ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Allosteric Modulation of Thrombin by Thrombomodulin: Insights from Logistic Regression and Statistical Analysis of Molecular Dynamics Simulations

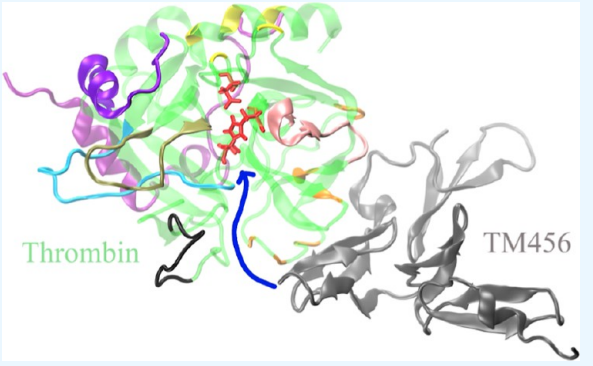
Dizhou Wu and Freddie R. Salsbury, Jr.*

Cite This: *ACS Omega* 2024, 9, 23086−23100

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Thrombomodulin (TM), a transmembrane receptor integral to the anticoagulant pathway, governs thrombin's substrate specificity via interaction with thrombin's anion-binding exosite I. Despite its established role, the precise mechanisms underlying this regulatory function are yet to be fully unraveled. In this study, we deepen the understanding of these mechanisms through eight independent 1 μs all-atom simulations, analyzing thrombin both in its free form and when bound to TM fragments TM456 and TM56. Our investigations revealed distinct and significant conformational changes in thrombin mediated by the binding of TM56 and TM456. While TM56 predominantly influences motions within exosite I, TM456 orchestrates coordinated alterations across various loop regions, thereby unveiling a multifaceted modulatory role that extends beyond that of TM56. A highlight of our study is the identification of critical hydrogen bonds that undergo transformations during TM56 and TM456 binding, shedding light on the pivotal allosteric influence exerted by TM4 on thrombin's structural dynamics. This work offers a nuanced appreciation of TM's regulatory role in blood coagulation, paving the way for innovative approaches in the development of anticoagulant therapies and expanding the horizons in oncology therapeutics through a deeper understanding of molecular interactions in the coagulation pathway.



INTRODUCTION

Thrombomodulin (TM), a transmembrane thrombin receptor comprising six contiguous epidermal growth factor-like domains (TM1−6), has the ability to modify thrombin's substrate specificity.[1,2] This modulation occurs as TM binds to the anion-binding exosite I, switching thrombin's function from a procoagulant to an anticoagulant. As a result, procoagulant substrates, including fibrinogen, PAR1, and factor V, are inhibited from binding, while protein C activation is facilitated.[1,2] The interaction of TM with thrombin significantly enhances the association rate (k_a) of various thrombin inhibitors targeting the active site,[3−7] and in the presence of TM, k_a for protein C binding witnesses a dramatic increase of over 1000-fold. However, the mechanisms contributing to this substantial increase in the association rate have yet to be comprehensively deciphered.[8−12]

This TM−thrombin interaction significantly enhances the association rate (k_a) of various thrombin inhibitors targeting the active site.[3−7] Furthermore, in the presence of TM, k_a for protein C binding witnesses a dramatic increase of over 1000-fold. However, the mechanisms contributing to this substantial increase in the association rate have yet to be comprehensively deciphered.[8−12]

Historically, the prevailing assumption was that TM binding did not trigger considerable allosteric structural alterations in thrombin's active site.[9,10,13] It was posited that TM either influenced the conformation of the bound protein C[9,13,14] or served as a scaffold promoting productive interactions between thrombin and protein C.[10]

The existing body of evidence largely aligns with the first hypothesis, asserting that TM has a negligible impact on thrombin's specificity sites S1, S2, and S3.[9] Chromogenic substrates, commonly utilized in enzymology studies to visualize enzymatic reactions, traditionally focus on these aforementioned sites.[9] Yet, their limitations become apparent when considering thrombin's specificity for protein C, which seemingly requires interaction with sites not typically targeted by the chromogenic substrate library.[9] This suggests that to enhance its specificity for protein C, thrombin must engage with these less-examined sites.[9] The residues within the

Received: April 10, 2024
Revised: May 1, 2024
Accepted: May 8, 2024
Published: May 17, 2024

sodium-binding loop, also referred to as the 220s loop, play a part in protein C cleavage,[15,16] even though they do not directly interact with small chromogenic substrates. The rise in specificity is unlikely to derive solely from structural domains situated outside the essential catalytic pocket sites that small chromogenic substrates probe.[13] Furthermore, it seems unlikely that TM would prompt a significant conformational shift in thrombin that would not be identifiable as a meaningful change in heat capacity.[13]

The second hypothesis principally relies on crystallographic observations of a TM fragment (TM456), which is critical for protein C activation, bound to thrombin; these observations revealed no significant conformational changes.[10] An inhibitor was present in the active site during this observation, potentially masking any modifications in the active site,[17] but there was no substantial evidence suggesting any TM-induced conformational modifications in thrombin beyond the active site. Additionally, TM56, the sole TM substructure interacting with thrombin in chondroitin-sulfate-free TM, did not enhance protein C activation when it bound solely to thrombin.[10] Nevertheless, when thrombin was bound to TM456, which includes an additional TM4 that does not seem to alter thrombin's structure, protein C activation was noted.[10,13] This led to the deduction that the expanded binding surface provided by TM4 is pivotal for protein C cleavage by thrombin.

Recent experimental[12,18,19] and computational research[20] have cast doubt on the validity of these previous models. These investigations challenge the former models' capacity to account for the observed 1000-fold enhancement in association rates, particularly the k_a increase for smaller substrates and inhibitors that can access the thrombin active site without directly engaging with TM4. Amide H/2H exchange experiments indicate that when TM binds to exosite I on the thrombin surface, it triggers a structural rearrangement involving the movement of various loops and helices, including the alteration of the 90s loop (residues 127−133).[18] TM binding opens up this loop, lessening the electrostatic repulsion between ASP133 and ASP135 in thrombin and the P3 ASP of protein C, thereby improving protein C's access to the active site of thrombin.[18] HDXMS studies demonstrate that the binding of TM456 to thrombin diminishes amide exchange across the thrombin molecule.[12] NMR investigations reveal dynamic modifications induced by TM456 binding, leading to the ordering of certain protein regions and disordering of others.[19] Computational studies, employing community network analysis, suggest a potential connection between TM4 and the thrombin active site via exosite I and TM5, due to TM4's notable association with TM5, which directly interacts with exosite I.[20]

To gain deeper insights into how TM enhances protein C cleavage, we performed 24 separate 1 μs all-atom simulations involving free thrombin, TM456-bound thrombin, and TM56-bound thrombin, eight for each system. For these simulations, we employed the CHARMM36 force field[21] in combination with GPU-enabled molecular dynamics. The use of GPU-enabled simulations conferred a significant advantage, enabling us to extend each simulation to 50 times the duration of prior computational studies on TM456.[20] We have used such simulations before to study the dynamic allostery of thrombin due to ion binding, mutation, and aptamer binding as well as to study dynamic allostery in other biomedically relevant protein complexes.[22−38]

In our study, a series of machine learning and statistical methods were used to analyze the data obtained from the simulations. These analyses were instrumental in supporting the idea that TM binding induces allosteric structural changes in both the active site and functional regions of thrombin,[12,18−20] thereby challenging the previous assumptions.[9,10,13,14] This approach not only elucidated the structural changes in thrombin as a result of TM binding but also provided novel insights into the allosteric pathways of TM binding.

Our clustering analysis identified significant conformational changes in thrombin induced by the binding of TM56 and TM456, with the latter inducing unique conformations. This differential impact of TM456 is likely responsible for the observed 1000-fold increase in the association rate constant (k_a) for protein C, an effect not replicated by TM56. Furthermore, our analysis of correlated motions revealed that TM56 binding predominantly influenced motions within exosite I, its binding site. In contrast, TM456 binding was found to significantly affect correlated motions between several functional regions, including the 170s loop, 220s loop, 180s loop, 60s loop, and exosite II, suggesting a more extensive allosteric influence. These observations provide a rationale for the limited ability of TM56 to stimulate a similar increase in k_a for protein C as compared to TM456.

The use of logistic regression, a supervised machine learning method, in our hydrogen-bond analysis—a technique we have previously developed and used elsewhere[24]—allowed us to identify critical hydrogen bonds that are either disrupted or formed upon the binding of TM56 and TM456. This analysis identified key hydrogen bonds, such as LEU132−ARG216 and ILE114−ASN380, offering insights into the allosteric pathway of TM4. The application of advanced machine learning techniques in this study not only deepens our understanding of the role of TM in blood coagulation but also paves the way for the development of innovative anticoagulant therapies. In addition, these findings may have significant implications for future therapeutic strategies in oncology.

MATERIALS AND METHODS

Simulation System. In order to scrutinize the distinct impacts of TM456 and TM56 binding on thrombin, we conducted simulations encompassing three different conditions: thrombin bound with TM456, thrombin bound with TM56, and thrombin in its unbound state. These simulations aimed to replicate the conformational ensemble of thrombin when subjected to the influence of TM456 and TM56 binding and to elucidate the innate thermodynamic properties of unbound thrombin. Through a comparative analysis of the simulations of TM456-bound, TM56-bound, and free thrombin, our objective is to pinpoint key conformational changes that could be instrumental to the enhanced association rate.

In this study, we utilized high-resolution crystal structures of the TM456−thrombin complex, which were obtained from the RCSB PDB (ID: 1DX5), as resolved by Fuentes-Prior et al. in 2000. It is noteworthy that the original PDB file contained several mutations such as ARG456 and HIS457 mutated to GLY456 and GLN457, respectively, for the purpose of preventing proteolysis,[10] and ASN364 was modified to ASP364 following PNGase treatment.[10] Additionally, THR91 was found to be mutated to ILE91. For our simulations to better reflect physiological conditions, these residues were

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

reverted to their wild-type state. From the original file, only the light chain, heavy chain, TM456, and the calcium ion were retained, discarding all other ligands and water molecules. Missing hydrogen atoms were appended using the psfgen package within VMD,[39] applying default parameters.

