HHS Public Access

Author manuscript

J Biomol Struct Dyn. Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

J Biomol Struct Dyn. 2019 March ; 37(4): 982–999. doi:10.1080/07391102.2018.1445032.

Probing Light Chain Mutation Effects on Thrombin via Molecular Dynamics Simulations and Machine Learning

Jiajie Xiao[1,2], **Ryan L. Melvin**[1,3], and **Freddie R. Salsbury Jr.**[1,*]

[1]Department of Physics, Wake Forest University, Winston-Salem, USA

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

[3]Department of Mathematics and Statistics, Wake Forest University, Winston-Salem,USA

Abstract

Thrombin is a key component for chemotherapeutic and antithrombotic therapy development. As the physiologic and pathologic roles of the light chain still remain vague, here, we continue previous efforts to understand the impacts of the disease-associated single deletion of LYS9 in the light chain. By combining supervised and unsupervised machine learning methodologies and more traditional structural analyses on data from 10 μs molecular dynamics simulations, we show that the conformational ensemble of the ΔK9 mutant is significantly perturbed. Our analyses consistently indicate that LYS9 deletion destabilizes both the catalytic cleft and regulatory functional regions and result in some conformational changes that occur in tens to hundreds of nanosecond scaled motions. We also reveal that the twoforms of thrombin each prefer a distinct binding mode of a Na+ ion. We expand our understanding of previous experimental observations and shed light on the mechanisms of the LYS9 deletion associated bleeding disorder by providing consistent but more quantitative and detailed structural analyses than early studies in literature. With a novel application of supervised learning, i.e. the decision tree learning on the hydrogen bonding features in the wild-type and ΔK9 mutant forms of thrombin, we predict that seven pairs of critical hydrogen bonding interactions are significant for establishing distinct behaviors of wild-type thrombin and its ΔK9 mutant form. Our calculations indicate the LYS9 in the light chain has both localized and long-range allosteric effects on thrombin, supporting the opinion that light chain has an important role as an allosteric effector.

Keywords

Thrombin; Generalized allostery; Molecular Dynamics; Machine learning; Ion binding modes

Introduction

Thrombin is a multifunctional serine protease with critical coagulant and anticoagulant properties in blood coagulation cascades (Crawley, Zanardelli, Chion, & Lane, 2007; Enrico

*To whom correspondence should be addressed., Telephone: (336)758-4975, Facsimile: (336)758-6142 salsbufr@wfu.edu.

The authors of this manuscript declare no conflicts of interest. Email addresses for all authors are xiaoj12@wfu.edu (Jiajie Xiao), melvrl13@wfu.edu (Ryan L. Melvin), and salsbufr@wfu.edu (Freddie R. Salsbury Jr.).

Di Cera, 2008). Meanwhile, thrombin plays important roles in tumor growth, metastasis, angiogenesis and invasion (Adams et al., 2015; Kobrinsky & Karpatkin, 2009; Nierodzik & Karpatkin, 2006; Radjabi et al., 2008). Therefore, thrombin is an attractive drug target for antithrombotic therapy (Ageno et al., 2012; Schwienhorst, 2006; Zavyalova & Kopylov, 2015) and chemotherapeutic development (Lechner, Kollars, Gleiss, Kyrle, & Weltermann, 2007; Mukherjee et al., 2009).

Activated from its precursor prothrombin, human alpha thrombin is composed of a 36-residue light and a 259-residue heavy chains (also respectively referred to as “A” and “B” chains, as shown in Figure 1). These two chains are linked to each other through a disulfide bridge between CYS1 and CYS122 (in the chymotrypsin numbering scheme) and a network of salt bridges, ionic interactions and hydrogen bonds (Bode, Turk, & Karshikov, 1992; Carter, Vanden Hoek, Pryzdial, & MacGillivray, 2010; Enrico Di Cera, 2008). In addition to the catalytic triad, several known functional sites are also located in the heavy chain of thrombin (Enrico Di Cera, 2008). Such functional sites include hydrophobic 60s and hydrophilic gamma loops, 220s loop, and anion-binding exosite I and II (Figure 1). The 220s loop is also known as “sodium binding loop” due to its specific accommodation of a Na+ ion (E Di Cera et al., 1995; E. Zhang & Tulinsky, 1997) (Figure 1), which maximizes thrombin’s catalytic function (Bush, Nelson, & Di Cera, 2006; Dang, Guinto, & Cera, 1997; R De Cristofaro, Picozzi, Morosetti, & Landolfi, 1996; Orthner & Kosow, 1980). However, unlike the catalytic heavy chain, the light chain has been less studied and considered an unnecessary activation remnant for thrombin’s function (DiBella, Maurer, & Scheraga, 1995; Hageman, Endres, & Scheraga, 1975) until the discovery of severe bleeding phenotype associated mutations in the light chain (S Akhavan et al., 2000; Sepideh Akhavan, Rocha, Zeinali, & Mannucci, 1999; Lefkowitz et al., 2000; Sun, Burkart, Holahan, & Degen, 2000).

Moreover, the thrombin light chain has also been proposed to serve as a sensitive marker for detecting gastric cancer due to a decreased level of the circulating thrombin light chain in gastric cancer patients’ serum (though, the precise mechanism is unknown) (Ebert et al., 2005). These results suggest the light chain of thrombin should have a functional importance in its enzymatic properties and cancer development, although -- again -- the physiologic and pathologic roles of the light chain and the mechanisms behind relevant experimental observations are not yet fully understood (Carter et al., 2010; Papaconstantinou, Bah, & Di Cera, 2008).

One of two lysine residues -- LYS9 or LYS10 -- has been suggested as one of the key residue in the 36-residue light as its mutation may cause hypoprothrombinemia (S Akhavan et al., 2000). It was seen that patients with a deletion of LYS9/LYS10 presented deficient antigen levels and prothrombin activities and experienced severe bleeding (S Akhavan et al., 2000). Several *in vitro* experiments have shown that mutations of LYS9/LYS10 can cause dramatic perturbations of thrombin’s catalytic activities, including severely impaired interactions between thrombin and its substrates and attenuated sensitivity to sodium ions (Raimondo De Cristofaro et al., 2004, 2006; Papaconstantinou et al., 2008).

Previous computational studies using a single run of 18 ns molecular dynamics simulation have made progress into the molecular-level details, which suggests the some insertion loops as well as the catalytic triad could have subtle conformational changes due to the deletion of LYS9/LYS10 (Raimondo De Cristofaro et al., 2004, 2006). More recently, in addition tofast dynamic motions at the ps-ns timescale, thrombin's slow time scale dynamics in surface loops, have been revealed by NMR experiments and conventional and accelerated MD calculations (Fuglestad et al., 2012), suggesting longer time scale MD simulations are required to study allosteric changes in thrombin. Given LYS9/LYS10's significant relevance to disease and the mechanism by which a LYS9/LYS10 deleetion impairs thrombin has not been fully revealed yet, in this work, we continue the progress in understanding the role of thrombin's light chain by combining microsecond all-atom molecular dynamics simulations of wild-type thrombin and its mutant with the deletion of the LYS9 residue with several recently developed machine learning techniques. We performed five independent 1 μs all-atom MD simulations (10 μs of simulations in total) for thrombin in the wild-type and LYS9 deletion mutant forms to probe the mutant effects regarding the conformational properties. Our calculations reveal consistent but more quantitative and detailed light chain mutation effects than previous relevant experimental and computational studies. The more extensive sampling and longer simulations in this study also suggest the deletion of LYS9 in the light chain may induce more significant perturbations on thrombin's conformational ensembles than what was revealed previously. Through a novel application of decision tree learning on the hydrogen interaction features, we predict several key residue interactions that provide a new mechanistic insight into how the naturally occurring LYS9 deletion causes severe bleeding disorders. Furthermore, the workflow (Figure 2) that combines MD simulations and machine learning methods in this study should be not just successful in the present test case of LYS9 deletion, but also applicable in future work on other thrombin's mutations (S Akhavan et al., 2000; Sepideh Akhavan et al., 1999; Lefkowitz et al., 2000; Papaconstantinou et al., 2008; Sun et al., 2000) and any other important disease-associated mutants of readers' interest (such as EGFR mutations described in Zhao, Yang, Xiang, Gao, & Zeng, 2017 and other import mutations (Krishnamoorthy, Gajendrarao, Olivotto, & Yacoub, 2017; Mandal, Panda, & Das, 2017; Sutthibutpong, Rattanarojpong, & Khunrae, 2017; Thirumal Kumar et al., 2017; Tompa & Kadhirvel, 2017)).

Materials and Methods

Simulation systems

To gain further understanding of the light chain mutations' structural influences at the atomic level and in the microsecond regime, we simulated human alpha thrombin in wild-type and LYS9 deletion (also referred to as ΔK9 in the following context) mutant forms.

The initial structure of the wild-type thrombin was based on the protein structure in PDB 4DIH (Russo Krauss et al., 2012). The missing residues of thrombin in this PDB (19 out of 295 residues) were modeled through Modeller (Šali & Blundell, 1993). Other missing hydrogen atoms were added via VMD's psfgen package (Humphrey, Dalke, & Schulten, 1996), using default parameters. As no crystal structure of alpha thrombin with light chain mutations has been yet solved, here, the initial structure of the LYS9 deletion mutant was

modeled via Modeller using the thrombin structure in the same PDB 4DIH as the structural template. Missing residues and hydrogen atoms for the mutant were added via the same procedures we performed for the wild-type. As a result, the initial structures of the wild-type and mutant forms have a large extent of overlaps due to the exactly same coordinates for most atoms (Figure 1). The relatively long (microsecond) time scales used in our study should ensure systems relax to their intrinsic conformational ensembles. Therefore, we expect the conformations of the wild-type and the LYS9 deletion mutant to be well-sampled by our simulations (10 runs of one microsecond simulations in total).

Simulation Configurations

Given previous identifications of slow dynamic motion of thrombin's surface loops (Fuglestad et al., 2012) and the fact that initial structure of the mutant form is modeled based on the wild-type one, access to these longer time scale is necessary to achieve full relaxation from the wild-type structure based model and probe events that take place only on longer time scales (R. C. Godwin, Melvin, & Salsbury, 2015; R. Godwin, Gmeiner, & Salsbury, 2016; Salsbury Jr, 2010). Leveraging recent developments in GPU parallelization, we performed five one-microsecond-long all-atom MD simulations for each system, i.e. a total of 5 microseconds for wild-type thrombin and ΔK9 mutant respectively. These simulations were performed using the GPU-enabled ACEMD simulation package (Harvey & De Fabritiis, 2009) and Titan GPUs in Metrocubo workstations produced by Acellera.

To mimic the aqueous, we solvated the protein by putting an explicit TIP3P (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) water box around thrombin using a 1 nm padding in VMD (Humphrey et al., 1996). Since the activation of thrombin requires Na^+ cations, via the autoionize package in VMD, sodium chloride ions were added to neutralize the systems and the concentration of Na^+ was set to a typical experimental value of 0.125 M (Russo Krauss et al., 2012). All ionizable residues were considered in their default protonation state at physiological pH. These procedures result in a simulation cell with more than 31,000 atoms and enable a close examination of each form of thrombin at the single molecular level. Prior to simulation, all systems underwent 1,000 cycles of conjugate gradient minimization to avoid un-physical initial configurations. All simulation runs were performed with the CHARMM27 force field that has a CMAP correction for proteins (MacKerell et al., 1998; Mackerell, Feig, & Brooks, 2004) under NPT ensemble (constant particle number, temperature and pressure). The pressure was set at 1 atm using Berendsen pressure control (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984) with a 400 fs relaxation time. Using Langevin thermostat (Lemons, 1997) with a damping coefficient of 0.1, the temperature was maintained at 300 K. A 0.9 nm cutoff and 0.75 nm switching distance was applied for van der Waals and electrostatic forces. The smooth particle mesh Ewald method as implemented in ACEMD (Darden, York, & Pedersen, 1993; Harvey & De Fabritiis, 2009) was used to calculate the long-range electrostatics with 72 evenly spaced grid points in all three directions. The time step was set to 4 fs during the simulations due to a usage of the hydrogen mass repartitioning scheme (Feenstra, Hess, & Berendsen, 1999). Bonds involving hydrogen atoms were constrained using the SHAKE algorithm (van Gunsteren & Berendsen, 1977). Conformations sampled in the simulations were saved in the MD trajectory file in every 2,500 steps (i.e. 10 ps) for the balance of temporal resolution and

cost in storage. As a result, we have an extensive sampling of 100,000 conformations across one-microsecond time scale for each simulation run.

