

Long-Range Changes in Neurolysin Dynamics Upon Inhibitor Binding

Uyar, A., Karamyan, V. T. and Dickson, A.*

Running title: Long-Range Changes in Neurolysin Motions

ABSTRACT

Crystal structures of neurolysin, which is a zinc metallopeptidase (neuropeptidase), do not show significant conformational changes upon the binding of an allosteric inhibitor. Neurolysin has a prolate ellipsoid shape with a deep channel that runs almost the entire length of the molecule. In this deep channel, neurolysin hydrolyzes the short neuropeptide neuropeptidase to create inactive shorter fragments and thus controls the neuropeptidase level in the tissue. The protein is of interest as a therapeutic target since changes in neuropeptidase level have been implicated in cardiovascular and neurological disorders and cancer, and inhibitors of neurolysin activity have been developed. Therefore, understanding the dynamical and structural differences between the apo and inhibitor-bound forms of neurolysin can aid in further design of potent inhibitors and activators. For this purpose, we performed several molecular dynamics (MD) simulations for both apo and neurolysin bound to an allosteric inhibitor. The collective dynamics of neurolysin is examined using Principal Component Analysis and Elastic Network Model calculations, and common collective motions are revealed in both neurolysin forms. MD simulations show that allosteric inhibitor binding induces additional low-frequency motions that are not observed in the apo form. ENM calculations reveal changes in inter- and intra-domain communication upon binding. Furthermore, differences are observed in the inhibitor-bound neurolysin contact network that are far from the active site, which reveals long-range allosteric behavior in neurolysin. Finally, a machine learning method Linear Discriminant Analysis is applied to reveal differences between the apo and inhibitor-bound ensembles in an automated way, and large differences are observed on residues that are far from both the active site and the inhibitor binding site, which are also in good agreement with the contact network results. This study, using different computational methods on neurolysin, can provide insight into the allosteric modulation of other neuropeptidases with similar folds.

INTRODUCTION

Upon binding, a ligand can modulate the function of a protein either directly (orthosteric) or indirectly (allosteric). Typically, an allosteric ligand works by causing a conformational change in the protein upon binding, although in other cases, the ligand can inhibit or activate the protein without an observable structural change as observed for hemoglobin (1), CheY (2), catabolite activator protein (3, 4), and several protein kinases (5). This can make the mechanism of allosteric inhibition especially difficult to discern. In such cases, there are both experimental and computational studies that suggest *dynamics* plays an important role (6–9). Neurolysin, which is a zinc metallopeptidase (neuropeptidase), is such an example, where a crystal structure with an allosteric inhibitor compared to the apo structure does not show any significant conformational change, although it is known to exhibit a large hinge-bending motion in solution. In this study, we use a set of multi-scale computational tools to examine changes in neurolysin dynamics that occur upon inhibitor binding.

Neurolysin is a monomeric enzyme, 704 residues in length, which belongs to the M3 family of metallopeptidases (10). The main function of this peptidase is to control extracellular levels of a number of short neuropeptides including *neurotensin*, by hydrolyzing them in its deep channel using a zinc ion as a cofactor (11). Changes in neurolysin level and that of its substrates have been documented in a number of cardiovascular and neurological disorders and cancer (12, 13) linking them to pathogenic mechanisms of these disorders. Based on these observations, modulation of neurolysin activity was deemed to carry a therapeutic potential which was partially demonstrated by experimental studies in animals (12–14). In this respect, there has been much interest in developing compounds that target the neurolysin active site (15–17). Nearly all known neuropeptidase inhibitors target neuropeptidase active sites, including inhibitors of angiotensin converting enzyme (ACE) in hypertension treatment (18), and inhibitors of thermolysin and pseudolysin (19). However, the recent crystal structure of neurolysin in complex with an allosteric inhibitor (PDB ID: 4FXY, (20)) reveals that there might be other attractive sites that can be targeted for the inhibition of neuropeptidases. This allosteric inhibitor is bound to a site approximately 20 Å away from the active site and a significant conformational change is not observed between apo and inhibitor-bound structures. Therefore, elaboration of short- and long-range effects of the binding of this allosteric inhibitor both on neurolysin structure and dynamics might open new opportunities for further drug design that specifically targets neuropeptidases.