Simulation Conformations. In order to sample a wide range of conformations, we performed eight separate 1 μs all-atom simulations for each of the three systems: TM456-bound thrombin, TM56-bound thrombin, and free thrombin. These simulations were executed using the GPU-accelerated ACEMD3 simulation package[40] on NVIDIA Titan GPUs housed within Metrocubo workstations. The simulation protocols described below are standard with ACEMD and have been successfully used in previous works on thrombin and other systems.[22−29,32−38]

We conducted the simulations in an explicitly aqueous environment, constructed via a TIP3P cubic water box[41] with a minimum buffer of 10 Å in all directions. Cl^- ions were added to neutralize the systems, and then the concentration of Na^+ and Cl^- was increased to 0.125 M,[42] adhering to protocols consistent with our previous research.[22−28] The ionizable residues retained their default protonation states at a physiological pH of 7. To mitigate any potential contact-related issues, all systems were subject to 1000 cycles of conjugate gradient minimization before running the simulations.

We conducted the simulations utilizing the CHARMM36 force field.[21] The system's pressure was maintained at 1 atm using the Berendsen pressure control method.[43] A constant temperature of 300 K was ensured through the Langevin thermostat approach,[44] applying a damping coefficient of 0.1. van der Waals and electrostatic forces were applied within a cutoff distance of 9 Å, with a switching distance set at 7.5 Å. The computation of long-range electrostatics was performed using the smooth-particle-mesh-Ewald (PME) method.[45,46]

To ensure efficient integration, we employed a simulation time step of 4 fs and incorporated hydrogen mass repartitioning.[47] The SHAKE algorithm,[48] a widely accepted approach for constraining bond lengths, facilitated the use of this larger integration step. System conformations were recorded every 2500 steps, corresponding to an interval of 10 ps, yielding a total of 100,000 frames per 1 μs simulation run.

We ran eight 1 μs simulations for each system, collectively amassing 800,000 frames per system. All simulations demonstrated adequate convergence, as evidenced by RMSD (Figure S1). The thrombin trajectories from all twenty-four 1 μs simulations were concatenated in the order of free thrombin, TM56-bound thrombin, and TM456-bound thrombin, forming a single 24,000-frame trajectory representing 24 μs. Rigid body rotations and translations were applied via a Python script to align all frames in the concatenated trajectory to the thrombin in the initial frame, facilitating a focus on the internal motions of thrombin. All subsequent analyses were conducted on this concatenated trajectory.

Root-Mean-Square Fluctuations. To quantify the atomic flexibility across each simulation, we calculated the root-mean-square fluctuations (RMSFs) for each residue. RMSFs measure the average deviation of each atom from its mean position in the protein structure over time, thus offering insights into the flexibility and mobility of distinct protein regions. Unless specified otherwise, all RMSFs in this study were computed for the alpha-carbons. The calculation of RMSF follows the equation

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_i(t) - \overline{r_i})^2} \quad (1)$$

where $r_i(t)$ is the position vector of the atom i at frame t, and $\overline{r_i}$ is the average position for atom i over all T frames.

Clustering Analysis. To elucidate the conformational dynamics of thrombin under the influence of TM456 and TM56, we utilized clustering analysis, a critical tool for studying molecular dynamics trajectories. Clustering analysis facilitates the categorization of molecular conformations based on structural resemblance, typically measured by calculating the root-mean-square distance (RMSD) between distinct conformations after alignment to negate the center-of-mass motion and the overall molecular rotation. This classification of conformations can provide insights into the potential conformations a molecular system can adopt.[30,49]

In our study, we employed two distinctive clustering techniques: hierarchical density-based spatial clustering of applications with noise (HDBSCAN)[50] and the Amorim−Hennig (AH) method.[51]

HDBSCAN is a nonparametric clustering technique that offers valuable insights into the overall structure and stability of disordered systems by identifying unique conformational ensembles. Unlike traditional density-based clustering approaches, HDBSCAN does not necessitate a predefined number of clusters, making it well-suited for analyses where the number of clusters is not known beforehand.[30]

Contrastingly, the AH method is particularly adept at analyzing structured, stable systems such as folded proteins. It captures alterations at a high-resolution level, including changes in relative positions of secondary structural elements, and can identify both local and global conformational changes accurately.[30] Together, HDBSCAN and AH can assess system stability and identify significant conformational shifts without requiring prior system-specific knowledge.[22,23,25,28,30,33,34]

In this analysis, we utilized HDBSCAN for the 60s loop, the 220s loop, and the catalytic triad, and the AH method for the 170s loop and the γ loop, acknowledging the importance of the latter from previous work. This approach leverages the unique strengths of both methods. To enhance computational efficiency, we focused on the heavy atoms (excluding hydrogen atoms) of these regions. The clustering analysis was performed on the concatenated trajectory at the Wake Forest University High-Performance Computing Facility,[52] sampling conformations every 1 ns to yield a total of 24,000 conformations, with 8000 conformations for each system under study.

For visualization, thrombin structures from each cluster were rendered using Tachyon in VMD 1.9.3.[39] Representative structures were depicted via the NewCartoon representation in transparent green, with other conformations within the same structural ensembles shown as shadows, thus illustrating the range of the conformational ensemble.[31] Each representative structure corresponds to the median structure of its cluster, which is the structure nearest to the average structure within the cluster. Shadows adhered to our established visualization rules,[31] wherein only conformers within 1 standard deviation in each cluster were displayed to represent variance.

Principal Component Analysis. Principal component analysis (PCA) is a vital tool for identifying the primary conformational changes in biological systems. This technique

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

reduces the dimensionality of data sets, enabling significant variations to be represented with fewer features, which simplifies interpretation.[53] A comprehensive overview of this method as applied to biological systems is available in our recent review.[49]

The fundamental steps for PCA are as follows:

1 Matrix construction: Construct a matrix A with T rows and 3N columns, where T denotes the number of frames in the simulation, and N represents the number of atoms in the system. Each row corresponds to the concatenated coordinates for each frame in the form (X_1, Y_1, Z_1, X_2, Y_2, Z_2, ..., X_N, Y_N, Z_N).

2 Covariance computation: Compute the covariance matrix C of matrix A using the formula:

$$\mathrm{cov}(i, j) = \sum_{t=1}^{T} \frac{(q_i(t) - \overline{q})(q_j(t) - \overline{q})}{T} \tag{2}$$

where q represents X, Y, or Z, and the range of i and j extends from 1 to 3N.

3 Covariance diagonalization: Diagonalize the covariance matrix C. The resulting eigenvectors V form the new feature basis, while the projections $Y = AV$ represent the new coordinates under this basis. The corresponding eigenvalues denote the significance of these new feature components. Larger eigenvalues imply greater data variation along the corresponding component. This step allows a significant portion of the original data's fluctuations to be represented by just a few components with the largest eigenvalues.

In this study, we implemented PCA using the "pca" module in the PYEmma[54] Python package, focusing on the regions of interest: exosite I, the 180s loop, and the γ loop for the heavy atoms. We extracted the subcovariance matrices corresponding to these regions. Principal components were then computed based on these subcovariance matrices, which allowed in-depth exploration of the dynamic behavior of each individual region.

Following this, we constructed conformational landscapes for these regions. We calculated the free energy ΔG_i via

$$\Delta G_i = -kT\ln\left(\frac{P_i}{P_0}\right) \tag{3}$$

where k denotes the Boltzmann constant, T represents the temperature, P_i is the occupancy in the ith bin, and P_0 is the maximum occupancy among all bins. In line with our prior research,[22,23,26,28] we used 50 bins in each dimension.

Correlation Matrix. A correlation matrix was used to evaluate the correlation of motion between pairs of residues throughout the simulations.[55] This matrix indicates how pairs of residues move in relation to each other, a critical aspect in understanding allosteric communication pathways in proteins and protein complexes.[22,23,26,31−34,36−38,55]

$$\mathrm{cov}(i, j) = \sum_{t=1}^{T} \frac{(r_i(t) - \overline{r_i}(t))\cdot(r_j(t) - \overline{r_j}(t))}{T} \tag{4}$$

$$\mathrm{corr}(i, j) = \frac{\mathrm{cov}(i, j)}{\sqrt{\mathrm{cov}(i, i)\cdot\mathrm{cov}(j, j)}} \tag{5}$$

where $r_i(t)$ is the position vector of the atom i at frame t, $\overline{r}(t)$ is the average atom position over all T frames. These equations were implemented in in-house Python scripts, available via GitHub https://github.com/salsburygroup/Thrombin_Thromodulin_LR.

The components of the correlation matrix range from −1 to +1. A correlation coefficient near 1 in absolute value signifies a strong coupling between a pair of atoms, suggesting either correlated or anticorrelated movements. In contrast, an uncorrelated pair of atoms will show a correlation coefficient near zero.

Unless otherwise specified, all correlation matrices were calculated for alpha-carbons.

Hydrogen-Bond Analysis. Hydrogen bonds are of paramount importance in maintaining the secondary and three-dimensional (3D) structure of proteins, offering strong and directional interactions.[56−58] They contribute to the stabilization of protein structures, forming between the backbone atoms of amino acids, and between amino acid side chains and other molecules.[56−58] Accordingly, changes in protein structures often result from the formation or disruption of hydrogen bonds.[56−58]

In this study, the Python programming language[59] and the MDAnalysis package[60,61] were utilized to analyze hydrogen bonding between polar atoms within the concatenated trajectory. The criteria to identify a hydrogen bond were a maximum heavy atom distance of 3.2 Å and a maximum angle of 120° from a heavy atom to a hydrogen atom and back to another heavy atom.[62] These parameters are as used in our prior research.[24]

To optimize the computational efficiency of our analysis, we implemented a targeted filtration of hydrogen bonds based on their frequency of occurrence within the concatenated trajectory frames. Our strategy was to exclude hydrogen bonds that were either too sporadic or too persistent across all systems, as such bonds are unlikely to be crucial for distinguishing between different systems. To this end, we set a conservative and inclusive threshold by removing hydrogen bonds present in less than 2.5% or more than 97.5% of the total frames. This selective process resulted in the exclusion of about 80% of the hydrogen bonds, which were considered unimportant. As a result, the computational load was reduced to no more than one-fifth of the original requirement. This reduction in computational demand is consistent with the O(n) complexity of the logistic regression model,[63,64] thereby validating our approach and corroborating the methods used in our previous research.[24]

After filtering, 312 hydrogen bonds served as explanatory variables in a two-dimensional matrix. Each column represented a hydrogen bond, each row represented a frame of the concatenated trajectory, and each entry was either 1 or 0, indicating the presence or absence of the hydrogen bond within that frame. The response variable indicated whether thrombin was bound to TM56 or TM456, labeled "thrombin" for single thrombin, "TM56" for the thrombin−TM56 complex, and "TM456" for the thrombin−TM456 complex. This matrix constituted the data set for our study.

In this study, we employed a logistic regression model for the analysis of hydrogen bonds, consistent with our previous methodology.[24] Logistic regression is a statistical model that uses a logistic function to model a binary response variable's relationship with one or more explanatory variables.[65,66] The logistic function, an S-shaped curve, maps any real-valued input to a value between 0 and 1, interpretable as the binary outcome's probability.

