Life Sci, 2008, 65, 1943–1947. [PubMed: 18470478]

59. De Cristofaro R, Carotti A, Akhavan S, Palla R, Peyvandi F, Altomare C and Mannucci PM, FEBS J, 2006, 273, 159–169. [PubMed: 16367756]

60. De Cristofaro R, Akhavan S, Altomare C, Carotti A, Peyvandi F and Mannucci PM, J. Biol. Chem, 2004, 279, 13035–13043. [PubMed: 14722067]

61. Handley LD, Fuglestad B, Stearns K, Tonelli M, Fenwick RB, Markwick PRL and Komives EA, Sci. Rep, 2017, 7, 39575. [PubMed: 28059082]

62. Xiao J, Melvin RL and Salsbury FR, J. Biomol. Struct. Dyn, 2018, 1102, 1–55.

63. Motlagh HN, Wrabl JO, Li J and Hilser VJ, Nature, 2014, 508, 331–339. [PubMed: 24740064]

64. Nussinov R and Tsai CJ, Curr. Opin. Struct. Biol, 2015, 30, 17–24. [PubMed: 25500675]

65. Csermely P, Palotai R and Nussinov R, Trends Biochem. Sci, 2010, 35, 539–546. [PubMed: 20541943]

66. Chakravorty DK, Wang B, Lee CW, Guerra AJ, Giedroc DP and Merz KM, J. Biomol. NMR, 2013, 56, 125–137. [PubMed: 23609042]

67. Zomot E, Gur M and Bahar I, J. Biol. Chem, 2015, 290, 544–555. [PubMed: 25381247]

68. Carnevale V, Treptow W and Klein ML, J. Phys. Lett, 2011, 2, 2504–2508.

69. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C, Proteins: Struct., Funct., Bioinf., 2006, 65, 712–725.

70. Joung IS and Cheatham TE, J. Phys. Chem. B, 2008, 112, 9020–9041. [PubMed: 18593145]

71. Joung IS and Cheatham TE, J. Phys. Chem. B, 2009, 113, 13279–13290. [PubMed: 19757835]

72. Kusalik PG and Svishchev IM, Science, 1994, 265, 1219–1221. [PubMed: 17787590]

73. Sumikama T and Oiki S, J. Am. Chem. Soc, 2016, 138, 10284–10292. [PubMed: 27454924]

74. Ahmed HU, Blakeley MP, Cianci M, Cruickshank DWJ, Hubbard JA and Helliwell JR, Acta Crystallogr., Sect. D: Biol. Crystallogr, 2007, 63, 906–922. [PubMed: 17642517]

75. Figueiredo AC, Clement CC, Zakia S, Gingold J, Philipp M and Pereira PJB, PLoS One, 2012, 7, 1–11.

76. Kurisaki I, Barberot C, Takayanagi M and Nagaoka M, J. Phys. Chem. B, 2015, 119, 15807–15812. [PubMed: 26634958]

77. Di Cera E, Page MJ, Bah A, Bush-Pelc LA and Garvey LC, Phys. Chem. Chem. Phys, 2007, 9, 1291. [PubMed: 17347701]

78. Di Cera E, Mol. Aspects Med, 2008, 29, 203–254. [PubMed: 18329094]
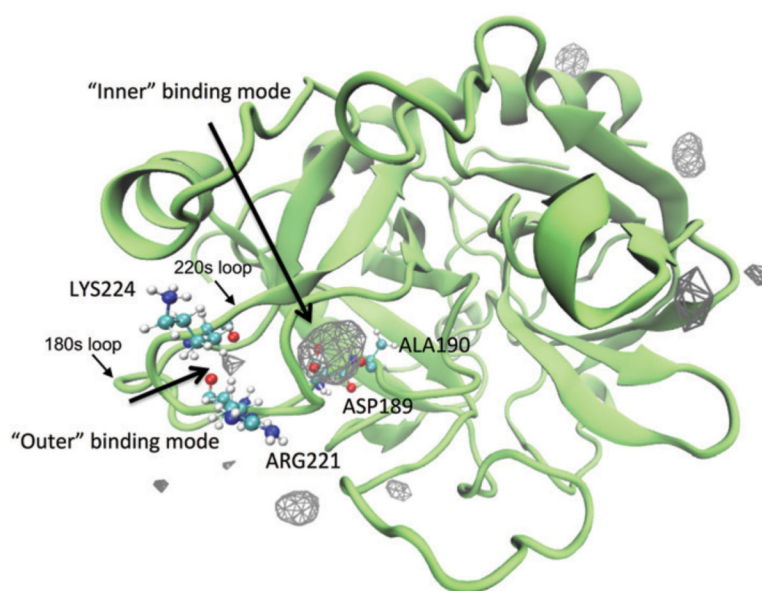
79. Lechtenberg BC, Freund SMV and Huntington JA, Biol. Chem, 2012, 393, 889–898. [PubMed: 22944689]