Using fluorescence polarization, it has been shown that neurolysin is shifted toward a more compact, closed conformation when the allosteric inhibitor is bound (20). In light of this finding, Hines *et al.* (20) proposed an inhibition mechanism suggesting that binding stabilizes an inactive neurolysin conformation, which is more closed compared to its active state (apo), and is incompatible with hydrolysis of the neurotensin peptide. However, the extent to which the inhibitor dampens the characteristic hinge bending motion of neurolysin is still unknown. To address this question, we examine neurolysin dynamics in both the apo and inhibitor bound states using computational tools.

In this paper we perform several independent CHARMM (21–23) all-atom molecular dynamics (MD) simulations (in total 775 ns for each form of neurolysin).

We also perform coarse-grained Elastic Network Model (ENM (24, 25)) calculations to examine the collective dynamics in the apo and inhibitor bound states, and to compare normal modes with principal components of the MD trajectories, obtained by Principal Component Analysis (PCA). MD simulations also reveal significant long-range changes in neurolysin contact network upon inhibitor binding. We then apply Linear Discriminant Analysis to automatically identify structural differences between apo and inhibitor-bound forms. These findings also enlighten structure-function relationships in other metallopeptidases, since neurolysin has both sequence and structural homology with other metallopeptidases such as thimet oligopeptidase (TOP), angiotensin I converting enzymes (ACE and ACE-2), thermolysin, dipeptidyl carboxypeptidase (Dcp), and carboxypeptidase Pfu.

METHODS

Elastic Network Model (ENM)

ENM (24, 25), which is a coarse-grained method for normal mode analysis, gives exact predictions of vibrational harmonic motions of proteins around a given single conformation. The system is represented by a network of nodes, which here represent the Ca atoms of each residue in the protein. The node pairs that fall within a certain cutoff distance (in this study, 10 Å) are connected by uniform harmonic springs to construct the elastic network. The total potential energy of the network is the summation of all harmonic interactions of the node pairs. The frequency of each harmonic mode (eigenvalues) and corresponding direction information (eigenvectors) are obtained from ENM analysis through diagonalization of the Hessian matrix (H), whose elements are the second partial derivative of the potential function with respect to each direction. For the elastic network of inhibitor-bound neurolysin, the ligand is also coarse-grained and represented by three atoms. The Zinc ion is not included in ENM calculations.

Molecular Dynamics Simulations (CHARMM)

We performed 32 independent CHARMM (21–23) MD simulations for apo and inhibitor-bound neurolysin crystal structures (PDB ID: 1III (11) and 4FXY (20), respectively) using the CHARMM36 force-field (26) (in total 775 ns for each system). The first 13 residues in the apo protein (14 for inhibitor-bound) are missing in the crystal structure and are omitted from the simulation. The inhibitor molecule is parameterized using the CHARMM Generalized Force Field (CGenFF) (27, 28). Each structure is solvated in a cubic box with 12 Å of distance between the protein and the cell boundaries. The zinc ion in the active site was preserved and 12 sodium ions were added to neutralize the system in both cases. Two histidine residues (residues 474 and 478), which are coordinating the zinc ion through hydrogen bonds (data from PDBsum (29, 30)), were protonated at NE2 position. The SHAKE algorithm was used to constrain hydrogen atoms with a tolerance of 10^{-8} . PME and a van der Waals switching function with a cutoff radius of 10 Å were used for non-bonded interactions. Each system was minimized with constraints on the protein atoms for 500 steps using steepest descent and adopted basis Newton-Raphson methods. The systems were heated from 50 to 300 K over 100 ps and equilibrated for 0.5 ns at 300 K. The final structures were used as initial coordinates for the production runs, and

different initial velocities were given to each of the 32 replicates. Langevin dynamics was performed at 300 K using CHARMM with OpenMM version 7.0 (31) for GPU acceleration.