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

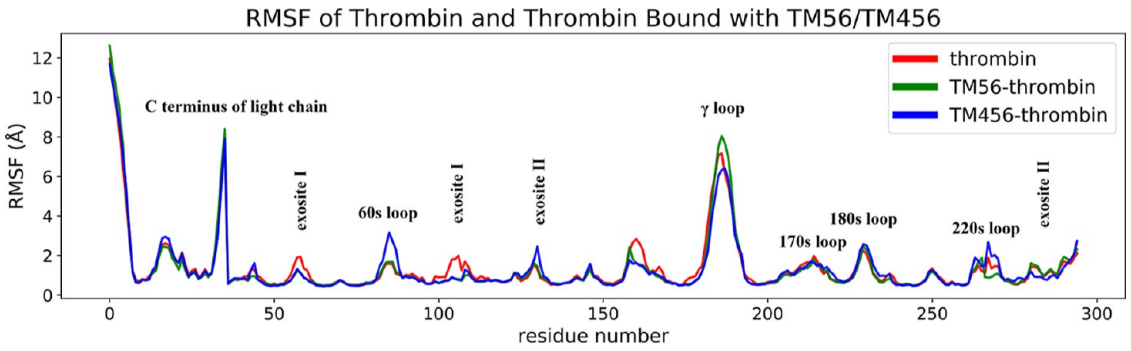


Figure 1. RMSFs for the α-carbons of thrombin, TM56-bound thrombin, and TM456-bound thrombin.

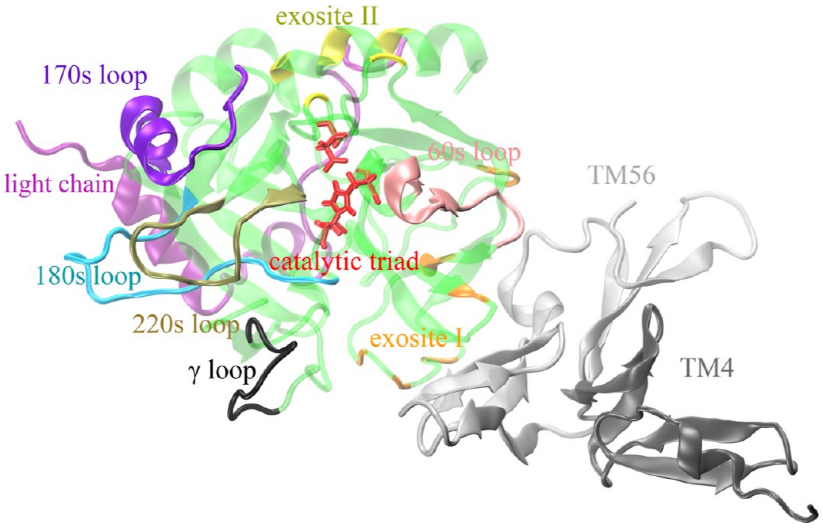


Figure 2. Depiction of TM456-bound thrombin structure. The heavy chain is visualized in transparent green, while the light chain is presented in purple. TM456 is rendered in gray, with TM56 and TM4 separately marked. Key functional sites include the 60s loop (pink), 170s loop (violet), 180s loop (cyan), 220s loop (tan), and the γ-loop (black), along with exosite I (orange) and II (yellow) that are distinctly colored. The side chains of the catalytic triad are highlighted in red.

The logistic regression model is mathematically expressed as

$$p = \frac{1}{1 + e^{-z}} \tag{6}$$

where p is the predicted probability of the binary outcome, z is a linear combination of the explanatory variables and their coefficients, and e is Euler's number.

The linear combination of independent variables and their coefficients can be expressed as

$$z = b_0 + \sum_{i=1}^{n} (b_i x_i) \tag{7}$$

where b_0 is the intercept term, and b_i are the coefficients for each explanatory variable x_i, which represents whether a hydrogen bond is present in our study.

The objective of logistic regression is to estimate the coefficient values that maximize the likelihood of observing the data given these coefficients. This process involves minimizing the elastic net penalized negative log-likelihood, which measures the alignment of the predicted probabilities with the actual binary outcomes in our data.

Once trained on our data, the coefficient values can be employed to represent the significance of each hydrogen bond. The larger the absolute value, the greater the importance of the corresponding hydrogen bond.

The main advantage of using a logistic regression model to analyze hydrogen-bond networks, compared to simpler methods such as calculating the fraction of H bonds in different systems, lies in the model's ability to account for the intricate interdependencies among hydrogen bonds. In proteins, correlations between hydrogen bonds are pervasive and cannot be ignored. This complexity is often overlooked by basic analytical approaches. By using logistic regression, we gain a more sophisticated insight into the variances within hydrogen-bond networks in different systems, allowing for a deeper understanding of their structural and functional implications. For a practical illustration of how the logistic regression model accounts for the correlations between hydrogen bonds, readers are encouraged to examine the example_Hbond.Rmd file. This file is accessible in our GitHub repository at https://github.com/salsburygroup/Salsbury_group_codes/tree/main/JCIM_2023_DW.

The logistic regression model is effective for binary outcomes, where the responses are limited to 0 and 1. This makes it ideal for comparing hydrogen-bond networks between two different systems. However, for the analysis of hydrogen-bonding networks across multiple systems, a more appropriate approach is the multinomial logistic regression model. This model extends the basic logistic regression framework to handle multiple categories. For a detailed discussion of this

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

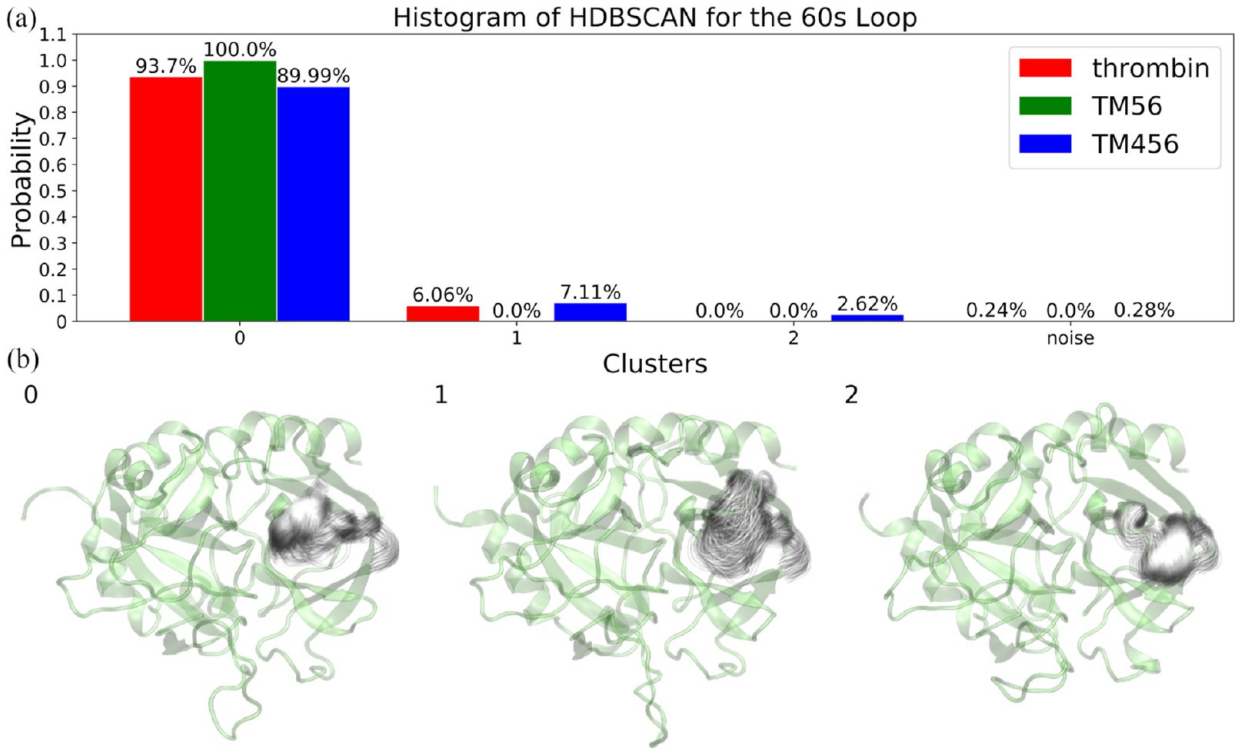


Figure 3. HDBSCAN clustering for the heavy atoms of the 60s loop in thrombin, TM56-bound thrombin, and TM456-bound thrombin. (a) Distribution of clusters. (b) Representative structures for each cluster. "Noise" designates the data points that are not assigned to any specific cluster, owing to their low density. The median structure of thrombin is represented in transparent green, while the variance of the 60s loop in each cluster is depicted by shadows, in accordance with the visual statistics method.[31]

method and its applications in our research, we direct the readers to our recent publication.[24]

RESULTS AND DISCUSSION

RMSF Analysis Revealing TM56 and TM456 Induced Atomic Flexibility Changes in Thrombin. Figure 1 shows the RMSFs, highlighting the atomic flexibility changes in thrombin as a result of binding with TM56 and TM456. Additionally, RMSFs of the α-carbons in free thrombin, as well as in TM56-bound and TM456-bound thrombin across all eight runs, are detailed in Figure S2 for comprehensive comparison. Notably, the binding of TM456 significantly enhances the flexibility of the 60s loop, which spans residues 82−94. Additionally, RMSF analysis reveals that exosite I (residues 57, 98, 104, 106, 109, 142, and 143) experiences marked stabilization upon the binding of both TM56 and TM456. Intriguingly, the 220s loop (residues 262−274) also exhibits flexibility changes when TM56 and TM456 are bound. Figure 2 presents a depiction of the TM456-bound thrombin structure with all functional regions labeled.

In order to illuminate potential conformational alterations in the functional domains of thrombin, we have performed clustering, correlation, hydrogen-bond, and principle component analyses. Figures 3−5 show the HDBSCAN clustering, while Figure 6 presents the results from the AH clustering. PCA, which is integral for understanding the primary variations in our data, is depicted in Figures 7−9. In Figure 10, we present the correlation analysis, and Figure 11 provides the analysis of hydrogen bonds. Taken together, these diverse analyses offer insights into the influence of TM56 and TM456 binding on the conformational dynamics of thrombin.