Processing and Analysis Methods

We first aligned all frames in MD trajectories to the initial structure of wild-type thrombin through rigid body translations and rotations to minimize the root-mean-square deviation to the alpha carbons of the common residues in the wild-type and ΔK9 forms. This alignment was performed with the 'align' class in Python library MDAnalysis (Gowers et al., 2016; Liu, Agrafiotis, & Theobald, 2009; Michaud-Agrawal, Denning, Woolf, & Beckstein, 2011; Theobald, 2005). As we found the relaxation time (~20 ns, according to the alpha carbon atom root-mean-squared distances to the initial structure shown in Figure S1) is much smaller than the simulation time (1000 ns) for all simulation runs, we performed our analyses on the whole trajectory for each system. Although we don't expect a global equilibrium can be reached and the full dynamics of thrombin can be captured in our simulations given the existence of μs~ms scale fluctuations revealed by experiments (Fuglestad et al., 2012), we observe that the extensive sampling in our 1 μs simulation can still result in very small block averaging errors and an approximate convergence (Figure S2). Moreover, our multiple runs of each type of simulations present similar sample distributions but not the same dynamic behaviors (Figure S1 & S3), which suggests a potential phase space sampling problem with previous computational work (Raimondo De Cristofaro et al., 2004, 2006).

Root-mean-squared fluctuations (RMSF) were computed for each alpha carbon in the wild-type and ΔK9 mutant thrombin as a first line examination of backbone mobility differences in these twoforms. The RMSF calculation was based on

$$RMSF(i) = \sqrt{\frac{1}{T}\Sigma_{t=1}^{T}\left(r_t^{\alpha} - < r_t^{\alpha} >\right)^2}, \quad [1]$$

where r_t^{α} is the coordinate vector of atom α in frame t and $< r_t^{\alpha} >$ denotes the average position vector of atom α over all frames.

To explore the similarities and differences of thrombin's conformational responses to the light chain mutation, we carried out Amorim-Hennig clustering analysis (De Amorim & Hennig, 2015; Ryan L. Melvin et al., 2016) on the concatenated trajectories of the common atoms in the wild-type and ΔK9 mutant thrombin. As a recently developed unsupervised learning technique, Amorim--Hennig (AH) clustering with a default Minkowski weight 2 (which intuitively corresponds to a Euclidean distance metric (De Amorim & Hennig, 2015)), has been shown its effectiveness and sensibilities on MD data (Ryan L. Melvin et al., 2016). In particular, without a requirement of specifying the number of clusters, this clustering method determines cluster sizes by optimizing the sihousette index -- a measure of how similar an element is to its own cluster relative to the next best cluster (Rousseeuw, 1987) -- of different grouping. As AH clustering iteratively adjusts the rescaling feature weights until convergence, it is able to identify non-spherical clusters among noise and the clustering results help differentiate local conformational changes sensitively (R.L. Melvin et

al., 2016). Therefore, we applied such an automatic and non-parametric clustering method to all MD trajectories via our Python implementation available on figshare (R. Melvin & Salsbury, 2016). Depending on the size of regions of interest, we clustered the common C \alpha atoms of (1) the entire molecule and the heavy atoms (not hydrogen) of (2) the catalytic triad and (3) the regulatory region, which explicitly includes the exosite I and II, 60s, 220s, and gamma loops (details about the residue ranges see Tables S1). To save computation time, all frames of conformations in the MD trajectories were re-sampled by a striding factor of 10 in the clustering analysis.

As an important metric of inter- and intra-molecular interactions, hydrogen bonds were detected and analyzed over all conformations we sampled in the simulations of wild-type and ΔK9 mutant thrombin. Using the HydrogenBondAnalysis class in MDAnalysis (Michaud-Agrawal et al., 2011), we searched for hydrogen bonds of at least intermediate strength as defined in Steiner, 2002. Under this definition, heavy atoms of the donor and acceptor pairs atoms must be within 0.32 nm and the bond angle of heavy atom-hydrogen-heavy must be greater than 120°. If such an intermediate-strength hydrogen bond was detected between two residues, we said these two residues formed a hydrogen bond pair. Then, given a frame of conformation, we constructed a binary feature vector of all possible pairs of residues with hydrogen bonding. This reduced feature set has been recently adopted as inputs for decision tree learning and helped identify allosteric response to three type of DNA damages (Ryan L. Melvin, Thompson, Godwin, Gmeiner, & Salsbury, 2017).

As one of the most interpretable machine learning method, the decision tree algorithm returns a fitted binary classifier based on the input features (Kingsford & Salzberg, 2008). By making a couple key decisions on the input feature's values, one can tell what class the input entity should belong to through the trained decision tree classifier. In the present study, we aim to employ this machine learning technique to identify key residue-residue hydrogen bonding motifs among all hydrogen-bonding pairs occurred in the wild-type and ΔK9 thrombin. Key decisions on the hydrogen bonding status were determined based on the comparisons on the Gini indexes from all splitting ways for each level of the decision tree. The Gini index (also called Gini impurity, denoted as I) is computed according to

$$I(X_m) = \sum_k p_{mk}(1 - p_{mk}), \quad [2]$$

, where p_{mk} is the probability of a conformation with an actual class label m (i.e. "WT" or "Mutant" here) being classified into class k in the node a training set X (Kingsford & Salzberg, 2008). If all conformations inside a node belong to the same class, the Gini index of that node is zero. A binary split on whether forming a hydrogen bond between a specific residue-residue pair can result in a more accurate classification if the Gini indexes of nodes in the splitting branch are closer to zero. A split resulting in smallest Gini indexes will be treated as top level in the decision tree and the conformations in that node will be separated into two branch of nodes accordingly. Conformations inside each node will be spited in the same way until the classification error rate reaches a user specific value. Therefore,

hydrogen bond motifs that distinguish the wild-type and mutant forms can be identified from the top several levels of the decision tree.

By calling the fitctree function in the Statistics and Machine Learning Toolbox of Matlab, we performed supervised learning on residue-residue hydrogen bonding features – an *N*-by-*M* input matrix, where *N* is the number of conformations and *M* is the number of residue-residue pairs with a possible hydrogen bond -- via a construction of decision trees. By fitting decision trees on the binary hydrogen bond trajectories (i.e. the *N*-by-*M* input matrix) of interactions among non-LYS9/LYS10 residues in the wild-type and mutant thrombin, we were able classify conformations of the wild-type and ΔK9 thrombin based on the knowledge of the presence of hydrogen bonds. In other words, construction of a binary decision tree classifier based on the features of hydrogen bonding statuses provides an interpretable model to differentiate the conformational ensembles of wild-type and ΔK9 mutant forms of thrombin. Furthermore, the trained decision tree suggests a list of important hydrogen bond interactions in the intramolecular communication network, offering testable predictions of critical interactions and suggestions for designs of further mutagenesis studies.

Since Na^+ binding has been thought as an allosteric effector of thrombin's enzymatic activation (Bush et al., 2006; Dang et al., 1997; R De Cristofaro et al., 1996; Orthner & Kosow, 1980; Xiao, Melvin, & Salsbury, 2017), we monitored the mean distance <d> between the sodium binding loop (220s loop) atoms and the nearest Na^+ ion. Based on the histograms of <d> in each type of simulations and structural examinations for each peak, a frame was defined as Na^+ bound (or say "Na^+ on") if the <d> associated with that frame was less than or equal to 0.93 nm and 0.95 nm respectively for wild-type and ΔK9 forms of thrombin. As seen in our previous work (Xiao et al., 2017), this mean distance captures the stable interior bindings rather than the surface ones and correlates well with the minimum distance (less than 0.4 nm) between the 220s loop atoms and the stably bound cation. Moreover, as we will see in the next section, rather than the minimum distance, the mean distance preserves some profound information about the relative position of the bound Na + to the 220s loop. Therefore, we used the mean distance <d> here to illustrate the existence of multiple interior binding modes.

For both the Na + bound and Na + unbound states, we calculated solvent accessible surface areas (SASA) of the catalytic triad and subpockets and S1–6 (residues involved see Table S1) and the whole protein. The measurement of SASA was carried out via the Shrake and Rupley algorithm (Shrake & Rupley, 1973) implemented by MDTraj. Comparisons of the distribution of these SASAs indicate direct and indirect impacts of the mutations in the light chain.

To dissect impacts of the light chain mutation on thrombin's conformational space, principle component analysis (PCA) was performed on the concatenated trajectories of the selected common atoms in the wild-type and mutant thrombin. This technique -- commonly used for dimension reduction in machine learning -- allowed us to project high dimensional structural information onto a reduced space. Here, we constructed the conformational free energy surfaces of the regulatory region (defined in the clustering analysis) and the catalytic cleft

that includes catalytic triad and S1–6 subpockets. By diagonalizing the covariance matrix for the heavy atoms in above regions of interest, we obtained dominant components capturing the majority of dynamic variance in the wild-type and mutant systems. By projecting the coordinates of corresponding heavy atoms in the aligned trajectories onto the eigenvectors with two largest eigenvalues, distinct conformations in the region of interest are expected to have distinguished projections in the reduced conformational space. Through binning the projections and converting the frequency of each bin intofree energy via

$$\Delta G = -\mathrm{k}T\ln\left(\frac{P}{P_0}\right), \quad [3]$$

where $\mathrm{k}T$ is the product of Boltzmann's constant k and temperature T and P_0 is the minimum frequency among all bins, we calculate the free energy change Delta G We then constructed the conformational free energy surfaces of the interested regions. The PCA decomposition was carried out via the "pca" module in PyEmma (Scherer et al., 2015) Python package.

The conformational ensembles in the free energy wells was visualized via the Tachyon render in VMD 1.9.2 (Humphrey et al., 1996). To show potential variances of the representative structure in the selected well, as previously suggested by R. L. Melvin & Salsbury, 2016, we visualized the conformational ensembles by showing the representative structure in solid and the other conformations in the same structural ensemble as shadows. Here, the representative structure was chosen as the one that is closest to the average conformation in the most populated bin in the selected well. 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, only the backbone using the new cartoon representation was displayed for simplicity of the image.

Results

LYS9 deletion reduces the structural rigidity of main regulatory regions

Root-mean-square fluctuations for the alpha carbons of ΔK9 (Mutant) and wild-type (WT) thrombin highlight what regions of thrombin's backbone flexibility are affected by the deletion of LYS9 (Figure 3). The time-averaged atomic fluctuations reveal most of these regions are known functional sites of thrombin, including 60s, 180s, 220s and gamma loops, exosite I and II. In particular, these sites exhibit increased mobilities in the ΔK9 mutant thrombin (Figure 3 a & b) than the wild-type one. As indicated by the colors in Figure 3b, the rigidity of the gamma and 60s loops is significantly reduced in mutant thrombin, though this mutation (ΔK9) is distant (> 4 nm) from the gamma and 60s loops.