Linear Discriminant Analysis (LDA)

LDA (32, 33) is a machine learning method for dimensionality reduction that is used in multi-class systems by minimizing intra-class information and maximizing the separation between classes. In machine learning applications it can be used as a predictor that, given a set of training data, can automatically assign new data to the set of classes. In the context of molecular simulation, this can also be useful to detect differences between datasets in an automated way (34).

Assuming that there are M classes representing different MD simulation conditions (in our case $M = 2$ as apo and inhibitor-bound form) that we would like to compare, the coordinates are given as q_m for each MD trajectory where m is the class index. Here we use the x , y and z coordinates for each of the N C α atoms in the system. Each class has n_m structures total. Using a projection matrix (W), one can transform the high-dimensional data q_m into low-dimensional data y_m , and the projected coordinates of the m th simulation and t th snapshot can be expressed as $y_{m,t} = W^T q_{m,t}$. The centroid of the whole data is determined (\bar{y}), as well as the centroid for each specific class (\bar{y}_m). Two measures of variance are then computed. The *within scatter* measures the average variance within the given classes:

$$\Phi_{\text{within}} = \sum_{m=1}^M \frac{1}{n_m} \sum_{t=1}^{n_m} \|y_{m,t} - \bar{y}_m\|^2 \quad (1)$$

where $\|\cdot\|^2$ is the Euclidean distance. The *between scatter* measures the variance between the classes:

$$\Phi_{\text{between}} = \sum_{m=1}^M \|\bar{y}_m - \bar{y}\|^2 \quad (2)$$

The ratio between these two quantities ($\rho = \phi_{\text{between}}/\phi_{\text{within}}$) is then iteratively maximized by changing the orthogonal projection matrix (W) as described in the work by Sakuraba and Kono (34). Here we use two LDA modes, and W has the dimensions $(2, 3N)$. The first LDA mode vector is used to separate the apo and inhibitor bound classes, and the elements of this vector show the specific residues that best describe the difference between the classes.

RESULTS

Collective dynamics from Elastic Network Model calculations

Apo (11) and inhibitor-bound (IB) (20) neurolysin crystal structures do not show a significant difference ($C\alpha$ RMSD = 0.6 Å), and both have the same prolate ellipsoid shape with a deep channel almost 60 Å length as shown in Fig. 1A. The deep channel divides the structure into two domains: I and II. The active site sequence motif of neurolysin (His-Glu-x-x-His, yellow), which is conserved in this family of metallopeptidases (35), and the catalytic zinc ion (green sphere) are located in domain II. The active site loops in domain I (fuchsia) and II (orange) possess key residues for peptide recognition (36). The allosteric inhibitor (shown in red licorice representation) is bound to a distant site (~20 Å) from the active site zinc ion. As there is no conformational change upon binding it has been hypothesized that the inhibitor might affect the enzyme activity by inhibiting the hinge-like motion of neurolysin and keeping the structure in an open but inactive form, preventing adoption of the fully closed form needed for hydrolysis (20). To explore differences in collective motions, including the hinge-bending opening and closing mode, we first apply ENM to apo and IB neurolysin crystal structures and extract harmonic vibrational modes.

The correspondence between ENM_{apo} and ENM_{IB} modes is calculated by taking dot products of eigenvector sets for each form considering the ten lowest-frequency modes, and given as an overlap matrix in Fig. 2A. Here, the upper limit of 1 indicates perfect overlap (white) and the lower limit of 0 indicates no overlap (black). Cumulative overlap values are also calculated by taking the sum of the squares over the ten modes (which has a theoretical maximum of 1) and are given on the top and on the right side of the overlap matrix for IB and apo modes, respectively. The overlap matrix shows almost one-to-one correspondence for individual ENM modes, and cumulative overlap values that are close to 1 meaning that each apo mode can be fully described by the cumulative ten modes of inhibitor-bound form, and *vice versa*.