Deciphering Allosteric Conformational Changes in Thrombin through HDBSCAN and AH Clustering: Differential Impact of TM56 and TM456 Binding. HDBSCAN clustering reveals distinct conformational shifts in the 60s loop upon binding with TM56 and TM456 (Figure 3). In the case of TM56, all observed conformations belong to cluster 0, suggesting minimal conformational changes (Figure 3a). In contrast, thrombin alone exhibits some conformations featuring a pronounced 180 degree rightward twist in the 60s loop (cluster 1, Figure 3b). Intriguingly, the binding of TM456 imparts a dual effect; while some conformations mimic the rightward twist seen in thrombin alone, others display an additional downward deformation of the 60s loop (cluster 2, Figure 3b). These observations suggest that TM456 binding potentially encourages a broader opening of the 60s loop, in contrast to the constrained openness observed with TM56 binding. This difference could contribute to the substantial (1000-fold) increase in the association constant (k_a) for protein C binding observed when TM456 is present, compared to when TM56 is bound.[8−12]

Likewise, the binding of TM456 significantly modifies the conformation of the 220s loop (Figure 4). Specifically, residues 263−265 exhibit a strand collapse in certain conformations (cluster 1, Figure 4b), an observation that aligns with previous experimental results.[17] This conformational adjustment indicates an allosteric reshaping of the thrombin structure toward its anticoagulant form when TM456 is bound.

Substantial conformational shifts are also apparent in the catalytic triad following the binding of TM456 (Figure 5). Following TM456 binding, around 22.3% of the conformations coalesce into cluster 1, a marked increase from 3.2 and 2.7% observed with wild-type thrombin and TM56, respectively

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

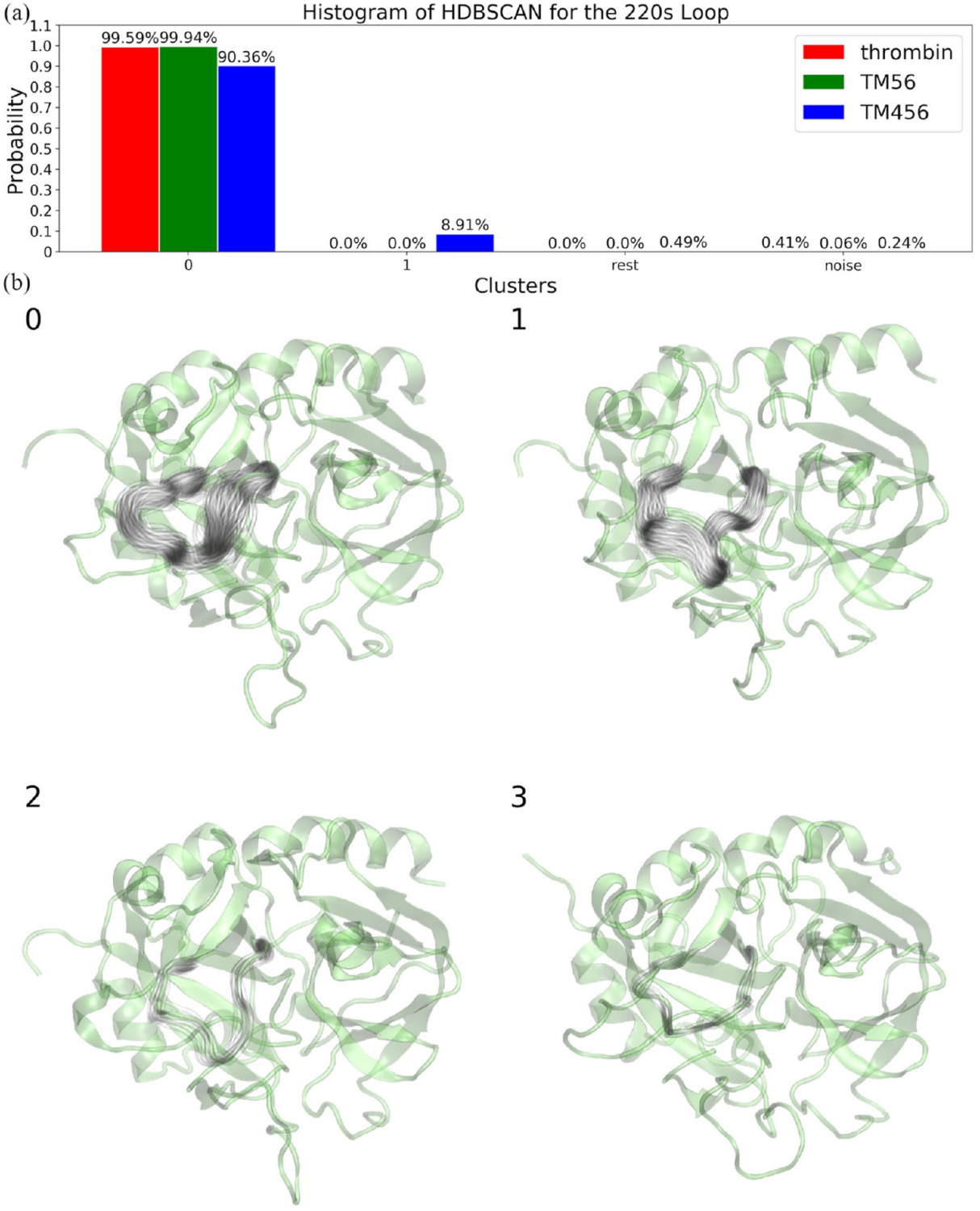


Figure 4. HDBSCAN clustering for the heavy atoms of the 220s loop in thrombin, TM56-bound thrombin, and TM456-bound thrombin. (a) Distribution of clusters. (b) Representative structures for each cluster. The term "rest" refers to the remainder of the small clusters, containing conformations that belong to thrombin, TM56-bound thrombin, or TM456-bound thrombin, each constituting less than 1% of the total. "Noise" designates the data points that are not assigned to any specific cluster, owing to their low density. The median structure of thrombin is represented via transparent green, while the variance of the 220s loop in each cluster is represented by shadows according to the visual statistics.[31]

(Figure 5a). Moreover, TM456 binding induces a unique conformation (cluster 2, Figure 5b), distinguished by significant positional deviations of residues 79 and 241 within the catalytic triad, indicating allosteric changes.

Lastly, the AH clustering illustrates a unique conformational shift within the 170s loop following TM456 binding. This interaction straightens the previously curved loop, forming a new conformation (cluster 1, Figure 6b).

Principal Component Analysis of Thrombin's Exosite I, 180s Loop, and γ Loop: Conformational Responses to TM56 and TM456 Binding. PCA effectively facilitates the construction of free-energy-like surfaces for exosite I, the 180s

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

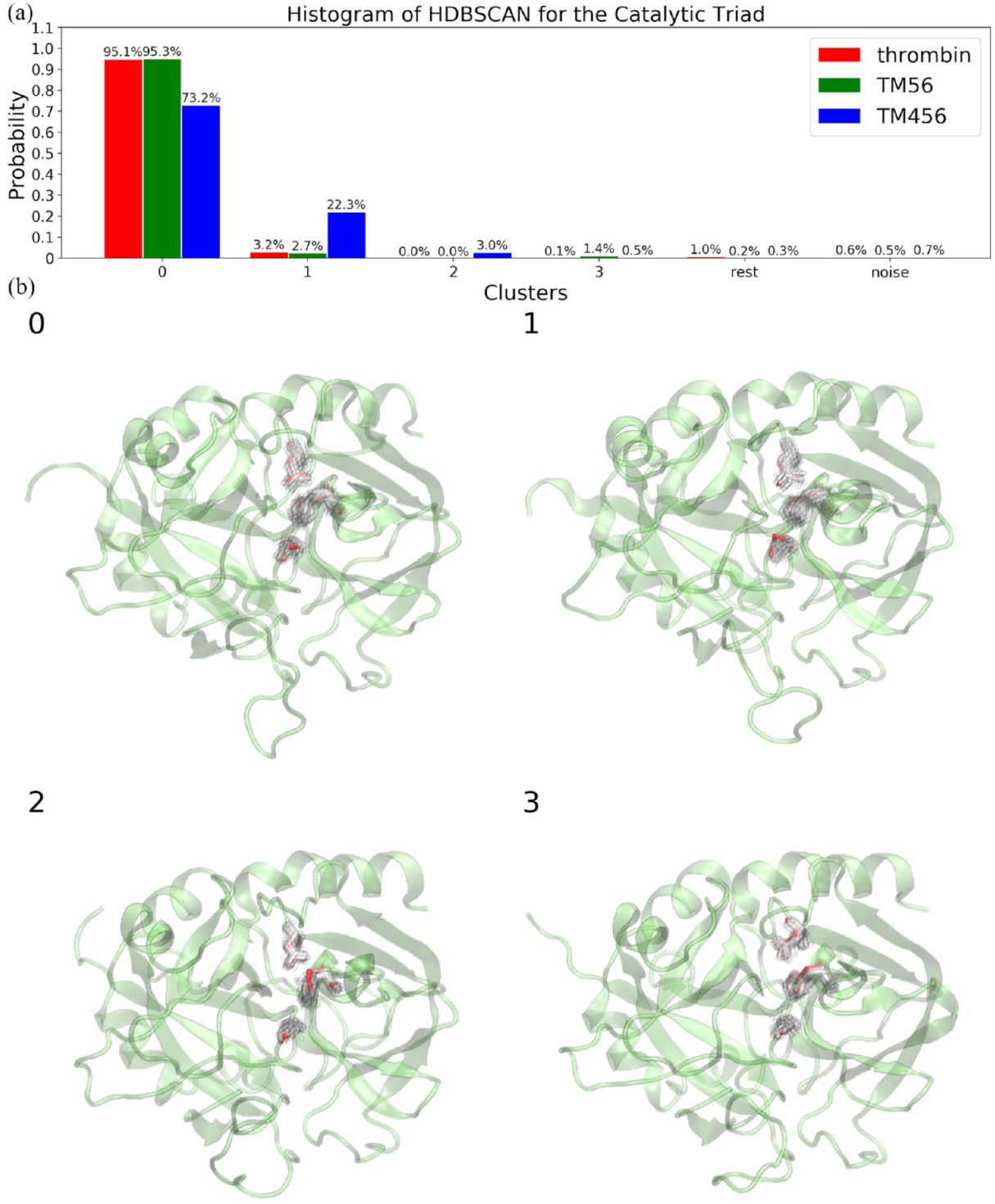


Figure 5. HDBSCAN clustering for the heavy atoms of the catalytic triad in thrombin, TM56-bound thrombin, and TM456-bound thrombin. (a) Distribution of clusters. (b) Representative structures for each cluster. The term "rest" refers to the remainder of the small clusters, containing conformations that belong to thrombin, TM56-bound thrombin, or TM456-bound thrombin, each constituting less than 1% of the total. "Noise" designates the data points that are not assigned to any specific cluster, owing to their low density. The median structure of thrombin is represented via transparent green, while the variance of the catalytic triad in each cluster is represented by shadows according to the visual statistics.[31]

loop, and the γ loop, allowing for detailed observation of conformational changes in each region. Specifically, PCA reveals a smaller free-energy well in exosite I following the binding of TM56 and TM456, suggesting the stabilization of this site (Figure 7). This is anticipated as exosite I is the binding locus for both TM56 and TM456. In contrast, unbound thrombin displays an additional free-energy well, potentially indicative of a favored conformation for the attachment of other biomolecules. Visual representations of the exosite I conformations in the two wells are provided in Figures S3a and S4.