80. Xu H, Schmidt AG, O'Donnell T, Therkelsen MD, Kepler TB, Moody MA, Haynes BF, Liao HX, Harrison SC and Shaw DE, Proteins: Struct., Funct., Bioinf., 2015, 83, 771–780.

**Fig. 1.**
A density map of Na$^+$ around thrombin. To show the density map as isosurfaces, a fractional occupancy of 0.05 was used as the isovalue. The protein is shown as the structure closest to the average structure in the simulations. Two different interior binding modes are identified between 220s and 180s loops. They are referred to as "outer" and "inner" binding modes in the following context. The residues accommodate these two binding modes are indicated by the labels beside their CPK representations.
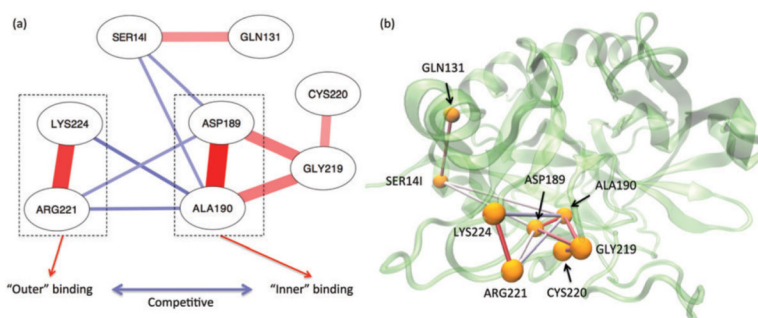
**Fig. 2.**
Cooperativity in Na$^+$ bindings. Panel (a) illustrates the giant component of the thrombin residues' coupling network with respect to number of bound Na$^+$. Panel (b) shows these coupling on the structure plot. The red and blue lines on both panels represent the positive and negative correlations respectively. A darker and thicker line indicates a stronger coupling.
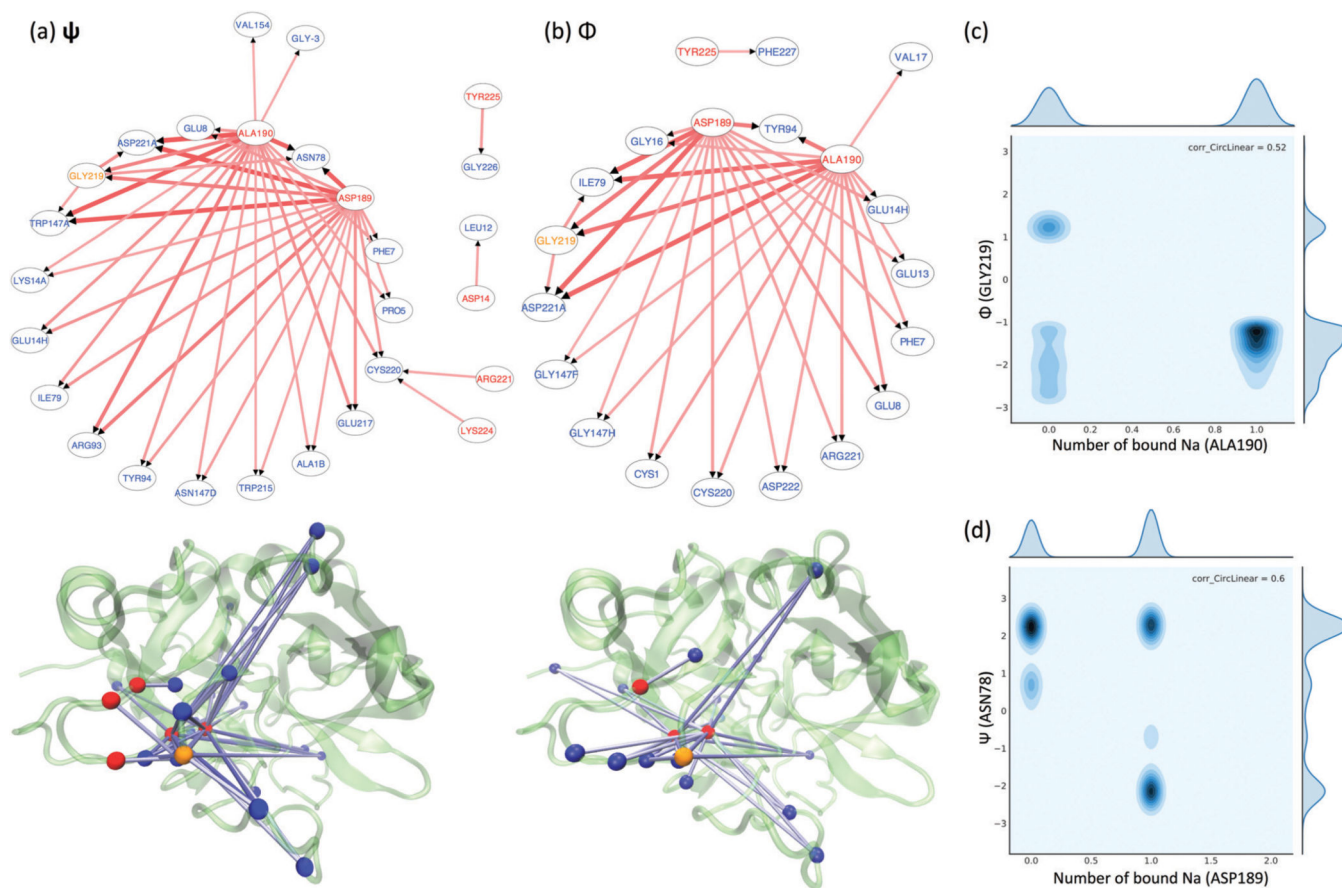
**Fig. 3.**
Allosteric responses to the Na$^+$-binding/unbinding. Correlation networks between residues' Na$^+$-occupancy and $\psi/\Phi$ angles are displayed in panels (a and b). The arrows point from the Na$^+$-bound residues to the highly responsive residues. In both the