On the hand, as shown by the left sub-figure in Figure 3b, the LYS9 deletion in the light chain appears to result in a slightly more rigid light chain. The residues near LYS9 present about 0.05 nm more fluctuations in the presence of LYS9. Similar to the C-terminus of the light chain, the backbones of the adjacent alpha helix (resid 164 to 168 in the sequential

numbering scheme, i.e SER129B to ALA132) in the heavy chain are also slightly more flexible in the wild-type thrombin.

Clustering across configurations highlights deceased stability of mutant thrombin

Clustering analysis of concatenated common residue trajectories for the wild-type and mutant thrombin offers quantitative estimations of the structural diversity. As shown in Figure 4, Amorim-Hennig (AH) clustering indicates the wild-type thrombin has more structurally stable ensemble and the mutant has more diverse structural ensembles. Only one primary backbone conformation is present in wild-type thrombin, while two extra conformational clusters contribute to 36% and 8% populations of the backbone conformations in the thrombin mutant (Figure 4a).

Except for C \alpha atoms, to take into account the side chain arrangement, we also clustered the heavy atoms of several regions of interest. Similar to the backbone clustering, the heavy atom clustering of the regulatory regions (60s, 220s and gamma loops, exosite I and II), catalytic triad demonstrates the conformational ground state is more likely to split into multiple conformationally metastable states for the thrombin mutant. In particular, the gamma loop establishes several different shapes in the mutant thrombin rather than the dominant one in the wild-type. The 60s loop also presents unique movement in the lack of LYS9 (Figure S4). While the conformational differences between the catalytic triad cluster 0 and 1 appear subtle (Figure S5), the side chains orientation of the catalytic triad in the AH detected cluster 1 becomes about 4-fold more likely to occur in the ΔK9 mutant thrombin than the wild-type.

LYS9 deletion results in perturbed hydrogen bonding network

Due to important roles of hydrogen bonding network in allosteric communication (Amor, Schaub, Yaliraki, & Barahona, 2016; Daily & Gray, 2009; Srivastava et al., 2016) and structural stability (Myers & Pace, 1996), hydrogen bonding statuses among residues were monitored and compared for both wild-type and ΔK9 mutant thrombin. While the LYS9 residue in the wild-type thrombin contributes to 15 hydrogen-bonding pairs that ΔK9 mutant thrombin can impossibly form, the mutant thrombin with a deletion of LYS9 presents 256 more unique residue-residue pairs with hydrogen bonds than the wild type. This increment in number of hydrogen-bonding pairs in the mutant thrombin is ascribable simply to the new hydrogen bonds formed in the heavy chain, since the numbers of hydrogen-bonding pairs within the light chain and between the light and heavy chains are larger in the wild-type due to an extra LYS9 (Table 1). On the other hand, as quantitatively assessed in Table 1, the wild-type thrombin has a higher mean value but slightly smaller standard deviation in the counts of the residue-residue hydrogen-bonding pairs regardless of the consideration of the LYS9 residue. These statistics imply the wild-type thrombin has a more stable conformational ensemble than the ΔK9 mutant thrombin and thrombin's hydrogen bonding network is perturbed by the deletion of the light chain residue LYS9 in terms of the occupancy and diversity of hydrogen bonds.

We compared the prevalence of each hydrogen bonding pair in the wild-type and mutant thrombin systems by subtracting each hydrogen bonding pair's occupancies in both systems.

Sixteen residues pairs exhibited a considerable difference -- having a difference in occupancy of more 33% between the two systems -- in hydrogen bonding interactions (Table S2). Out of 16 pairs, 10 have a much greater occupancies in the wild-type, indicating the wild-type thrombin have a more stable structure. The hydrogen bonds between the LYS9 and ASP116, and LYS9 and PRO5 are entirely missing in the mutant thrombin due to the lack of LYS9. Moreover, the wild-type form preserves several hydrogen bonds formed within the light chain (LEU12-PHE7, LYS10-LEU6) and heavy chains (LEU40-GLY147F, TYR94-LEU60, LYS147G-GLU39, ASP221A-ASP189, LEU60-ALA56, LEU144-GLN151). While the deletion of LYS9 destabilizes these hydrogen bonding interactions, the hydrogen bonds involved the LYS9 in the wild-type form are compensated by the nearby LYS10. Instead of LYS9, in the ΔK9 mutant thrombin, LYS10 starts selectively interacting with ASP116. However, this hydrogen bond interaction between the light and heavy chain is about half likely than the hydrogen bond between LYS9 and ASP116, which is selected by the wild-type thrombin. In addition, several hydrogen bonds (SER11-LEU6, LEU12-LEU6, ARG73-GLU39, TYR94-ASP102, THR147-GLU192) are preferential by the ΔK9 mutant thrombin (Table S2), which implying the significantly changes in hydrogen bonding network are not only located beside the LYS9 residue but also distant residues around catalytic cleft.

To identify underlying key hydrogen-bonding pairs of residues differentiating the wild-type and ΔK9 mutant thrombin, decision tree classifiers were constructed based on residue-residue hydrogen bonding features of the wild-type and mutant thrombin here. Via a decision tree with 169 levels of hydrogen-bonding pair decisions, we can at most classify the two types of thrombin's conformations as accurate as 99.9985%. Pruning such full tree by 166 levels, a simple decision tree as shown in Figure 5a still yields 96.21% classification accuracy (Figure S6).

The presences of hydrogen bonds among these six pairs of residues, which are (1) LEU12-PHE7 (2) LEU40-GLY147F (3) LEU12-LEU6 (4) TYR94-LEU59 (5) ASP221A-ASP189 (6) THR147-GLU192, establish distinct hydrogen bonding motifs of the wild-type and ΔK9 mutant thrombin. For example, the hydrogen bond motifs of (LEU12-PHE7:Yes) and (LEU12-PHE7:No, LEU40-GLY147F:Yes) cover 74.12% and 13.94% sampled conformations of wild-type thrombin, while these two motifs only occur in 3.54% of observed conformations without the K9 residue. Instead, (LEU12-PHE7:No, LEU40-GLY147F:No, LEU12-LEU6:Yes), (LEU12-PHE7:No, LEU40-GLY147F:No, LEU12-LEU6:No, TYR94-LEU59:No) and (LEU12-PHE7:No, LEU40-GLY147F :No, LEU12-LEU6:No, TYR94-LEU59:Yes, ASP221A-ASP 189:No) are mostly preferable by 92.67% sampled conformations of ΔK9 mutant thrombin. Further splits of the hydrogen bonding decision tree -- such as ASP221A-ASP189 and THR147-GLU192 -- yield a higher coverage of each type of conformations, though the purities in these leafs are less than the top level splits. The requirement of these hydrogen bonding decisions for a better coverage of the ΔK9 mutant implies a distribution shift in the conformational space due to the deletion effect. As we have seen above, the occupancies of hydrogen bonds among these six pairs of residues are 33.33% greater in one system than another (Table S2). Again, it is clear that considerable changes in hydrogen bonding network are not only localized but also long ranged allosteric (Figure 5b & c).

Weakened and perturbed Na^+ binding in LYS9 deletion mutant thrombin

The association and dissociation of a Na^+ ion have been observed in both wild-type and ΔK9 mutant thrombin. As assessed by the distance between the 220s loop (also known as sodium binding loop) and the nearest Na^+ ion, the bound Na^+ ion presents two binding modes onto the sodium binding loop in both simulations of wild-type and mutant thrombin (Figure 6). However, the first binding mode, where Na^+ is binding outer region of the 220s loop (as shown in Figure 6b), is preferable to the ΔK9 mutant thrombin. The second binding mode (Figure 6c), where the bound Na^+ is located at the inner the 220s loop and almost entering the S1 subpocket of thrombin, is more selected by the wild-type thrombin. The former binding mode has been seen in the crystal studies of thrombin (E Di Cera et al., 1995; E. Zhang & Tulinsky, 1997); the later one is consistent with the density distribution of Na^+ illustrated in the previous computational work (Kurisaki, Takayanagi, & Nagaoka, 2015).

Here, we simply defined the Na^+ bound and unbound states of thrombin without differentiating binding mode 1 and 2. The numbers of frames of the Na^+ bound and unbound states for wild-type and mutant thrombin are listed in Table 2. This statistics clearly indicates that the binding of Na^+ ion becomes about half likely for the ΔK9 mutant thrombin than the wild-type, suggesting a weakened affinity of Na^+ binding.

LYS9 deletion perturbs solvent accessible surface areas

As a reflection of conformations and meaningful quantity for interacting substrates, solvent accessible surface areas (SASA) of catalytic subpockets S1–6, catalytic triad and the whole protein were computed for wild-type and ΔK9 mutant thrombin in Na + bound and Na + unbound states. As shown in Figure 7, the wild-type and mutant thrombins exhibit significantly distinct distributions of SASAs of the catalytic cleft and event the whole protein.

The ΔK9 mutant thrombin in general presents a wider distribution of SASAs than the wild-type (Figure 7), suggesting more dynamic structural ensembles. In particular, when there is no Na + attaching to the interior of sodium binding loop, the mutant thrombin exhibits highly perturbed SASA distributions of S2, S6 and catalytic triad. The binding of a Na + ion make the mutant thrombin turn more wild-type-like in terms of the distribution of the catalytic subpockets' SASAs.

The catalytic triad's SASA of the mutant thrombin is altered dramatically, regardless of the status of Na + binding. As seen in Figure 7g, the multiple distinct peaks than the single one of the wild-type indicate the catalytic triad residues and their vicinity area establish less stable conformations than they have in the wild-type thrombin. The catalytic triad is more exposed when K9 is not present in the light chain. Similarly, although the wild-type thrombin has one residue than the ΔK9 mutant one, the deletion of the K9 in fact leads to a more solvent exposed surface area for the whole protein molecule.

LYS9 deletion disrupts thrombin's conformational free energy profiles

Conformational free energy profiles of the wild-type and ΔK9 mutant thrombin were estimated via principal component analysis, which is a very sensitive approach to reflect

dominant structural differences in the wild-type and mutant thrombin. Na + bound/unbound states were also considered as what we did in the SASA analysis, aiming to reveal the independent and combined influences of the LYS9 deletion and the known allosteric effector of Na + ions.

As shown in Figure 8, the conformational free energy surface of the regulatory regions, including the 60s, 220s and gamma loops and exosite I and II, is dramatically perturbed by the deletion of LYS9. The primary conformational free energy well of above regulatory sites shifts to a large region that is inaccessible for wild-type thrombin. While the wild-type thrombin primarily establishes one narrow well, the mutant thrombin presents multiple spread wells. Furthermore, the binding of Na^+ mainly alters populations of the wild-type thrombin among several conformations with subtle structural differences in the regulatory regions (Figure 8e). However, unlike wild-type thrombin, the Na^+ binding stabilizes some conformations of the regulatory regions that are only presented by the mutant thrombin rather than the wild-type. These conformations have a greater blockage of the catalytic cleft by the gamma loop but a more exposed catalytic triad due to the extension of the 60s loop (Figure 8e).

The conformational free energy surfaces of the catalytic cleft, including the catalytic triad and substrate pockets S1–6, also reveal that the LYS9 deletion causes a striking disruption of the shape of the catalytic site. While the wild-type thrombin illustrates a major stable conformational free energy well in Figure 9a & b, the mutant thrombin displays multiple separate wells in the free energy surfaces (Figure 9c & d). The catalytic clef of the ΔK9 mutant thrombin is thus not as stable as the one in the wild-type thrombin. The mutant's 60s loop is able toflip up and the end of the 220s loop exhibits more variances in its orientations (Figure 9e). Again, the binding of Na^+ to the mutant thrombin stabilizes the conformations that are not seen in the wild-type thrombin.