The mode shape of the first ENM_{apo} mode in Fig. 2B and Movie S1 exhibits the expected hinge-bending type motion, which is functionally important and also explicitly observed in other neuropeptidases by crystallography (37–40). The second mode represents a scissor-like motion where domain I (yellow) and II (pale blue) form the blades of the scissor and undergo a rotation as shown in Fig. 2C and Movie S2. The third mode in Fig. S1A and Movie S3 represents a bending type mode, where active site loops on domain I and II move in opposite directions, making more space in the center of the deep channel, while the bottom part of domain I and the active site helix move toward the center of deep channel. In the fourth mode (Fig. S1B), we observe a unique motion where the upper helices are negatively correlated with the beta-sheets and their surrounding loops which are adjacent to the active site in domain II.

In summary, while ENM gives us a clean basis set of motions in neurolysin, it does not reveal significant differences in the collective motions between the apo and inhibitor bound forms. It is important to note that the ENM method is limited by its low-resolution, and largely reflects only the backbone dynamics since the method does not consider side chain information. It is also based on the dynamics surrounding a single structure. Considering this, we next turn to all-atom MD simulations for both apo and inhibitor-bound neurolysin, and perform a number of analyses to extract differences in the dynamics and conformational ensembles that occur upon inhibitor binding.

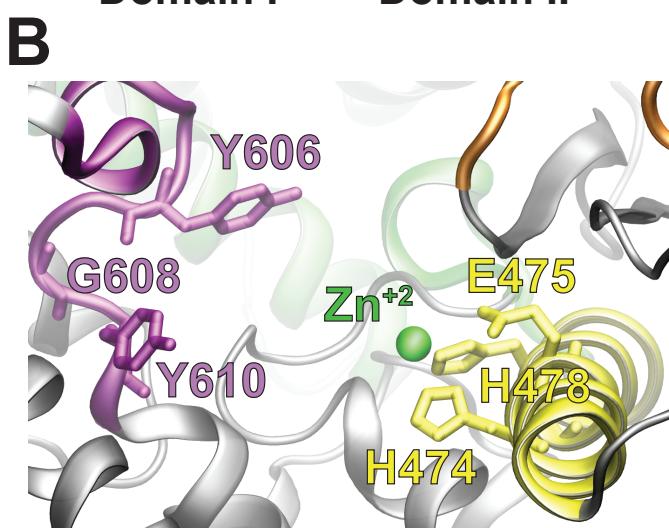
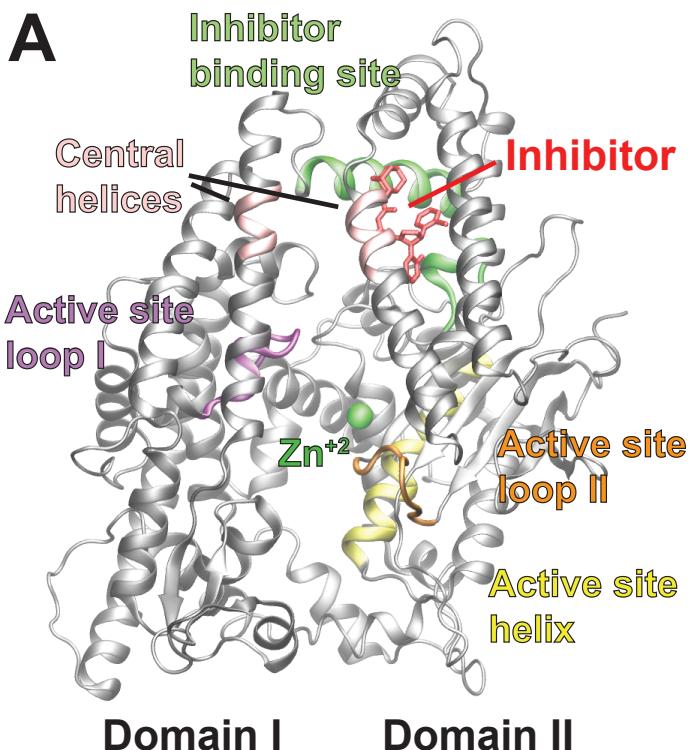