Additionally, the binding of TM56 induces a new free-energy well for the 180s loop, while the attachment of TM456 yields another unique well for the same loop (Figure 8).

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

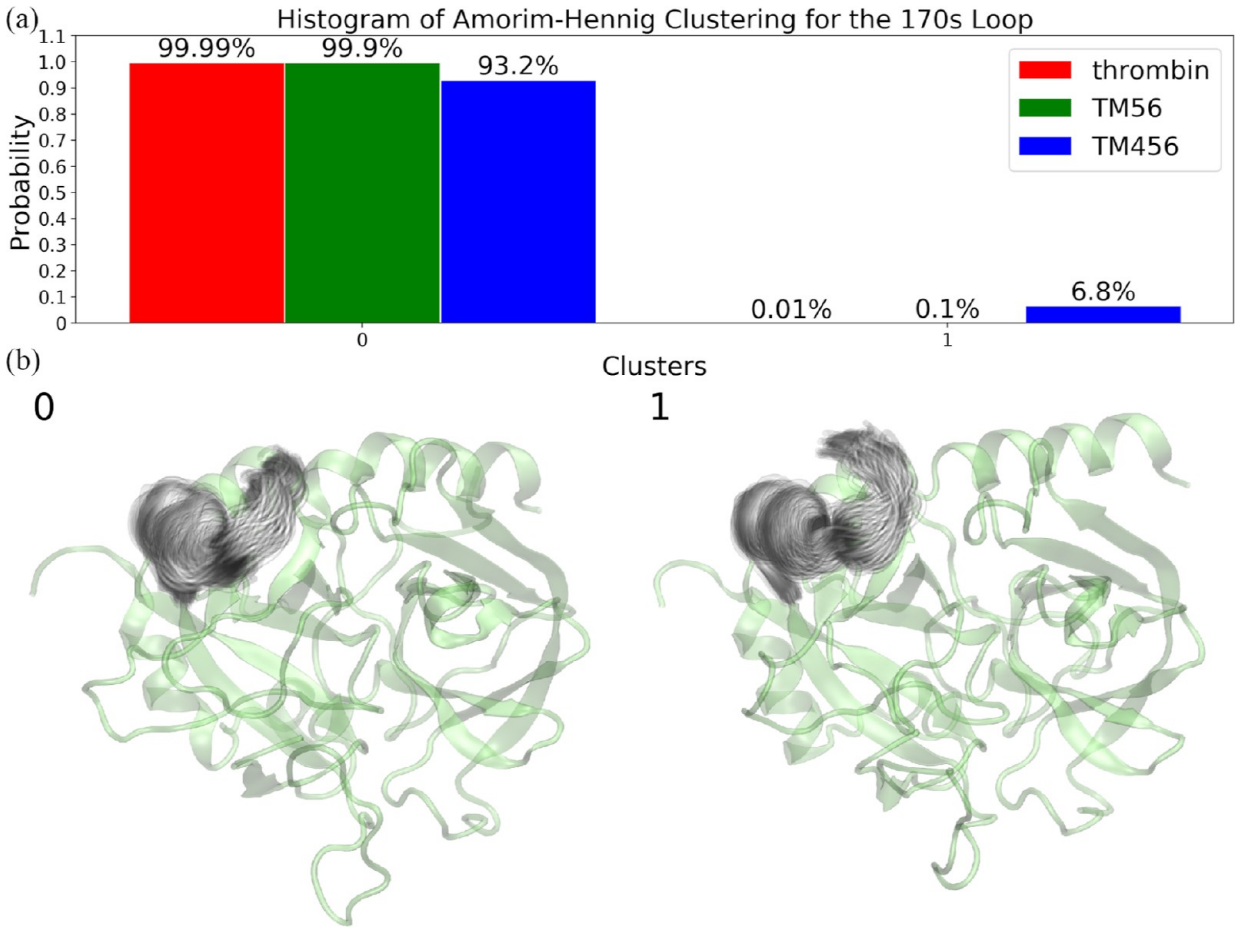


Figure 6. AH clustering for the heavy atoms of the 170s loop in thrombin, TM56-bound thrombin, and TM456-bound thrombin. (a) Distribution of clusters. (b) Representative structures for each cluster. The median structure of thrombin is represented via transparent green, while the variance of the 170s loop in each cluster is represented by shadows according to the visual statistics.[31]

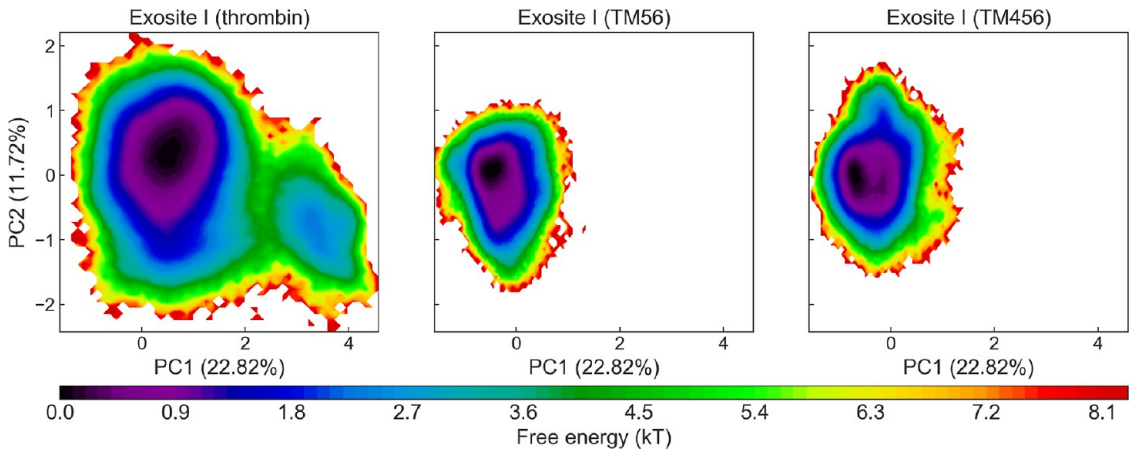


Figure 7. Conformational free-energy-like surfaces of exosite I in thrombin, TM56-bound thrombin, and TM456-bound thrombin.

Considering the role of the 180s loop in Na^+ binding, these new wells may represent alternative Na^+ binding configurations. Illustrations of the 180s loop structures in the four distinct wells are depicted in Figures S3b and S5.

Furthermore, PCA elucidates divergent impacts on the free-energy well of the γ loop resulting from TM56 and TM456 binding (Figure 9). The attachment of TM56 leads to an expansion of the energy well, while TM456 binding results in contraction and centralization of the well. The structures of the gamma-loop located in 12 separate wells are exhibited in Figures S3c and S6. Consistently, RMSF analysis (Figure 1) substantiates the increased atomic stability of the γ-loop post-TM456 binding, relative to TM56 binding or unbound thrombin. The atomic stability ranking is as follows: TM456 > *rbin* TM56 > thrombin. Consequently, the stabilized γ-loop following TM456 binding may provide a better scaffold for protein C interaction.

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

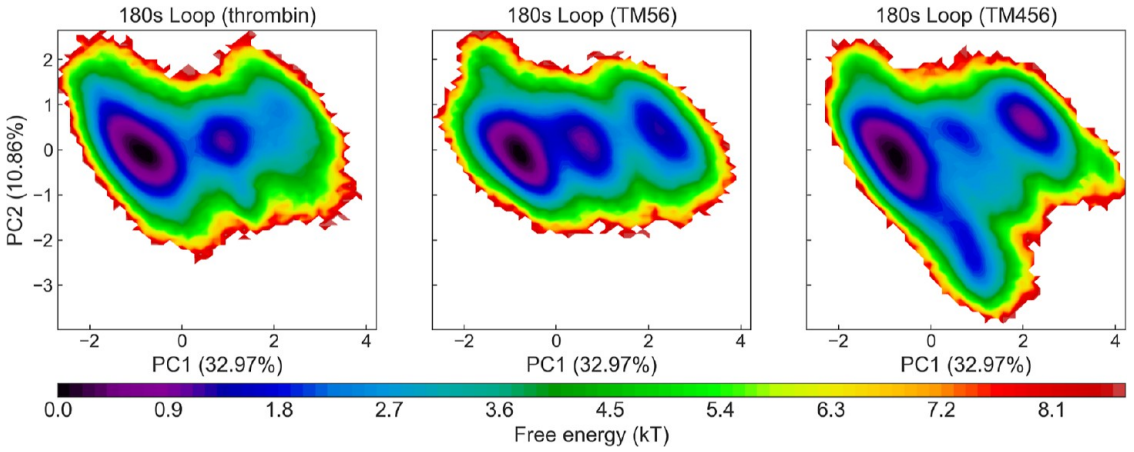


Figure 8. Conformational free-energy-like surfaces of the 180s loop in thrombin, TM56-bound thrombin, and TM456-bound thrombin.

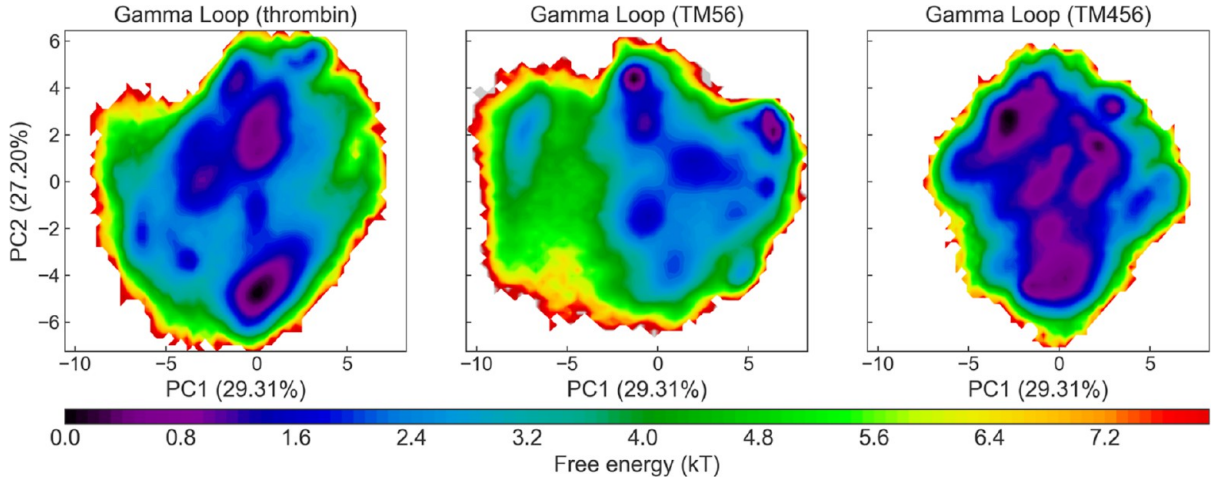


Figure 9. Conformational free-energy-like surfaces of the γ-loop in thrombin, TM56-bound thrombin, and TM456-bound thrombin.