network graphs and structural mapping plots in panels (a and b), the Na$^+$-bound residues are indicated by red labels or red beads, while the responsive residues are indicated by blue ones. GLY219 is the only residue in orange, representing its Na$^+$-binding can affect other residues' torsion angles and its torsion angles are also responding other residues' Na$^+$-binding. Panels (c and d) illustrate the heat maps of the joint distributions of one residues' Na$^+$-occupancy and another residue's $\psi/\Phi$ angles. (c) Shows the most common heat map representing generalized allosteric responses, while (d) is the only case exhibiting "induced-fit"-like responses.
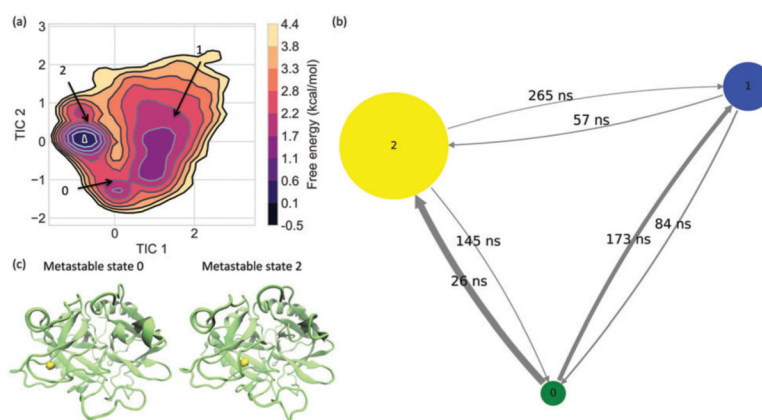
**Fig. 4.**

Metastable states from Hidden Markov modeling. (a) Three metastable states are identified from the free energy surface constructed along the first TICA components. The transition among these three metastable states are illustrated in panel (b), where the size of nodes and the thickness of the transition arrows are respectively proportional to the stationary probability of the corresponding states and transition matrix. The quantities above the arrows indicate the mean first passage times from one state to another. Structural representations for the $Na^+$-bound metastable states are shown in panel (c), where the bound $Na^+$ are indicated by the yellow isosurfaces with a number density of 0.5.