FIGURE 1. (A) Neurolysin structure and (B) active site. Functionally important regions are colored as follows, active site helix: yellow, zinc ion: green sphere, active site loop in domain I: fuchsia, active site loop in domain II: orange, central helices: light pink, and allosteric inhibitor: red licorice representation. Important residues in/around the active site are shown and colored based on the region.

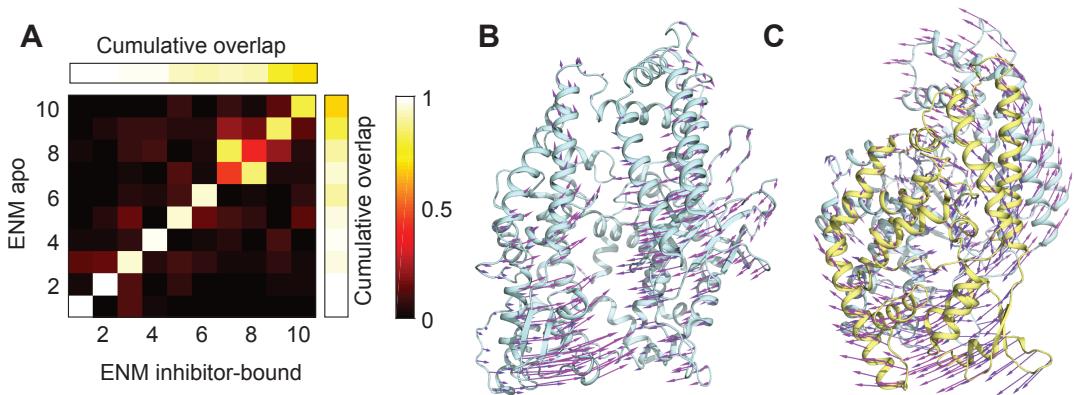


FIGURE 2. (A) Overlap of ENM modes for apo and inhibitor-bound neurolysin. (B) The first ENM mode shows a hinge-bending type opening-closing motion. (C) The second ENM mode (side-view) shows a scissor-like motion.

Essential dynamics from Principal Component Analysis and comparison with Elastic Network Model

Principal component analysis (PCA), (41) is applied separately to apo and inhibitor-bound neurolysin trajectories to extract essential modes (principal components, *PCs*). Ca coordinates of frames, which are aligned to the initial structure of each system, are used to construct the covariance matrix in PCA. Eigenvector sets representing *PCs* are obtained for both apo and inhibitor-bound forms. We again examine the overlap between the eigenvector sets in the same manner as the ENM overlap calculation, and this is shown in Fig. 3A. The first three slow modes are significantly overlapping with a value greater than 0.6.

The overlap matrix indicates that the presence of inhibitor does not greatly affect the first mode dynamics of neurolysin; the overlap coefficient between the apo and inhibitor bound forms was 0.96. The $PC1_{apo}$ mode shape is given in Fig. S2A, showing a hinge-type opening/closing motion that is very similar to the first ENM_{apo} mode. The second mode ($PC2$, overlap=0.77) represents a scissor-like motion (Fig. S2B), which is again similar to the second ENM mode. The third mode ($PC3$, overlap=0.63) represents a bending type motion as shown in Fig. S2C, and here we start to see significant differences between apo and inhibitor bound modes. In $PC3_{apo}$, the active site loops move in opposite directions similar to $ENM3$. On the other hand, $PC3_{IB}$ shows a slightly different motion where the active site loop on domain I moves toward the center of deep channel, not toward the active site loop in domain II. We also find that the inhibitor bound system does not exhibit a significant decay in the amplitude of these motions. $PC1_{IB}$ in fact shows a higher amplitude of motion along mode 1, and modes 2 and 3 show only small decreases in amplitude (Fig. S3).