Complementing our PCA findings, we further present the AH clustering results for the γ-loop in Figure S7a, overlaid on principal components 1 and 2 in Figure S7b. Figure S8 showcases the γ-loop structures specific to each cluster. The AH clustering results demonstrate good agreement with the free-energy well (Figure S3c). A significant portion of the γ-loop conformations in TM456-bound thrombin is found in clusters 0−3, suggesting these conformations could be beneficial for protein C binding. In contrast, free thrombin and TM56-bound thrombin exhibit 6.8% and 5.7% of their conformations within cluster 6, respectively. However, only a minor fraction (0.2%) of conformations in TM456-bound thrombin is observed in cluster 6, indicating that γ-loop structures in cluster 6 might not be conducive for protein C binding.

Dissecting the Correlated Motions within Thrombin: Differential Impacts of TM56 and TM456 Binding and the Crucial Role of PHE383. In addition to structural changes, we examined the correlated movements of thrombin, both with and without the binding of TM56 and TM456. The visualization of the correlation matrix reveals that TM56 binding predominantly affects the correlated movements within exosite I (Figure 10e), with little impact on the other areas of thrombin. Conversely, the binding of TM456 extends beyond exosite I, influencing correlated movements between the 170s loop and the 220s loop, the 180s loop and the 220s loop, the γ-loop and the 220s loop, and the 60s loop and exosite II (Figure 10f). This wider impact may account for why TM456 binding leads to a 1000-fold increase in the observed association rate for protein C, an effect not seen with TM56. Thus, while TM56 primarily affects the correlated movements of exosite I, TM456 significantly influences the correlated movements not only within exosite I but also among the various functional regions of thrombin.

The difference in the correlation matrix between TM56-bound and TM456-bound thrombin is depicted in Figure 10g. Notably, TM4 greatly amplifies the correlation within TM6 and between TM6 and thrombin. The residue PHE383 appears pivotal in transmitting the changes in correlated movements from TM4 to thrombin, thus signifying a potential target for experimental investigation. Based on this observation, we hypothesize that mutation of PHE383 may either eliminate or substantially attenuate the increase in the observed association rate for protein C.

Identification of Key Hydrogen Bonds in Thrombin: Insights into the Allosteric Effects of TM56 and TM456 Binding via Logistic Regression Models. To distinguish among thrombin, TM56-bound thrombin, and TM456-bound thrombin, we adopted a logistic regression model, using thrombin's hydrogen bonds as explanatory variables. This approach facilitated the discovery of key hydrogen bonds that define each unique state and could hold biological significance.

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

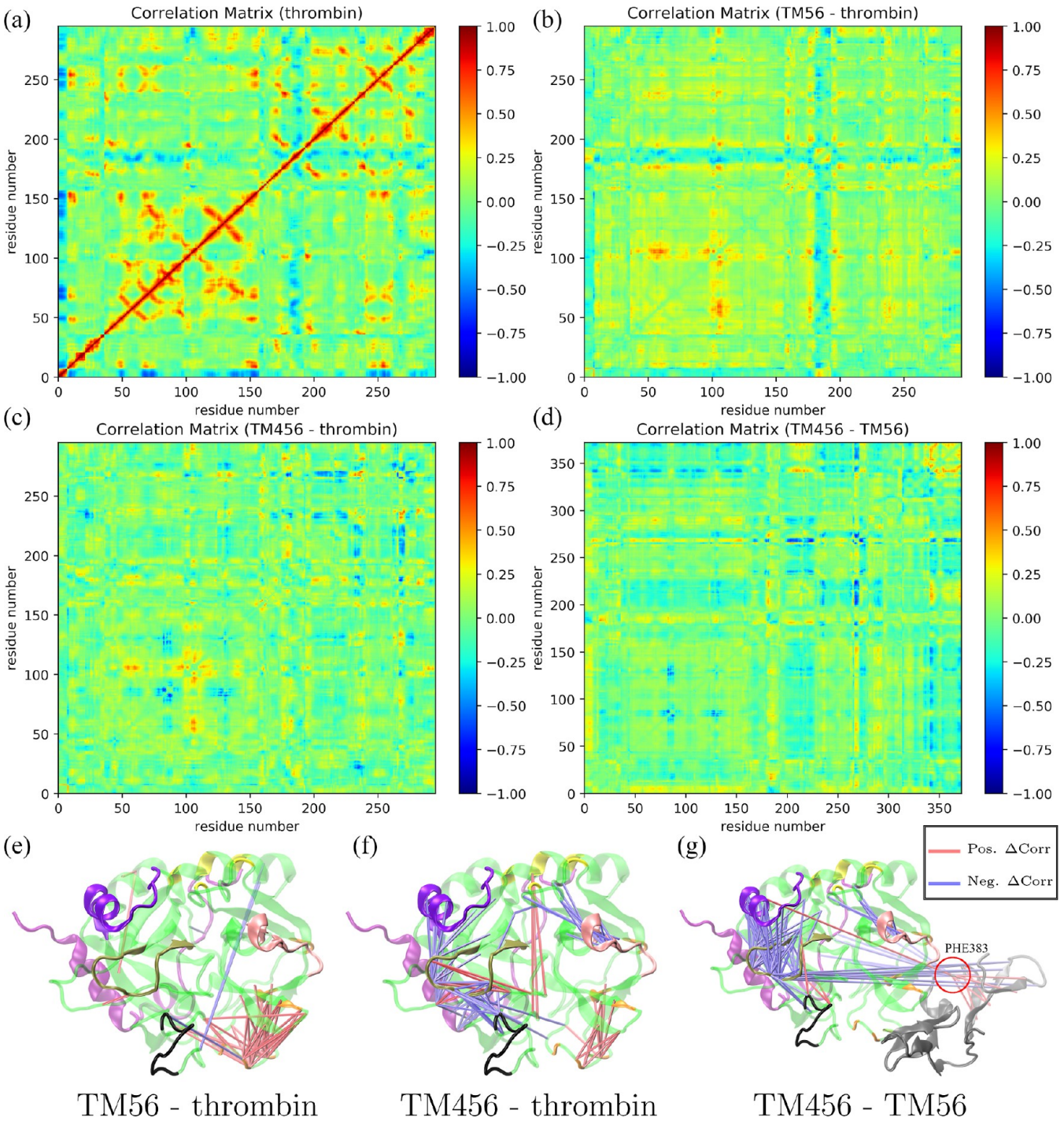


Figure 10. (a) Correlation matrix for thrombin. (b) Correlation matrix for TM56, with the correlation matrix of thrombin subtracted. (c) Correlation matrix for TM456, with the correlation matrix of thrombin subtracted. (d) Correlation matrix for TM456, with the correlation matrix of TM56 subtracted. Panels (e)−(g) depict the visualizations of the correlation matrices as illustrated in panels (b)−(d). In these visualizations, light-red and light-blue lines represent pairs of residues with positive and negative matrix subtractions, respectively, where the absolute value of the subtraction exceeds 0.5.

We extended the logistic regression model's application to distinguish TM56-bound thrombin from TM456-bound thrombin, based on hydrogen bonds within TM56 and those linking thrombin and TM56. This analysis illuminates the essential bonds involved in the allosteric pathway from TM4 to thrombin.

The logistic regression model, utilized to differentiate among thrombin, TM56-bound thrombin, and TM456-bound thrombin, achieved a prediction accuracy of 92.5%. Using just the top 29 hydrogen bonds within the model yielded an 83.8% accuracy, maintaining over 90% of the original accuracy, thereby underlining these 29 hydrogen bonds as vital differentiators. For thrombin, the hydrogen bond of greatest significance was identified as SER58-GLN60 (Table 1), disrupted upon TM56 and TM456 binding. This disruption aligns with TM56 and TM456 binding to exosite I, as it is

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

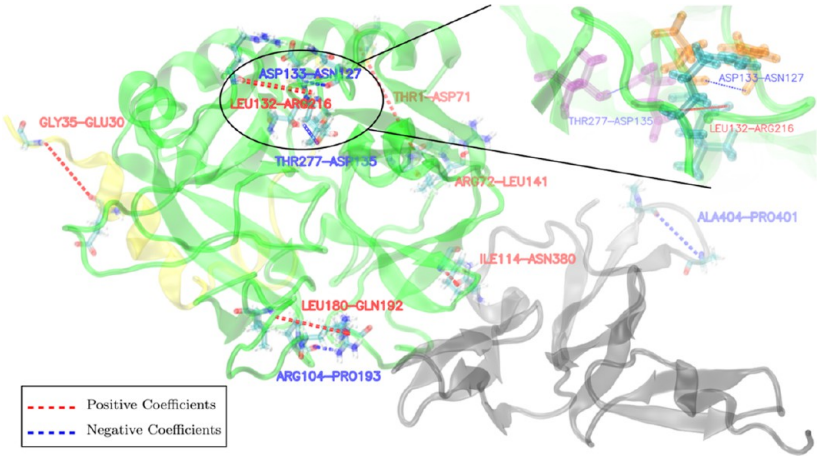


Figure 11. Detailed depiction of eight critical hydrogen bonds identified by the logistic regression model as either forming or breaking upon the binding of TM456 to thrombin. The light and heavy chains of thrombin are visually represented via NewCartoon in transparent yellow and green hues, respectively, while TM456 is illustrated in a similar style in transparent gray. Associated residues are displayed in licorice. Hydrogen bonds are indicated by dashed lines, with blue denoting negative beta values and red indicating positive beta values. The thickness of these dashed lines represents the absolute value of the beta values. The names of the hydrogen bonds are labeled directly on the corresponding dashed lines. An inset at the top right corner provides an enlarged view of the specific hydrogen bonds LEU132−ARG216, THR277−ASP135, and ASP133−ASN127.