Discussion

Our quantitatively comparisons of the wild-type and ΔK9 mutant thrombin above consistently indicate significant impacts of the light chain residue LYS9 on thrombin in molecular level and structural ensemble aspect. By quantifying the distributions of possible conformational states in the wild-type and mutant forms, we reveal that thrombin's structural stability decreases in terms of more diverse population of the conformational states in the absence of the LYS9 (Figure 4). Note that previous urea-induced denaturation and disulfide scrambbling experiments also suggest that the overall structural stability of the wild-type thrombin should be higher than the ΔK9 form (Raimondo De Cristofaro et al., 2006). Our Amorim-Hennig clustering analysis presents experimentally consistent results, suggesting the effectiveness of this recently developed non-parametric clustering algorithm in MD analysis.

Through a total of 10 microseconds of all-atom molecular dynamics simulations, we see that the reduced stability in the ΔK9 mutant thrombin was ascribed to the increased mobilities of thrombin's regulatory regions including the 60s, 180s, 220s and gamma loops, and exosite I and II (Figure 3). The conformational free energy surfaces of these regulatory regions

confirm that the increased mobility results in conformational changes and metastable states. While previous single run of 18ns MD simulation reveals the deletion of LYS9 may lead to subtle conformational changes in the thrombin's catalytic triad and the 60s loop (Raimondo De Cristofaro et al., 2004), our multiple runs of longer simulations show that the ΔK9 form should have a much more perturbed catalytic pocket and 60s loop. Figure 9 illustrates several different distortions of the catalytic cleft in the mutant thrombin. Basically, the 60s loop is destabilized by the light chain deletion of LYS9. When a Na+ ion is not bound to the 220s loop, 60s loop can flip up and lead to a fully exposed S2 subpocket (Figure 9e). In the mutant thrombin, the movement of residues in S3 subpocket also can result in a self-inhibited conformation that has previously been seen in the crystal structure of thrombin mutant D102N (Gandhi, Chen, Mathews, & Di Cera, 2008).

As indicated in Figure 7, all subpockets are distorted by the LYS9 deletion. As a long range allosteric effect, different preferential orientations of the side chains in the mutant thrombin also lead to population shifts in the conformational free energy surfaces of the catalytic cleft comparing to the wild-type thrombin (Figure 9). Moreover, the hydrophilic gamma loop, which restricts the access of substrate to the catalytic cleft, is destabilized in the ΔK9 form and establishes a variety of poses in the slow time scale (tens to hundreds of nanoseconds) motions (Figure 8). Therefore, our multiple runs of microsecond simulations and comprehensive analyses reveal additional and more significant conformational changes that have not been previously uncovered. This is probably due to the limitation of the crystallization as well as the limited time scale and sampling in previous studies.

The hydrogen bond detection provide further insights of the structure and intra-molecular interactions, considering the fact that thrombin is an allosteric enzyme and that the light chain interacts with the heavy chain via a network of hydrogen bonding interaction and salt bridges. Our statistics of the hydrogen bonds indicate an overall weakened hydrogen bonding network in the absence of LYS9, consisting with the consequence of destabilization we have discussed before. Our calculation suggests the salt bridge between LYS9 and ASP116 is a critical interaction between the light chain and heavy chain. Although LYS10 tends to compensate the lost of this interaction, the likelihood of forming this salt bridge in the ΔK9 mutant is about a half of the wild-thrombin.

In order to understand how the light chain deletion conducts an allosteric signal to the distant region of the protein, it is necessary to identify hidden patterns that can distinguish the conformational ensembles of the wild-type and mutant forms. As hydrogen bonds play an important role in allosteric communication (Amor, Schaub, Yaliraki, & Barahona, 2016; Daily & Gray, 2009; Srivastava et al., 2016), we searched the simplest hydrogen bonding motifs via decision tree learning. Via this machine learning methodology, in additional to the hydrogen bond between LYS9 and ASP116, we predict six pairs of such interactions -- (1) LEU12-PHE7, (2) LEU40-GLY147F, (3) LEU12-LEU6, (4) TYR94-LEU59, (5) ASP221A-ASP189, and (6) THR147-GLU192 -- play a profound role in thrombin's intra-molecular signaling and structural stability. Different hydrogen bonding statuses of these pairs are likely responsible for propagating the signal from the light chain and causing allosteric effects. We suggest future experimental and computational mutation studies should first consider these residues for further understanding of thrombin's allostery.

Na^+ ions, known as an allosteric modulator of thrombin, has been revealed ~2-fold frequent to bind to the thrombin in wild-type form than the ΔK9 one in our simulations. Such results are consistent with the observation that ΔK9 thrombin has a half gain in the intrinsic fluorescence intensity of upon sodium concentration than wild-type form does (Raimondo De Cristofaro et al., 2006). Furthermore, although Na^+ ions are able to recognize the sodium-binding loop, it is the first time to see that the bound Na^+ ion presents a different preference of binding in the wild-type and mutant forms. In the ΔK9 mutant thrombin, a Na^+ tends to stay in the outer site of the 220s loop that has been seen in the crystal structure (E Di Cera et al., 1995; E. Zhang & Tulinsky, 1997). For the wild-type thrombin, the interior of the 220s loop is more favorable for the bound Na^+. The statistics also suggest the deeper binding is more thermodynamically stable than the one solved by the crystallography studies before. This find provides a mechanistic insight into the reduced affinity of Na^+ to thrombin considering the fact that the more stable deep binding is not preferred by the Δ K9 mutant form. This result triggers us interest to investigate the association/dissociation pathways of Na + and functional differences of these two binding modes in the future study.

From the SASA analysis and free energy surface construction, we can see the binding of Na + ion stabilizes the structural ensembles of thrombin. Subpockets S2, S3 and S6 appear more wild-type like after Na + binding according to the SASA distribution (Figure 7). On the other hand, according to figure 8 & Figure 9, the regulatory regions and the whole catalytic cleft tends to be stabilized into some unique conformations that are inaccessible for the wild-type form. Therefore, the LYS9 deletion appears to cause a primary influence on the entire protein, while the binding of a Na^+ ion offers a secondary perturbation.

Conclusions

Thrombin has been considered as an allosteric enzyme (see the review by Enrico Di Cera, 2008). In this study, we see significant conformational differences in the wild-type and ΔK9 forms of thrombin. The deletion of the LYS9 perturbs the interaction network within the light, heavy chains as well as between them. Such perturbations cause significant and profound conformational changes in the catalytic and regulatory sites. Our calculations presented here are consistent with previous experimental and computational work. Our results support the opinion that the light chain is the allosteric effector instead of the activation remnant discussed by Carter et al., 2010. Moreover, it expands our understanding of how the disease-associated LYS9 deletion impairs thrombin's biophysical properties from the following aspects.

Firstly, our study shows that there are more significant conformational changes occurring in the longer time scale (tens to hundreds of nanoseconds) in the ΔK9 thrombin. For instance, the 60s loop can completely flip up and result in a fully exposed S2 subpocket in the absence of LYS9. The gamma loop also establishes more diverse poses in the ΔK9 thrombin, which may further affect substrates' recognition to the catalytic pocket. We provide a more detailed dynamic perspective of ΔK9 thrombin.

Secondly, we identify a list of residues with critical hydrogen bonds. These residues, i.e. LYS9, ASP116, LEU12, PHE7, LEU40, GLY147F, LEU6, TYR94, LEU59, ASP221A,

ASP189, THR147 and GLU192, are likely playing an important role in the intramolecular interaction network of thrombin. We hypothesize that the hydrogen bonds among these residues are responsible for transmitting the allosteric signals from the light chain to the entire protein. We suggest these residues should be first looked at in the future mutagenesis study in order to understand thrombin's allosteric pathway and response to mutations.

Thirdly, we reveal a second Na+ binding mode in thrombin for the first time. We show that the wild-type and ΔK9 thrombins present different preferences between the two Na + binding modes. These observations further suggest a mechanistic explanation of the reduced sensitivity to the Na+ binding for the ΔK9 thrombins. Meanwhile, we clarify that the binding of a Na+ ion should play a secondary role in regulating thrombin's conformational ensembles comparing with the LYS9 deletion.

Moreover, this work combines microsecond molecular dynamics simulations and machine learning techniques to probe the LYS9/LYS10 deletion's influences on thrombin and the corresponding mechanisms. The decision tree learning uncovers hydrogen bond motifs that distinguish the wild-type thrombin from its ΔK9 mutant. These critical hydrogen bonds are likely key components that transmit the allosteric signals to the either protein from the single deletion in the light chain. Although only one light chain mutation has been discussed here, it is necessary to study this important mutation first as a test case for understanding other disease-related mutations in the light chain. Given the new insights into how and why the LYS9 deletion perturbs thrombin, we will work on other thrombin mutations in the future. The methodology presented in this work should be also applicable for other systems if one interests in understanding the molecular response to any type of perturbations such as mutations and ligand binding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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

Abbreviations and symbols

ΔK9	LSY9 deletion form
MD	Molecular dynamics
RMSF	root-mean-square fluctuations
Na+	sodium
PCA	principal component analysis

NPT	isothermal-isobaric ensemble
PME	particle mesh Ewald

References

Adams GN, Rosenfeldt L, Frederick M, Miller W, Waltz D, Kombrinck K, … Palumbo JS (2015). Colon cancer growth and dissemination relies upon thrombin, Stromal PAR-1, and fibrinogen. Cancer Research, 75(19), 4235–4243. 10.1158/0008-5472.CAN-15-0964 [PubMed: 26238780]

Ageno W, Gallus AS, Wittkowsky A, Crowther M, Hylek EM, & Palareti G (2012). Oral Anticoagulant Therapy. Chest, 141(2), e44S–e88S. 10.1378/chest.11-2292 [PubMed: 22315269]

Akhavan S, Mannucci PM, Lak M, Mancuso G, Mazzucconi MG, Rocino A, … Perkins SJ (2000). Identification and three-dimensional structural analysis of nine novel mutations in patients with prothrombin deficiency. Thrombosis and Haemostasis, 84(6), 989–97. [PubMed: 11154146]

Akhavan S, Rocha E, Zeinali S, & Mannucci PM (1999). Gly319 → Arg substitution in the dysfunctional prothrombin Segovia. British Journal of Haematology, 105(3), 667–669. 10.1046/j.1365-2141.1999.01423.x [PubMed: 10354128]

Amor BRC, Schaub MT, Yaliraki SN, & Barahona M (2016). Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. Nature Communications, 7, 12477 10.1038/ncomms12477

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, & Haak JR (1984). Molecular dynamics with coupling to an external bath. Journal of Chemical Physics, 81(8), 3684–3690. 10.1063/L448118

Bode W, Turk D, & Karshikov A (1992). The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. Protein Science : A Publication of the Protein Society, 1(4), 426–71. 10.1002/pro.5560010402 [PubMed: 1304349]

Bush LA, Nelson RW, & Di Cera E (2006). Murine thrombin lacks Na+ activation but retains high catalytic activity. Journal of Biological Chemistry, 281(11), 7183–7188. 10.1074/jbc.M512082200 [PubMed: 16428384]

Carter ISR, Vanden Hoek AL, Pryzdial ELG, & MacGillivray RTA (2010). Thrombin A-Chain: Activation Remnant or Allosteric Effector? Thrombosis, 2010, 1–9. 10.1155/2010/416167

Crawley JTB, Zanardelli S, Chion CKNK, & Lane DA (2007). The central role of thrombin in hemostasis. Journal of Thrombosis and Haemostasis, 5(SUPPL. 1), 95–101. 10.1111/j.1538-7836.2007.02500.x [PubMed: 17635715]