Table 1. Logistic Regression Beta-Values

	hydrogen bonds	beta-values
thrombin	SER58−GLN60	6.56
	LEU132−ARG216	−3.75
	TYR32−LYS248	−3.22
	GLU112−GLU108	−2.00
	TYR32−LYS171	−1.97
	ARG98−ARG106	−1.64
	ARG56−LYS57	1.46
	GLN60−LYS57	−1.45
	ARG56−GLU61	−1.43
	TYR83−LYS88	−1.30
	ARG173−ASP22	−1.28
	ARG109−TYR107	−1.27
	GLY46−HIS102	−1.24
	ARG56−LYS57	−4.68
	ARG56−PRO59	−2.99
	TYR126−LEU82	1.97
TM56	PHE275−TRP263	−1.88
	ARG233−GLU39	−1.77
	HIS123−PRO124	1.65
	ARG216−ASP211	−1.41
	ARG104−GLU61	−1.19
	LEU132−ARG216	3.26
	ARG104−PRO193	−1.76
	GLY35−GLU30	1.65
TM456	THR1−ASP71	1.53
	THR277−ASP135	−1.39
	ARG72−LEU141	1.33
	ASP133−ASN127	−1.24
	LEU180−GLN192	1.16

likely to influence the loop where exosite I is located. For TM56-bound thrombin, the top two hydrogen bonds—ARG56-LYS57 and ARG56-PRO59 (Table 1)—are located within the loop hosting exosite I, accounting for why TM56 binding mainly impacts the correlated movement within exosite I (Figure 10e).

As for the TM456-bound thrombin, eight pivotal hydrogen bonds are depicted in Figure 11, including the formation of the hydrogen bond LEU132−ARG216 and the disruption of the THR277−ASP135 and ASP133−ASN127 bonds (Figure 11, top right). These changes potentially clarify the observed effects of TM456 binding on the catalytic triad (residues 79, 135, and 241), as revealed in earlier clustering analysis (Figure 5a). This concurs with previous experimental observations, suggesting that the alteration of ASP135 might be transmitted from the 90s loop (residues 127−133) by a pair of β-strands,[19] enabling catalysis. The hydrogen-bond formation between LEU132 and ARG216 may explain the unique conformational shift observed in the 170s loop upon TM456 binding (Figure 6). Changes in the γ loop's free-energy well following TM456 binding, as shown in the earlier PCA analysis (Figure 9), might be attributed to the formation of the hydrogen bond LEU180−GLN192 and disruption of the ARG104−PRO193 bond. The formation of hydrogen bonds GLY35−GLU30 and THR1−ASP71 emphasizes the light chain's significance in TM456 binding's allosteric effects. The formation of hydrogen bonds ARG72−LEU141 and THR1−ASP71 suggests a potential allosteric pathway from exosite I (residues 142 and 143) to the light chain.

When differentiating between TM56-bound thrombin and TM456-bound thrombin, the logistic regression model achieved a prediction accuracy of 71.7%. Upon using just the top two hydrogen bonds, the model's accuracy dropped slightly to 67.5%, so it still retained over 90% of the original accuracy. These hydrogen bonds, ALA404−PRO401 and ILE114−ASN380 (Table 2), suggest how TM4 could exert

Table 2. Logistic Regression Beta-Values for Differentiating between Thrombin Bound to TM56 and TM456 considering Intra-TM56 Hydrogen Bonds and Hydrogen Bonds between TM56 and Thrombin

	hydrogen bonds	beta-values
TM56−TM456	ALA404−PRO401	−1.00
	ILE114−ASN380	0.94

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

allosteric influence on TM56 and thrombin. The formation of the ILE114−ASN380 hydrogen bond suggests it could be key in how PHE383 transmits correlated motions from TM4 to thrombin (Figure 10g).

Insights from Comparative Analysis with Previous Works. The γ loop, 60s loop, and exosite I are regions substantially influenced by TM binding and are particularly responsive to mutations or deletions, albeit with different outcomes.[22] This underscored the functional importance of these regions, emphasizing their susceptibility to TM binding or sequence modifications, and their consequential regulatory role on thrombin function.

Our study, from a hydrogen-bonding perspective, uncovers different allosteric pathways for TM binding and for effects elicited by mutations or deletions.[24] This variation is to be expected since TM binding and mutations occur at distinct locations.

However, a noteworthy hydrogen bond, THR277−ASP135, is significant for both TM456 and E8K, hinting at a potential shared allosteric pathway leading to the catalytic triad.

In an earlier study, we noted that the 60s loop only undergoes a twisting motion when a single Na^+ is bound to thrombin, possibly conferring a procoagulant property to thrombin.[23] In the current investigation, we found that even when TM456 is bound to thrombin, the 60s loop retains its ability to twist, an ability lost when TM56 is bound.

We thus hypothesize that the twisting of the 60s loop may also be critical for protein C binding. This conjecture is supported by the observation that the thrombin−TM456 complex boosts the k_a value for protein C binding more than 1000-fold, an effect not replicated by the binding of TM56.

CONCLUSIONS

In summary, our study provides a detailed investigation of the allosteric modulation of thrombin by the TM fragments TM56 and TM456 using a combination of MD simulations, statistical analysis, and machine learning methods. This investigation supports the notion that TM binding triggers allosteric structural changes in both the active site and functional regions of thrombin,[12,18−20] rather than merely affecting the conformation of bound protein C[9,13,14] or acting as a scaffold to facilitate thrombin−protein C interactions.[10]

Our application of HDBSCAN and AH clustering techniques has revealed that the binding of TM456, the smallest active cofactor fragment of TM, leads to significant allosteric conformational changes in thrombin's key loops and catalytic triad. This is in stark contrast to the minimal changes observed upon TM56 binding, which does not enable thrombin to activate protein C. Correlation analysis suggests that while TM56 modifies correlated motions within exosite I, TM456 induces more extensive changes across multiple functional regions of thrombin, underscoring the critical role of TM4 in allosteric regulation. Logistic regression was instrumental in identifying critical hydrogen bonds such as LEU132−ARG216 and THR277−ASP135, shedding light on their potential roles in thrombin's biological functions and allosteric pathways. Through comparative analysis, our study also highlights the influence of TM binding on the crucial regions of thrombin, suggesting possible common allosteric pathways, thereby enhancing our understanding of thrombin regulation.

Based on our findings, future research should focus on the hydrogen bonds and critical residues identified in our study, such as PHE383. Future investigations could use accelerated sampling techniques, such as replica-exchange with selective variables chosen based on our hydrogen-bond analysis. Additionally, virtual screening based on the key residues pinpointed in our research could be productive. Another promising avenue is to explore how mutations influence thrombin's allostery. Our current research examines the influence of W215A/E217A mutations on thrombin's allostery in response to TM binding. This sheds light on the intricate allosteric regulation of thrombin.[67]

ASSOCIATED CONTENT

Data Availability Statement

All associated code utilized in this study is readily accessible at a Github repository at https://github.com/salsburygroup/Thrombin_Thromodulin_LR. Due to their substantial size, the raw trajectory files are not directly available online. However, they can be obtained upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c03375.

Time series of the RMSD of alpha-carbons relative to the initial structure of thrombin across eight runs, for free thrombin, TM56-bound thrombin, and TM456-bound thrombin, respectively; RMSF of the α-carbons of thrombin in free thrombin, TM56-bound thrombin, and TM456-bound thrombin, across all eight runs; conformational free-energy-like surfaces of exosite I, 180s loop, and gamma-loop for the concatenated trajectory of free thrombin, TM56-bound thrombin, and TM456-bound thrombin; representative structures for exosite I in wells 0 and 1; representative structures for 180s loop in wells 0−3; representative structures for gamma-loop in wells 0−11; performing AH clustering on the heavy atoms of the gamma-loop in thrombin and TM56-bound thrombin; and gamma-loop representative structures for each cluster derived from AH clustering results (PDF)

AUTHOR INFORMATION

Corresponding Author

Freddie R. Salsbury, Jr. − *Department of Physics, Wake Forest University, Winston-Salem, North Carolina 27106, United States;* orcid.org/0000-0002-2699-829X; Email: salsbufr@wfu.edu

Author

Dizhou Wu − *Department of Physics, Wake Forest University, Winston-Salem, North Carolina 27106, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c03375

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the Wake Forest Baptist Comprehensive Cancer Center Crystallography & Computational Biosciences Shared Resource, supported by the National Cancer Institute's Cancer Center Support Grant award number P30CA012197. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute. Portions of

the computations were performed on the Wake Forest University DEAC Cluster, a centrally managed resource with support provided in part by Wake Forest University.[52] F.R.S. would also like to thank the Scott Family for the Scott Family Fellowship. D.W. would like to thank the Center for Molecular Signaling at Wake Forest University for a fellowship.