Daily MD, & Gray JJ (2009). Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. PLoS Computational Biology, 5(2). 10.1371/journal.pcbi.1000293

Dang QD, Guinto ER, & Cera E Di. (1997). Rational engineering of activity and specificity in a serine protease. Nature Biotechnology, 15(2), 146–149. 10.1038/nbt0297-146

Darden T, York D, & Pedersen L (1993). Particle mesh Ewald: An N· log(N) method for Ewald sums in large systems. Journal of Chemical Physics, 95(12), 1008910092. 10.1063/L464397

Davie E, & Kulman J (2006). An Overview of the Structure and Function of Thrombin. Seminars in Thrombosis and Hemostasis, 32(S 1), 003–015. 10.1055/s-2006-939550

De Amorim RC, & Hennig C (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. Information Sciences, 324, 126145. 10.1016/j.ins.2015.06.039

De Cristofaro R, Akhavan S, Altomare C, Carotti A, Peyvandi F, & Mannucci PM (2004). A Natural Prothrombin Mutant Reveals an Unexpected Influence of A-chain Structure on the Activity of Human α-Thrombin. Journal of Biological Chemistry, 279(13), 13035–13043. 10.1074/jbc.M312430200 [PubMed: 14722067]

De Cristofaro R, Carotti A, Akhavan S, Palla R, Peyvandi F, Altomare C, & Mannucci PM (2006). The natural mutation by deletion of Lys9 in the thrombin A-chain affects the pKa value of catalytic

residues, the overall enzyme's stability and conformational transitions linked to Na+ binding. FEBS Journal, 273(1), 159169. 10.1111/j.1742-4658.2005.05052.x [PubMed: 16367756]

De Cristofaro R, Picozzi M, Morosetti R, & Landolfi R (1996). Effect of sodium on the energetics of thrombin-thrombomodulin interaction and its relevance for protein C hydrolysis. Journal of Molecular Biology, 258(1), 190–200. 10.1006/jmbi.1996.0242 [PubMed: 8613987]

Di Cera E (2008). Thrombin. Molecular Aspects of Medicine, 29(4), 203–254. 10.1016/j.mam.2008.01.001 [PubMed: 18329094]

Di Cera E, Guinto ER, Vindigni A, Dang QD, Ayala YM, Wuyi M, & Tulinsky A (1995). The Na+ Binding Site of Thrombin. Journal of Biological Chemistry, 270(38), 22089–22092. 10.1074/jbc.270.38.22089 [PubMed: 7673182]

DiBella EE, Maurer MC, & Scheraga HA (1995). Expression and folding of recombinant bovine prethrombin-2 and its activation to thrombin. Journal of Biological Chemistry, 270(1), 163–169. 10.1074/jbc.270.L163 [PubMed: 7814368]

Ebert MPA, Lamer S, Meuer J, Malfertheiner P, Reymond M, Buschmann T, … Seibert V (2005). Identification of the thrombin light chain a as the single best mass for differentiation of gastric cancer patients from individuals with dyspepsia by proteome analysis. Journal of Proteome Research, 4(2), 586–590. 10.1021/pr049771i [PubMed: 15822938]

Feenstra KA, Hess B, & Berendsen HJC (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. Journal of Computational Chemistry, 20(8), 786–798. 10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B

Fuchs JE, Huber RG, Waldner BJ, Kahler U, Von Grafenstein S, Kramer C, & Liedl KR (2015). Dynamics govern specificity of a protein-protein interface: Substrate recognition by thrombin. PLoS ONE, 10(10), 1–14. 10.1371/journal.pone.0140713

Fuglestad B, Gasper PM, McCammon JA, Markwick PRL, & Komives EA (2013). Correlated Motions and Residual Frustration in Thrombin. The Journal of Physical Chemistry B, 117(42), 12857–12863. 10.1021/jp402107u [PubMed: 23621631]

Fuglestad B, Gasper PM, Tonelli M, McCammon JA, Markwick PRL, & Komives EA (2012). The Dynamic Structure of Thrombin in Solution. Biophysical Journal, 103(1), 79–88. https://doi.org/10.1016Zj.bpj.2012.05.047 [PubMed: 22828334]

Gandhi PS, Chen Z, Mathews FS, & Di Cera E (2008). Structural identification of the pathway of long-range communication in an allosteric enzyme. Proceedings of the National Academy of Sciences, 105(6), 1832–1837. 10.1073/pnas.0710894105

Godwin RC, Melvin R, & Salsbury FR (2015). Molecular Dynamics Simulations and Computer-Aided Drug Discovery. In Zhang W (Ed.), Methods in Pharmacology and Toxicology (pp. 1–30). Springer New York 10.1007/7653_2015_41

Godwin R, Gmeiner W, & Salsbury FR (2016). Importance of long-time simulations for rare event sampling in zinc finger proteins. Journal of Biomolecular Structure and Dynamics, 34(1), 125–134. 10.1080/07391102.2015.1015168 [PubMed: 25734227]

Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, … Beckstein O (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference, (Scipy), 98–105.

Hageman TC, Endres GF, & Scheraga HA (1975). Mechanism of action of thrombin on fibrinogen on the role of the a chain of bovine thrombin in specificity and in differentiating between thrombin and trypsin. Archives of Biochemistry and Biophysics, 171(1), 327–336. 10.1016/0003-9861(75)90039-9 [PubMed: 1103742]

Harvey MJ, & De Fabritiis G (2009). An implementation of the smooth particle mesh Ewald method on GPU hardware. Journal of Chemical Theory and Computation, 5(9), 2371–2377. 10.1021/ct900275y [PubMed: 26616618]

Humphrey W, Dalke A, & Schulten K (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14(October 1995), 33–38. 10.1016/0263-7855(96)00018-5 [PubMed: 8744570]

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983). Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics, 79(2), 926–935. 10.1063/L445869

Kingsford C, & Salzberg S (2008). What are decision trees? Nature Biotechnology, 26(9), 1011–1013. 10.1038/nbt0908-1011.What

Kobrinsky B, & Karpatkin S (2009). The Role of Thrombin in Tumor Biology In Thrombin (pp. 161–172). New York, NY: Springer New York 10.1007/978-0-387-09637-7_9

Krishnamoorthy N, Gajendrarao P, Olivotto I, & Yacoub M (2017). Impact of disease-causing mutations on inter-domain interactions in cMyBP-C: a steered molecular dynamics study. Journal of Biomolecular Structure and Dynamics, 35(9), 1916–1922. 10.1080/07391102.2016.1199329 [PubMed: 27267291]

Kurisaki I, Takayanagi M, & Nagaoka M (2015). Toward understanding allosteric activation of thrombin: A conjecture for important roles of unbound Na+ molecules around thrombin. Journal of Physical Chemistry B, 119(9), 3635–3642. 10.1021/jp510657n

Lechner D, Kollars M, Gleiss A, Kyrle PA, & Weltermann A (2007). Chemotherapy-induced thrombin generation via procoagulant endothelial microparticles is independent of tissue factor activity. Journal of Thrombosis and Haemostasis, 5(12), 2445–2452. 10.1111/j.1538-7836.2007.02788.x [PubMed: 17922809]

Lefkowitz JB, Haver T, Clarke S, Jacobson L, Weller A, Nuss R, … Hathaway WE (2000). The prothrombin Denver patient has two different prothrombin point mutations resulting in Glu-300Lys and Glu-309Lys substitutions. British Journal of Haematology, 108(1), 182–187. 10.1046/j.1365-2141.2000.01810.x [PubMed: 10651742]

Lemons DS (1997). Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. American Journal of Physics, 65(11), 1079 10.1119/L18725

Liu P, Agrafiotis DK, & Theobald DL (2009). Fast determination of the optimal rotational matrix for macromolecular superpositions. Journal of Computational Chemistry, 31(16), 1561–1563. 10.1002/jcc.21439

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, … Karplus M (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. The Journal of Physical Chemistry B, 102(18), 3586–3616. 10.1021/jp973084f [PubMed: 24889800]

Mackerell AD, Feig M, & Brooks CL (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. Journal of Computational Chemistry, 25(11), 1400–1415. 10.1002/jcc.20065 [PubMed: 15185334]

Mandal RS, Panda S, & Das S (2017). In silico prediction of drug resistance due to S247R mutation of Influenza H1N1 neuraminidase protein. Journal of Biomolecular Structure and Dynamics, 1102, 1–15. 10.1080/07391102.2017.1305295

Melvin RL, Godwin RC, Xiao J, Thompson WG, Berenhaut KS, & Salsbury FR (2016). Uncovering Large-Scale Conformational Change in Molecular Dynamics without Prior Knowledge. Journal of Chemical Theory and Computation, 12(12), 6130–6146. 10.1021/acs.jctc.6b00757 [PubMed: 27802394]

Melvin RL, Godwin RC, Xiao J, Thompson WG, Berenhaut KS, & Salsbury FR (2016). Uncovering Large-Scale Conformational Change in Molecular Dynamics without Prior Knowledge. Journal of Chemical Theory and Computation, 12(12). 10.1021/acs.jctc.6b00757

Melvin RL, & Salsbury FR (2016). Visualizing ensembles in structural biology. Journal of Molecular Graphics and Modelling, 67, 44–53. 10.1016/jjmgm.2016.05.001 [PubMed: 27179343]

Melvin RL, Thompson WG, Godwin RC, Gmeiner WH, & Salsbury FR (2017). MutSα's Multi-Domain Allosteric Response to Three DNA Damage Types Revealed by Machine Learning. Frontiers in Physics, 5(3), 10 10.3389/fphy.2017.00010

Melvin R, & Salsbury F (2016, 1 1). HDBSCAN and Amorim-Hennig for MD. Figshare. 10.6084/m9.figshare.3398266.v1

Michaud-Agrawal N, Denning EJ, Woolf TB, & Beckstein O (2011).MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of Computational Chemistry, 32(10), 2319–2327. 10.1002/jcc.21787 [PubMed: 21500218]

Mukherjee SD, Swystun LL, Mackman N, Wang J-G, Pond G, Levine MN, & Liaw PC (2009). Impact of Chemotherapy on Thrombin Generation and on the Protein C Pathway in Breast Cancer Patients. Pathophysiology of Haemostasis and Thrombosis, 37(2–4), 88–97. 10.1159/000324166

Myers JK, & Pace CN (1996). Hydrogen bonding stabilizes globular proteins. Biophysical Journal, 71(4), 2033–2039. 10.1016/S0006-3495(96)79401-8 [PubMed: 8889177]

Nierodzik ML, & Karpatkin S (2006). Thrombin induces tumor growth, metastasis, and angiogenesis: Evidence for a thrombin-regulated dormant tumor phenotype. Cancer Cell, 10(5), 355–362. https://doi.org/10.1016Zj.ccr.2006.10.002 [PubMed: 17097558]

Orthner CL, & Kosow DP (1980). Evidence that human α-thrombin is a monovalent cation-activated enzyme. Archives of Biochemistry and Biophysics, 202(1), 63–75. 10.1016/0003-9861(80)90406-3 [PubMed: 7396537]

Papaconstantinou ME, Bah A, & Di Cera E (2008). Role of the A chain in thrombin function. Cellular and Molecular Life Sciences, 65(12), 1943–1947. 10.1007/s00018-008-8179-y [PubMed: 18470478]

Radjabi AR, Sawada K, Jagadeeswaran S, Eichbichler A, Kenny HA, Montag A, … Lengyel E (2008). Thrombin Induces Tumor Invasion through the Induction and Association of Matrix Metalloproteinase-9 and 1-Integrin on the Cell Surface. Journal of Biological Chemistry, 283(5), 2822–2834. 10.1074/jbc.M704855200 [PubMed: 18048360]