REFERENCES

(1) Davie, E. W.; Fujikawa, K.; Kisiel, W. The Coagulation Cascade: Initiation, Maintenance, and Regulation. *Biochemistry* **1991**, *30*, 10363−10370.
(2) Di Cera, E. Thrombin. *Mol. Aspects Med.* **2008**, *29*, 203−254.
(3) Rezaie, A. R.; Cooper, S. T.; Church, F. C.; Esmon, C. T. Protein C Inhibitor Is a Potent Inhibitor of the Thrombin-Thrombomodulin Complex. *J. Biol. Chem.* **1995**, *270*, 25336−25339.
(4) Van De Locht, A.; Bode, W.; Huber, R.; Le Bonniec, B. F.; Stone, S. R.; Esmon, C. T.; Stubbs, M. T. The Thrombin E192Q−BPTI Complex Reveals Gross Structural Rearrangements: Implications for the Interaction with Antithrombin and Thrombomodulin. *EMBO J.* **1997**, *16*, 2977−2984.
(5) Rezaie, A. R.; He, X.; Esmon, C. T. Thrombomodulin Increases the Rate of Thrombin Inhibition by BPTI. *Biochemistry* **1998**, *37*, 693−699.
(6) Myles, T.; Church, F. C.; Whinna, H. C.; Monard, D.; Stone, S. R. Role of Thrombin Anion-Binding Exosite-I in the Formation of Thrombin-Serpin Complexes. *J. Biol. Chem.* **1998**, *273*, 31203−31208.
(7) De Cristofaro, R.; Landolfi, R. Allosteric Modulation of BPTI Interaction with Human α-and ζ-Thrombin. *Eur. J. Biochem.* **1999**, *260*, 97−102.
(8) Esmon, C. T. The Roles of Protein C and Thrombomodulin in the Regulation of Blood Coagulation. *J. Biol. Chem.* **1989**, *264*, 4743−4746.
(9) Vindigni, A.; White, C. E.; Komives, E. A.; Di Cera, E. Energetics of Thrombin-Thrombomodulin Interaction. *Biochemistry* **1997**, *36*, 6674−6681.
(10) Fuentes-Prior, P.; Iwanaga, Y.; Huber, R.; Pagila, R.; Rumennik, G.; Seto, M.; Morser, J.; Light, D. R.; Bode, W. Structural Basis for the Anticoagulant Activity of the Thrombin−Thrombomodulin Complex. *Nature* **2000**, *404*, 518−525.
(11) Xu, H.; Bush, L. A.; Pineda, A. O.; Caccia, S.; Di Cera, E. Thrombomodulin Changes the Molecular Surface of Interaction and the Rate of Complex Formation Between Thrombin and Protein C. *J. Biol. Chem.* **2005**, *280*, 7956−7961.
(12) Handley, L. D.; Treuheit, N. A.; Venkatesh, V. J.; Komives, E. A. Thrombomodulin Binding Selects the Catalytically Active Form of Thrombin. *Biochemistry* **2015**, *54*, 6650−6658.
(13) Di Cera, E.; Dang, Q. D.; Ayala, Y. M. Molecular Mechanisms of Thrombin Function. *Cell. Mol. Life Sci.* **1997**, *53*, 701−730.
(14) Hayashi, T.; Zushi, M.; Yamamoto, S.; Suzuki, K. Further Localization of Binding Sites for Thrombin and Protein C in Human Thrombomodulin. *J. Biol. Chem.* **1990**, *265*, 20156−20159.
(15) Di Cera, E.; Guinto, E. R.; Vindigni, A.; Dang, Q. D.; Ayala, Y. M.; Wuyi, M.; Tulinsky, A. The Na+ Binding Site of Thrombin. *J. Biol. Chem.* **1995**, *270*, 22089−22092.
(16) Dang, Q. D.; Guinto, E. R.; Cera, E. D. Rational Engineering of Activity and Specificity in a Serine Protease. *Nat. Biotechnol.* **1997**, *15*, 146−149.
(17) Pineda, A. O.; Chen, Z.-W.; Caccia, S.; Cantwell, A. M.; Savvides, S. N.; Waksman, G.; Mathews, F. S.; Di Cera, E. The Anticoagulant Thrombin Mutant W215A/E217A Has a Collapsed Primary Specificity Pocket. *J. Biol. Chem.* **2004**, *279*, 39824−39828.
(18) Koeppe, J. R.; Seitova, A.; Mather, T.; Komives, E. A. Thrombomodulin tightens the thrombin active site loops to promote protein C activation. *Biochemistry* **2005**, *44*, 14784−14791.
(19) Peacock, R. B.; McGrann, T.; Tonelli, M.; Komives, E. A. Serine Protease Dynamics Revealed by NMR Analysis of the Thrombin-Thrombomodulin Complex. *Sci. Rep.* **2021**, *11*, 9354.
(20) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R. L.; McCammon, J. A. Allosteric Networks in Thrombin Distinguish Procoagulant vs. Anticoagulant Activities. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 21216−21222.
(21) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(22) Wu, D.; Xiao, J.; Salsbury, F. R. Light Chain Mutation Effects on the Dynamics of Thrombin. *J. Chem. Inf. Model.* **2021**, *61*, 950−965.
(23) Wu, D.; Salsbury, F. R. Simulations Suggest Double Sodium Binding Induces Unexpected Conformational Changes in Thrombin. *J. Mol. Model.* **2022**, *28*, 120−216.
(24) Wu, D.; Salsbury, F. R. Unraveling the Role of Hydrogen Bonds in Thrombin via Two Machine Learning Methods. *J. Chem. Inf. Model.* **2023**, *63*, 3705−3718.
(25) Xiao, J.; Melvin, R. L.; Salsbury, F. R. Mechanistic Insights Into Thrombin's Switch Between "Slow" and "Fast" Forms. *Phys. Chem. Chem. Phys.* **2017**, *19*, 24522−24533.
(26) Xiao, J.; Salsbury, F. R. Molecular Dynamics Simulations of Aptamer-Binding Reveal Generalized Allostery in Thrombin. *J. Biomol. Struct. Dyn.* **2017**, *35*, 3354−3369.
(27) Xiao, J.; Salsbury, F. R. Na+-Binding Modes Involved in Thrombin's Allosteric Response as Revealed by Molecular Dynamics Simulations, Correlation Networks and Markov Modeling. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4320−4330.
(28) Xiao, J.; Melvin, R. L.; Salsbury, F. R. Probing Light Chain Mutation Effects on Thrombin via Molecular Dynamics Simulations and Machine Learning. *J. Biomol. Struct. Dyn.* **2019**, *37*, 982−999.
(29) Melvin, R. L.; Gmeiner, W. H.; Salsbury, F. R. J. All-Atom Molecular Dynamics Reveals Mechanism of Zinc Complexation with Therapeutic F10. *J. Phys. Chem. B* **2016**, *120*, 10269−10279.
(30) Melvin, R. L.; Godwin, R. C.; Xiao, J.; Thompson, W. G.; Berenhaut, K. S.; Salsbury, F. R. Uncovering Large-Scale Conformational Change in Molecular Dynamics Without Prior Knowledge. *J. Chem. Theory Comput.* **2016**, *12*, 6130−6146.
(31) Melvin, R. L.; Salsbury, F. R. Visualizing Ensembles in Structural Biology. *J. Mol. Graphics Modell.* **2016**, *67*, 44−53.
(32) Melvin, R. L.; Gmeiner, W. H.; Salsbury, F. R. J. All-Atom MD Predicts Magnesium-Induced Hairpin in Chemically Perturbed RNA Analog of F10 Therapeutic. *J. Phys. Chem. B* **2017**, *121*, 7803−7812.
(33) Melvin, R. L.; Gmeiner, W. H.; Salsbury, F. R. All-Atom MD Indicates Ion-Dependent Behavior of Therapeutic DNA Polymer. *Phys. Chem. Chem. Phys.* **2017**, *19*, 22363−22374.
(34) Melvin, R. L.; Thompson, W. G.; Godwin, R. C.; Gmeiner, W. H.; Salsbury, F. R., Jr. Mutsα's Multi-Domain Allosteric Response to Three Dna Damage Types Revealed by Machine Learning. *Front. Phys.* **2017**, *5*, 10.
(35) Godwin, R.; Gmeiner, W.; Salsbury, F. R. Importance of Long-Time Simulations for Rare Event Sampling in Zinc Finger Proteins. *J. Biomol. Struct. Dyn.* **2016**, *34*, 125−134.
(36) Godwin, R. C.; Melvin, R. L.; Gmeiner, W. H.; Salsbury, F. R. Binding Site Configurations Probe the Structure and Dynamics of the Zinc Finger of Nemo (Nf-κB Essential Modulator). *Biochemistry* **2017**, *56*, 623−633.
(37) Godwin, R. C.; Macnamara, L. M.; Alexander, R. W.; Salsbury, F. R. Structure and Dynamics of tRNAMet Containing Core Substitutions. *ACS Omega* **2018**, *3*, 10668−10678.
(38) Godwin, R. C.; Gmeiner, W. H.; Salsbury, F. R. All-atom Molecular Dynamics Comparison of Disease-associated Zinc Fingers. *J. Biomol. Struct. Dyn.* **2018**, *36*, 2581−2594.
(39) Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(40) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. A. C. E. M. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632−1639.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100

(42) Russo Krauss, I.; Merlino, A.; Randazzo, A.; Novellino, E.; Mazzarella, L.; Sica, F. 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 Res.* **2012**, *40*, 8119−8128.
(43) Berendsen, H. J. C.; Postma, J. P. M. v.; van Gunsteren, W. F.; DiNola, A. R. H. J.; Haak, J. R. Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(44) Lemons, D. S.; Gythiel, A. Paul Langevin's 1908 Paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," CR Acad. Sci. (Paris) 146, 530−533 (1908)]. *Am. J. Phys.* **1997**, *65*, 1079−1081.
(45) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(46) Harvey, M. J.; De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory Comput.* **2009**, *5*, 2371−2377.
(47) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20*, 786−798.
(48) Van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311−1327.
(49) Godwin, R. C.; Melvin, R.; Salsbury, F. R. *Computer-Aided Drug Discovery*; Springer, 2015; pp 1−30.
(50) Campello, R. J. G. B.; Moulavi, D.; Sander, J. *Advances in Knowledge Discovery and Data Mining*; Springer Berlin Heidelberg, 2013; pp 160−172.
(51) De Amorim, R. C.; Hennig, C. Recovering the Number of Clusters in Data Sets With Noise Features Using Feature Rescaling Factors. *Inf. Sci.* **2015**, *324*, 126−145.
(52) Systems, I.; University, W. F. WFU High Performance Computing Facility, 2021. https://hpc.wfu.edu (accessed May 16, 2024).
(53) Abdi, H.; Williams, L. J. Principal Component Analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433−459.
(54) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(55) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins: Struct., Funct., Bioinf.* **1991**, *11*, 205−217.
(56) Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. Hydrogen Bonds Involving Sulfur Atoms in Proteins. *Proteins: Struct., Funct., Bioinf.* **1991**, *9*, 99−107.
(57) Fersht, A. R.; Serrano, L. Principles of Protein Stability Derived from Protein Engineering Experiments. *Curr. Opin. Struct. Biol.* **1993**, *3*, 75−83.
(58) Schell, D.; Tsai, J.; Scholtz, J. M.; Pace, C. N. Hydrogen Bonding Increases Packing Density in the Protein Interior. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 278−282.
(59) van Rossum, G. *Python Reference Manual*; Department of Computer Science [CS], 1995.
(60) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(61) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*; SciPy, 2016; p 105.
(62) Steiner, T. The Hydrogen Bond in the Solid State. *Angew. Chem., Int. Ed.* **2002**, *41*, 48−76.
(63) Schein, A. I.; Ungar, L. H. Active Learning for Logistic Regression: An Evaluation. *Mach. Learn.* **2007**, *68*, 235−265.
(64) Sani, H. M.; Lei, C.; Neagu, D. Computational Complexity Analysis of Decision Tree Algorithms. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*; Springer International Publishing, 2018; pp 191−197.
(65) LaValley, M. P. Logistic Regression. *Circulation* **2008**, *117*, 2395−2399.
(66) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer, 2013; Vol. *112*.
(67) Wu, D.; Salsbury, F., Jr. Unraveling the Impact of W215A/E217A Mutations on Thrombin Dynamics and Thrombomodulin Binding through Molecular Dynamics Simulations. *bioRxiv* **2023**, 2023−2112.

https://doi.org/10.1021/acsomega.4c03375
ACS Omega 2024, 9, 23086−23100