Rousseeuw PJ (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20(C), 5365. 10.1016/0377-0427(87)90125-7

Russo Krauss I, Merlino A, Randazzo A, Novellino E, Mazzarella L, & Sica F (2012). High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. Nucleic Acids Research, 40(16), 8119–8128. 10.1093/nar/gks512 [PubMed: 22669903]

Šali A, & Blundell TL (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. Journal of Molecular Biology, 234(3), 779–815. 10.1006/jmbi.1993.1626 [PubMed: 8254673]

Salsbury FR, Jr (2010). Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Current Opinion in Pharmacology, 10(6), 738–744. https://doi.org/10.1016Zj.coph.2010.09.016 [PubMed: 20971684]

Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, … Noé F (2015). PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. Journal of Chemical Theory and Computation, 11(11), 5525–5542. 10.1021/acs.jctc.5b00743 [PubMed: 26574340]

Schwienhorst A (2006). Direct thrombin inhibitors - a survey of recent developments. Cellular and Molecular Life Sciences : CMLS, 63(23), 2773–91. 10.1007/s00018-006-6219-z [PubMed: 17103113]

Shrake A, & Rupley JA (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. Journal of Molecular Biology, 79(2), 351–371. 10.1016/0022-2836(73)90011-9 [PubMed: 4760134]

Srivastava A, Tracka MB, Uddin S, Casas-Finet J, Livesay DR, & Jacobs DJ (2016). Mutations in Antibody Fragments Modulate Allosteric Response Via Hydrogen-Bond Network Fluctuations. Biophysical Journal, 110(9), 1933–1942. 10.1016/j.bpj.2016.03.033 [PubMed: 27166802]

Steiner T (2002). The hydrogen bond in the solid state. Angew. Chem. Int. Ed, 41(1), 49–76. 10.1002/1521-3773(20020104)41:1<48::AID-ANIE48>3.0.C0;2-U

Sun WY, Burkart MC, Holahan JR, & Degen SJ (2000). Prothrombin San Antonio: a single amino acid substitution at a factor Xa activation site (Arg320 to His) results in dysprothrombinemia. Blood, 95(2), 711–4. [PubMed: 10627484]

Sutthibutpong T, Rattanarojpong T, & Khunrae P (2017). Effects of helix and fingertip mutations on the thermostability of xyn11A investigated by molecular dynamics simulations and enzyme activity assays. Journal of Biomolecular Structure and Dynamics, 1102, 1–15. 10.1080/07391102.2017.1404934

Theobald DL (2005). Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. Acta Crystallographica Section A: Foundations of Crystallography, 61(4), 478–480. 10.1107/S0108767305015266 [PubMed: 15973002]

Thirumal Kumar D, George Priya Doss C, Sneha P, Tayubi IA, Siva R,Chakraborty C, & Magesh R (2017). Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular dynamics approach. Journal of Biomolecular Structure and Dynamics, 35(5), 917–928. 10.1080/07391102.2016.1166456 [PubMed: 27125723]

Tompa DR, & Kadhirvel S (2017). Molecular dynamics of a far positioned SOD1 mutant V14M reveals pathogenic misfolding behavior. Journal of Biomolecular Structure and Dynamics, 1102, 1–14. 10.1080/07391102.2017.1407675

van Gunsteren WF, & Berendsen HJC (1977). Algorithms for macromolecular dynamics and constraint dynamics. Molecular Physics, 34(5), 1311–1327. 10.1080/00268977700102571

Xiao J, Melvin RL, & Salsbury FR (2017). Mechanistic insights into thrombin's switch between "slow" and "fast" forms. Phys. Chem. Chem. Phys, 19(36), 2452224533. 10.1039/C7CP03671J [PubMed: 28849814]

Zavyalova E, & Kopylov A (2015). Exploring potential anticoagulant drug formulations using thrombin generation test. Biochemistry and Biophysics Reports, 5, 111–119. 10.1016/j.bbrep.2015.11.011 [PubMed: 28955812]

Zhang E, & Tulinsky A (1997). The molecular environment of the Na+ binding site of thrombin. Biophysical Chemistry, 63(2–3), 185–200. 10.1016/S0301-4622(96)02227-2 [PubMed: 9108691]

Zhao F-L, Yang G-H, Xiang S, Gao D-D, & Zeng C (2017). In silico analysis of the effect of mutation on epidermal growth factor receptor in non-small-cell lung carcinoma: from mutational analysis to drug designing. Journal of Biomolecular Structure and Dynamics, 35(2), 427–434. 10.1080/07391102.2016.1146165 [PubMed: 26813338]

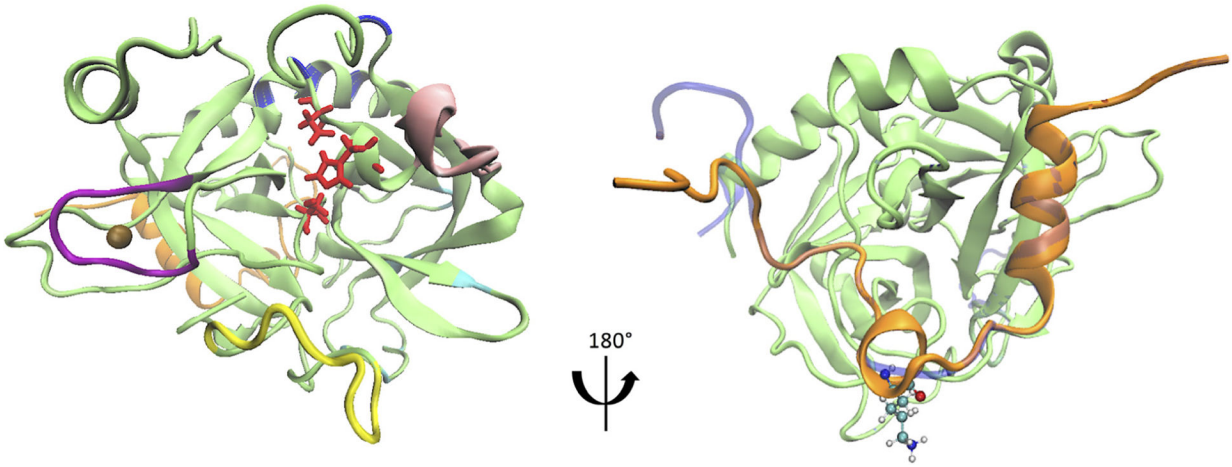

Figure 1.

Thrombin-Na+ complex structure and thrombin mutant with the ΔLYS9 deletion. (a) The thrombin structure is visualized based on PDB 4DIH and is shown in green and orange for the heavy and light chains. The 60s (in pink), 220s (in purple), and gamma (in yellow) loops and exosite I (in cyan) and II (in blue) are highlighted, as these are known regulatory regions. The side chains of the catalytic triad are displayed in red. The bound Na {+} is shown as a brown bead. (b) The side chain of LYS9 in the light chain is indicated in CPK representation. The modeled initial structure of ΔK9 mutant thrombin is shown in blue.

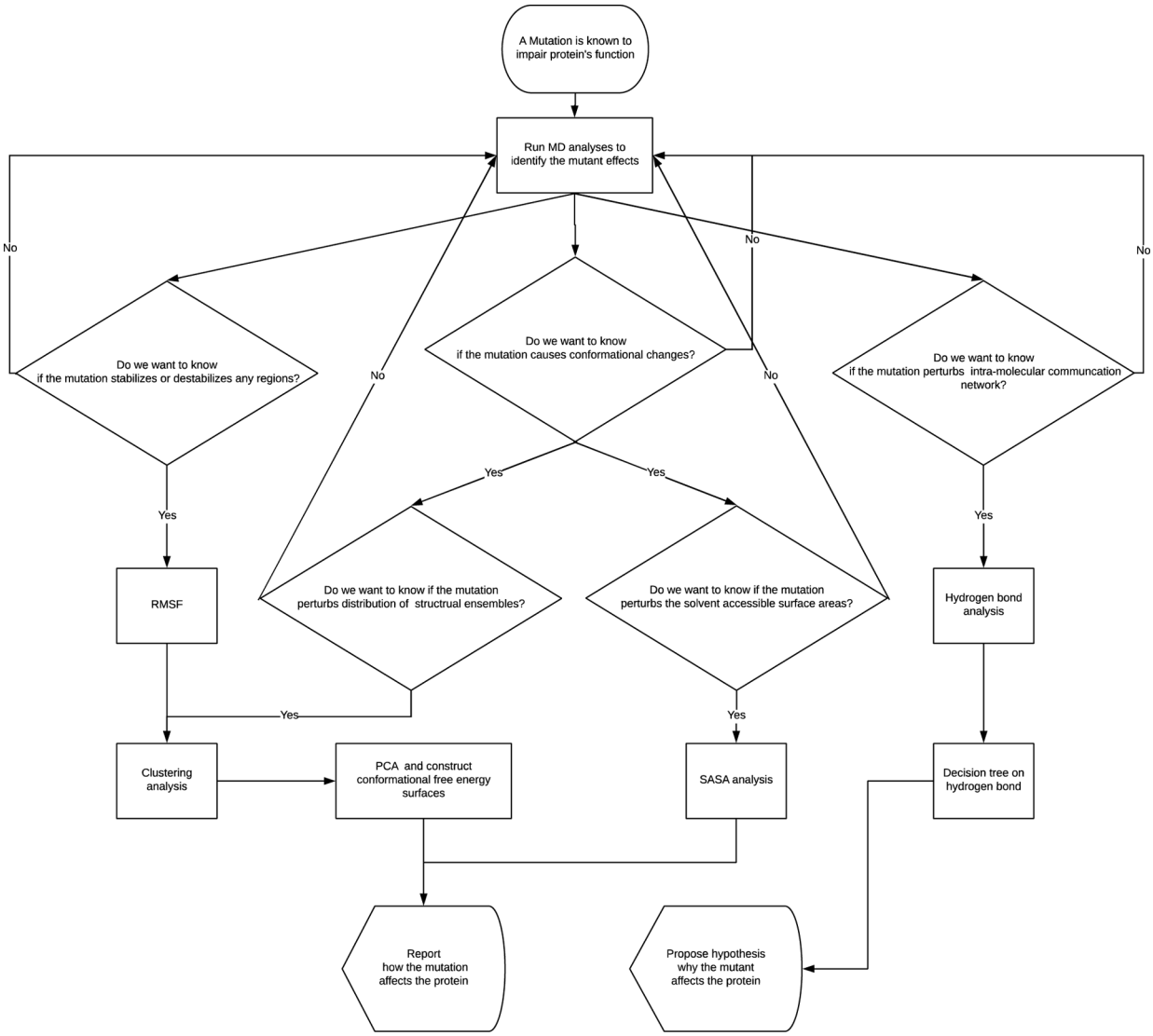

Figure 2.
A working flow chart.

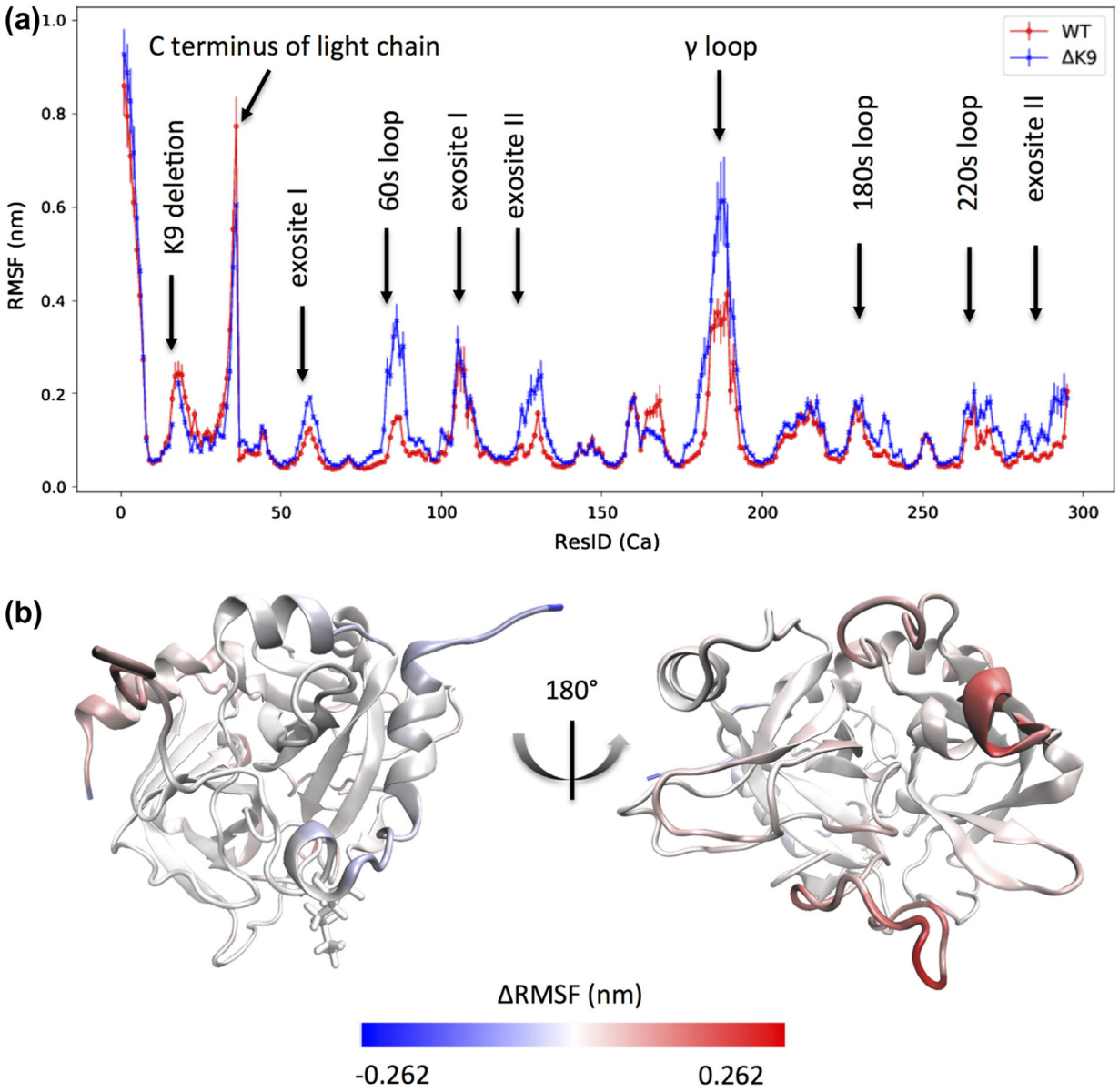

Figure 3.

Root-mean-square fluctuations (RMSF) for the alpha carbons of ΔK9 (Mutant) and wild-type (WT) thrombin. The blue and red colors in (a) depict the ΔK9 mutant and wild-type thrombin respectively. The residue indexes in (a) follow our sequential residue numbering scheme that all residues in the light and heavy chains are numbered from 1 to 295 and the LYS9 in the light chain thereby has a residue ID of 17. The known sites with distinct RMSF are indicated by labels. The thrombin molecule is colored based on the subtractions of RMSF (Mutant-WT) in (b) to indicate the location of the significantly affected regions.

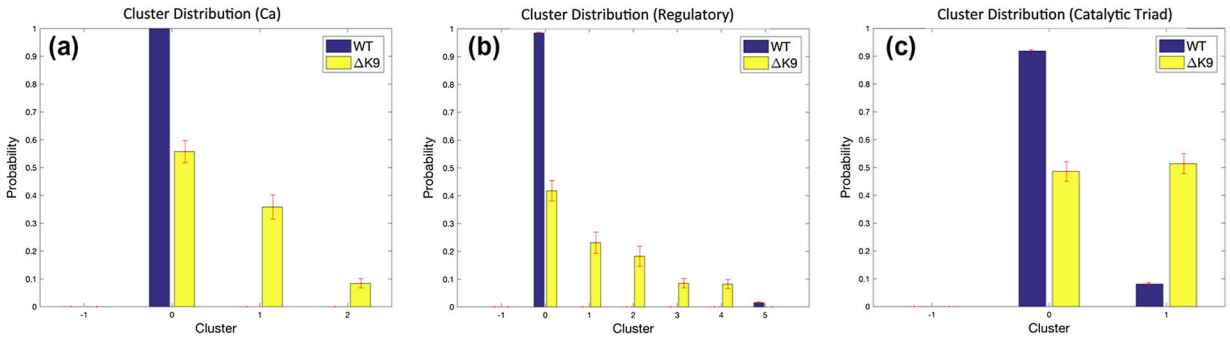

Figure 4.

Amorim-Hennig (AH) Clustering Distribution of ΔK9 (Mutant) and wild-type (WT) thrombin. Panels (a)-(c) respectively illustrate the clustering results of the Cα atoms, heavy atoms of the regulatory regions, and catalytic triad.

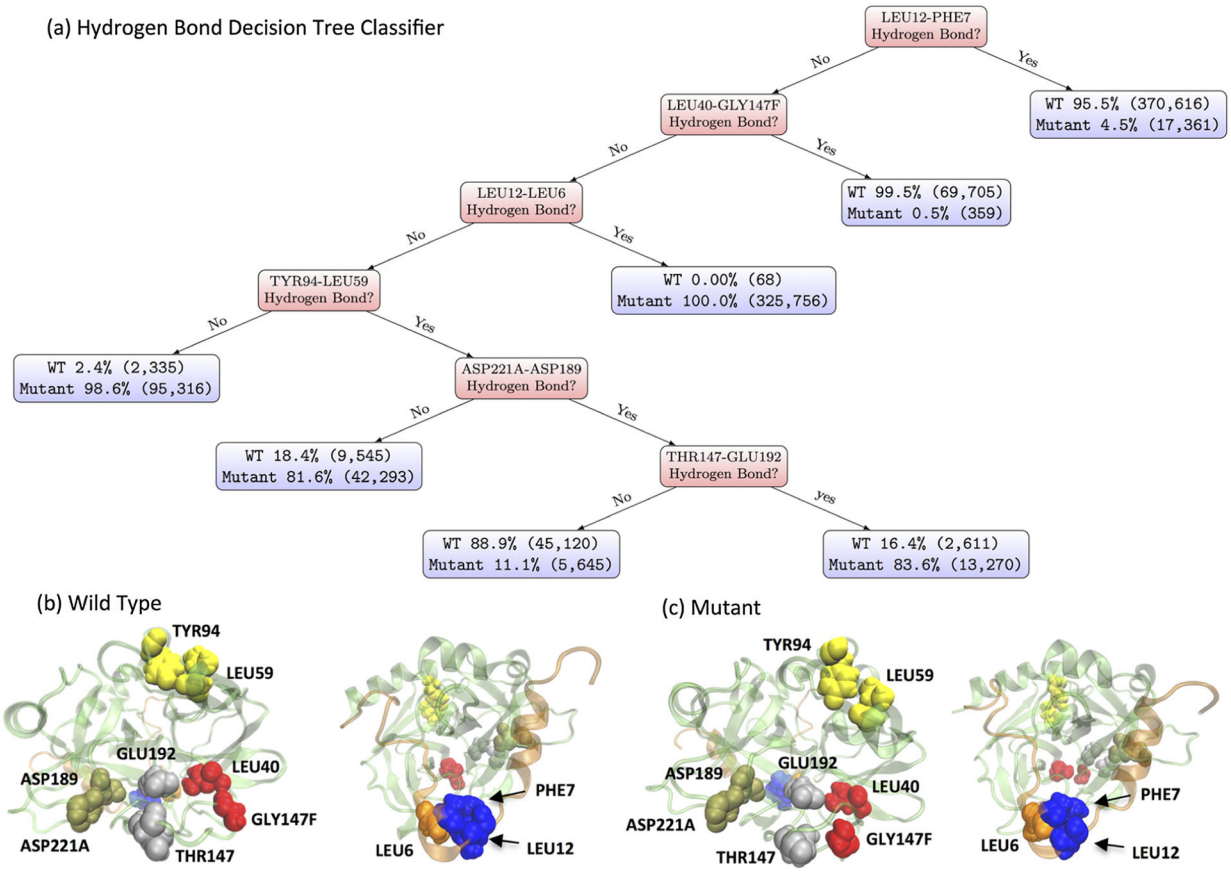

Figure 5.

Hydrogen Bond Analysis. (a) A decision tree classifier (96.21% accuracy) of ΔK9 (Mutant) and wild-type (WT) thrombin based on their residue-residue hydrogen bonding features. The percentages in the blue box denote the relative population of corresponding simulation type. The numbers within the following parentheses indicate the actual counts of structures of corresponding simulation type. Panel (b) and (c) illustrate the hydrogen bonds involved in the decision splits in (a). The wild-type and mutant thrombin structures were picked as the frames with the smallest root-mean-square distance to the average structure in each type of simulations. Each pair of residues has the same color and the order of colors (blue, red, orange, yellow, tan, silver, pink and purple) follows the decision levels.

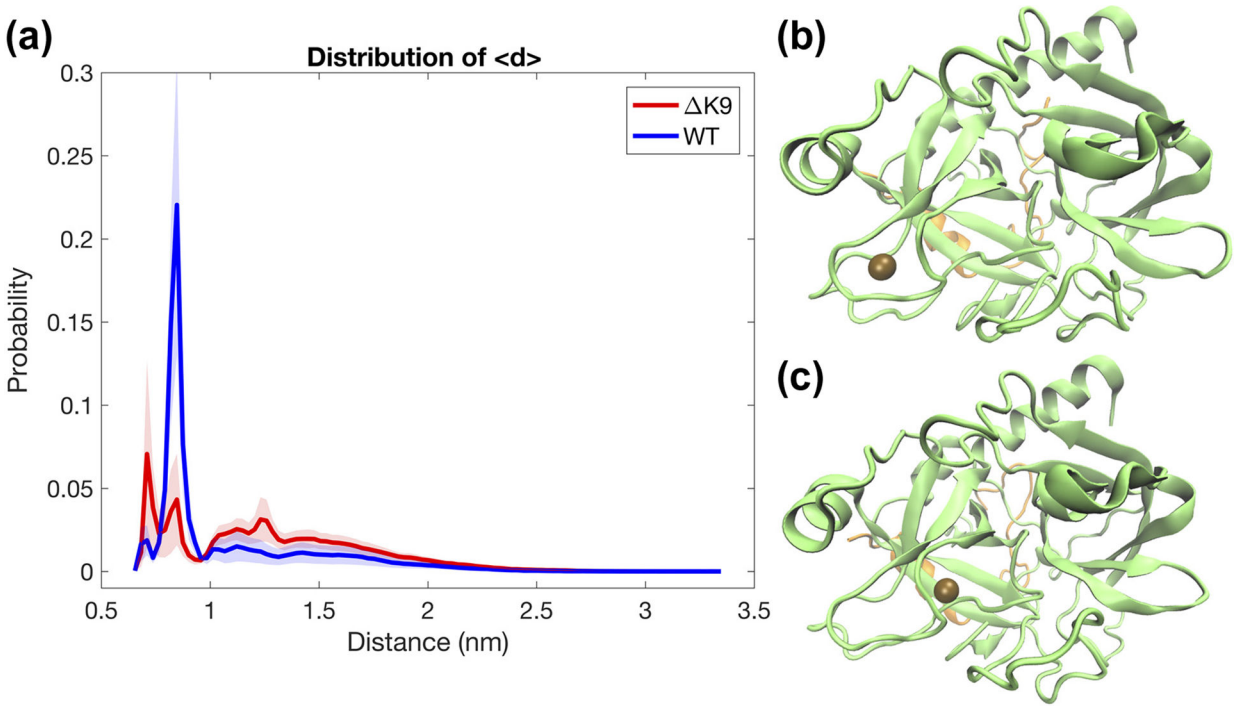

Figure 6.

Distribution of the distance between the sodium loop and the nearest Na + ion. The wild-type (WT) and ΔK9 mutant thrombin present two striking peaks representing Na + binding in the distribution plot (a). The peaks at 7 Å and 8.5 Å correspond to two binding modes of Na + as respectively illustrated in panel (b) and (c), where the bound Na + ion is shown as the brown bead.

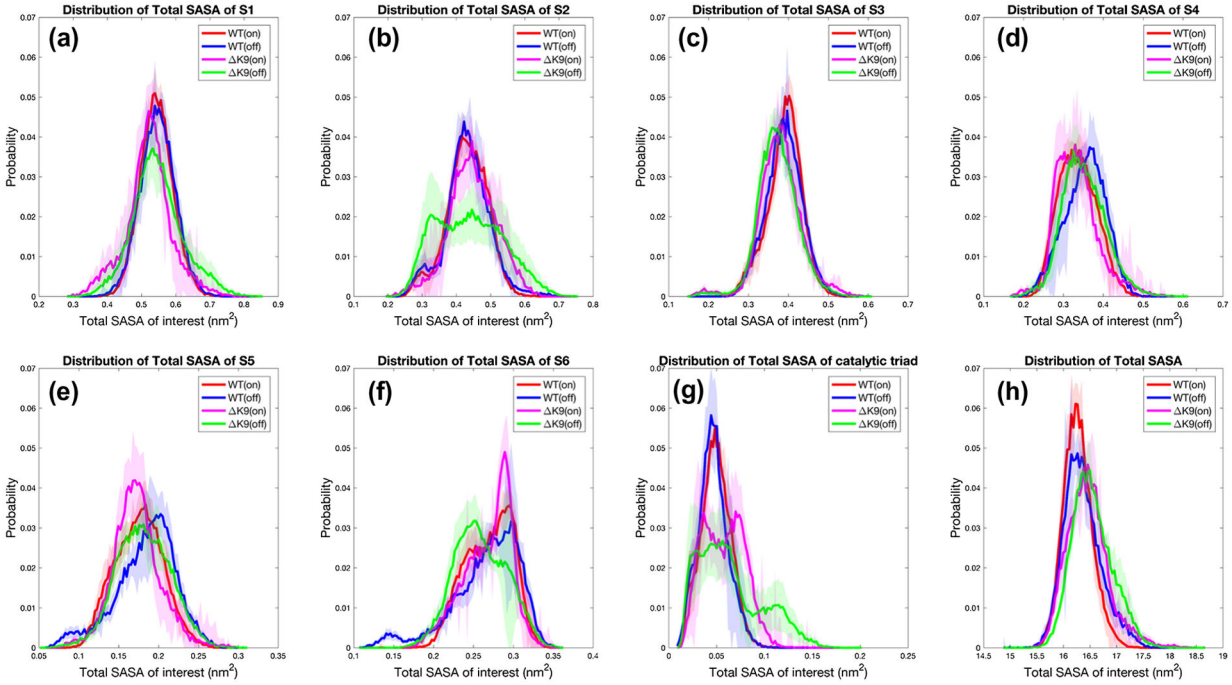

Figure 7.

Distributions of solvent accessible surface area (SASA). Regarding conformations with a bound/unbound Na {+} ion (denoted as on/off), the SASA distribution of the catalytic sub-pocket S1–6 residues are plotted respectively for wild-type (WT) and ΔK9 mutant thrombin 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).

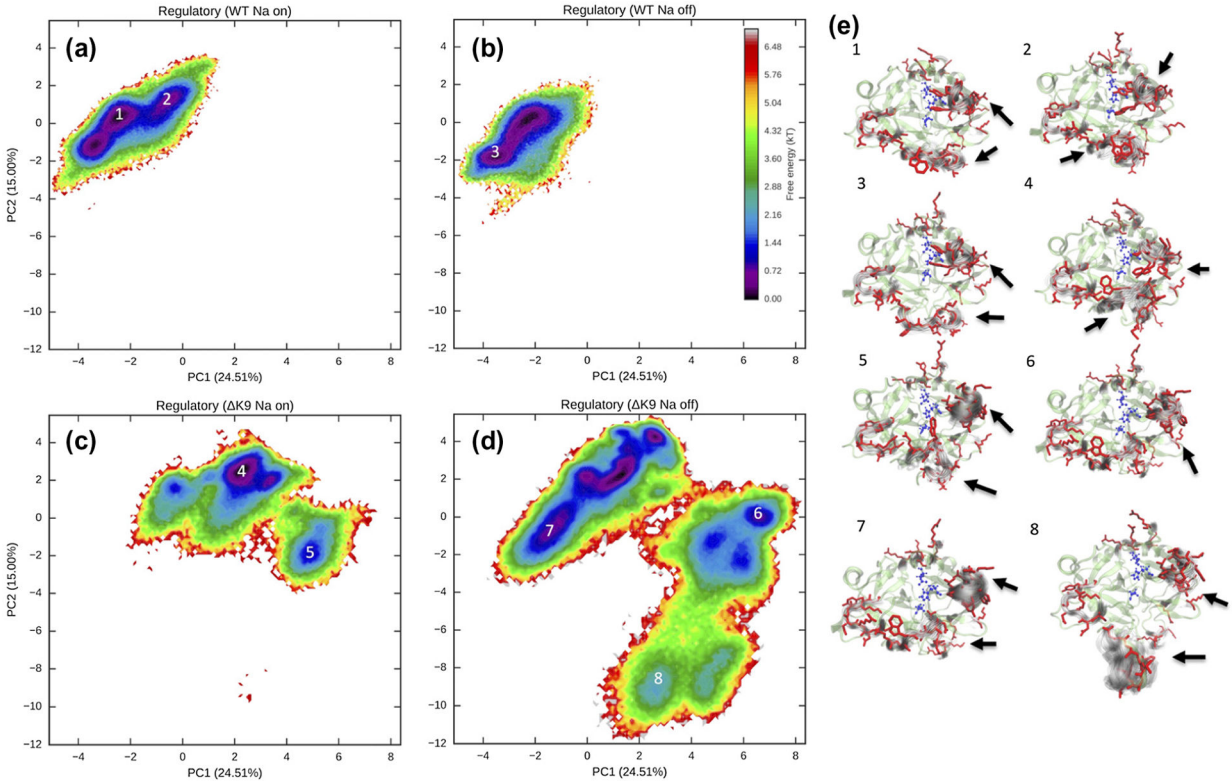

Figure 8.

Conformational free energy surfaces of the regulatory regions. Regarding conformations with a Na+ binding/unbinding, four free energy surfaces are plotted for wild-type (WT) and ΔK9 mutant thrombin respectively in panels (a), (b), (c), and (d) to compare the mutant and 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.

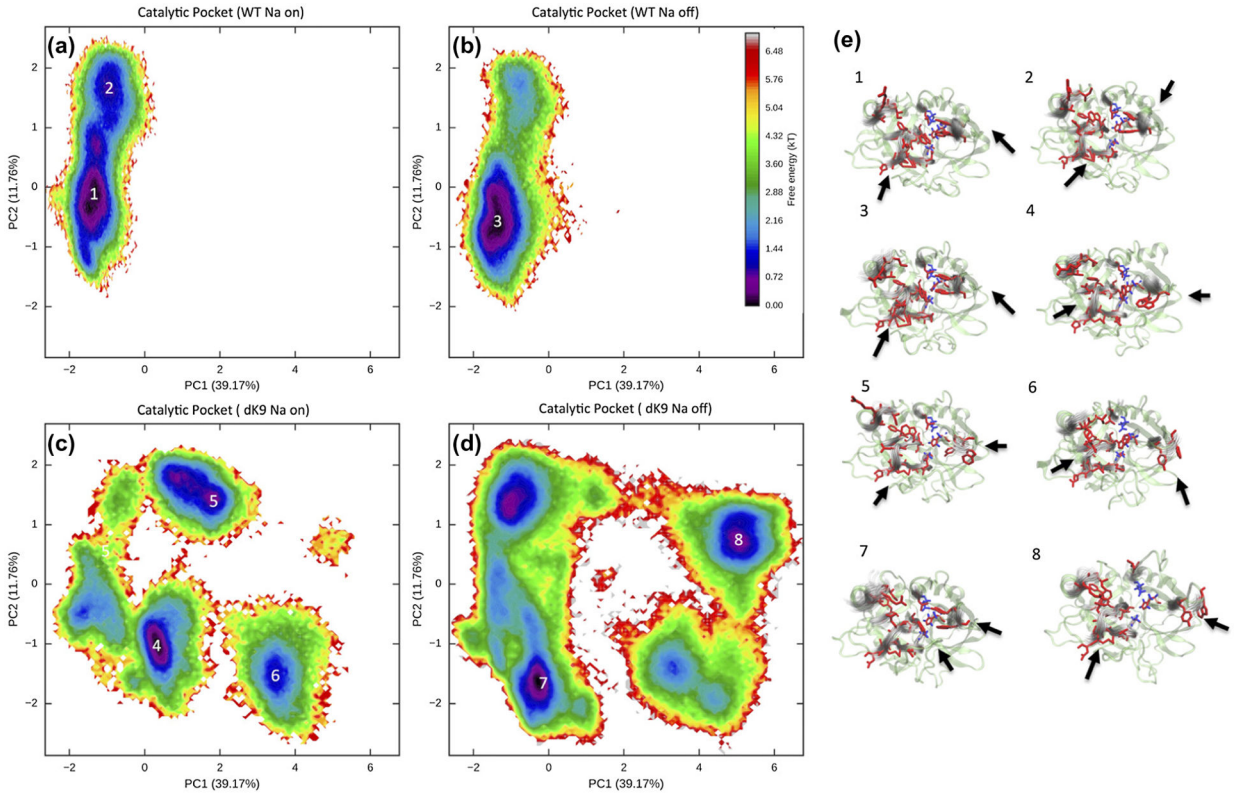

Figure 9.

Conformational free energy surfaces of the catalytic pocket. Regarding conformations with a Na+ binding/unbinding, four free energy surfaces are plotted for wild-type (WT) and ΔK9 mutant thrombin respectively in panels (a), (b), (c), and (d) to compare the mutant and 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 Figure 7. 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 residue pairs with hydrogen bonds

Conf.	WT all	WT w/i A	WT w/i B	WT b/t A&B	WT all but K9	ΔK9 all	ΔK9 w/i A	ΔK9 w/i B	ΔK9 b/t A&B
Unique	2006	203	1520	283	1991	2262	191	1811	260
Mean	184.35	14.30	158.80	13.26	183.59	181.18	13.84	154.81	12.53
Std. Dev.	7.56	2.86	6.73	2.44	7.56	7.96	2.81	7.09	1.99

*Heavy and light chains are labeled as ‘A’ and ‘B’. ‘w/i’ denotes ‘within’; ‘b/t’ denotes ‘between’.

Table 2.

Statistics of frames with Na+-binding/unbinding.

	Bound (frames/percentage)	Unbound (frames/percentage)
Wild-type	306,521 / 61.30%	193,479 / 38.70%
ΔK9	141,974 / 28.39%	358,026 / 71.61%