The first four modes of ENM_{apo} and PCA_{apo} show significant overlap (> 0.6) (Fig. 3B). The 4th mode appears to be common between ENM and PCA_{apo} and similar overlap is observed between $ENM4$ and $PC6_{IB}$ (overlap value=0.66) (Fig. 3C), and no overlap is observed between $PC4_{apo}$ and $PC4_{IB}$ (Fig. 3A). $PC4_{IB}$, in addition to $PC5_{IB}$ form two new low-frequency *PCs*, which are not visited in the apo state, and are not detected in the slowest 10 modes of ENM.

To test the significance of the PCA modes, we divide the datasets into two parts and compute the overlap. The comparison of the modes from the two apo subsets with ENM_{apo} modes shows that the first two ENM modes are reproduced in a one-to-one correspondence in both subsets (Fig. S4 A and B), and the first four ENM modes can be described using a combination of several PCA_{apo} modes. However, the comparison of the two apo subsets indicates that only the first mode is common (Fig. S4 C). For the PCA_{IB} trajectories, the overlaps indicate that the first three ENM modes are reproducible in a one-to-one correspondence (Fig. S4 D and E). The comparison of the two subsets of eight inhibitor bound trajectories shows that three modes are common (Fig. S4 F).

In summary, PCA and ENM results show that both apo and inhibitor-bound neurolysin exhibit three distinct modes, namely hinge-like opening/closing (clamshell-like) ($PC1$), scissor-like ($PC2$) and bending motions ($PC3$). We conclude that binding of the allosteric inhibitor does not significantly affect the slowest collective motions in neurolysin. In the next section we analyze the structural changes in apo and inhibitor-bound simulations in more detail to better understand the effect of inhibitor binding on the ensemble of neurolysin structures observed in solution.

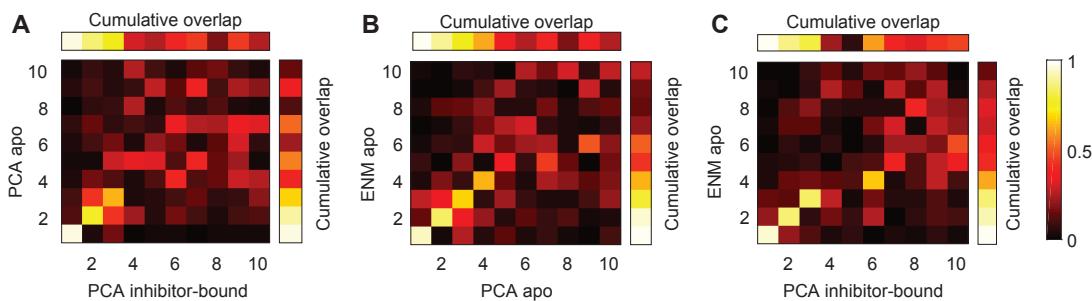


FIGURE 3. Overlap of collective modes (A) PCA_{apo} vs. PCA_{IB} , (B) ENM_{apo} vs. PCA_{apo} and (C) ENM_{apo} vs. PCA_{IB} modes.

Large scale structural differences observed in molecular dynamics simulations

It was previously shown that the inhibitor stabilizes a so-called inactive conformation that is more closed than the apo form (20). In this respect, we explored the large scale structural changes during the dynamics of apo and inhibitor-bound neurolysin based on the radius of gyration (R_g). We observed that lower R_g structures can be characterized by a smaller distance between the centers of mass of the upper part of central helices (residues 104-111 in domain I and residues 166-177 in domain II as colored light pink in Fig. 1A). R_g is plotted against the helix-helix distances as free energy maps for both forms in Fig. 4. The correlation between the distance and R_g can be clearly seen for both neurolysin forms. Although R_g values fall in quite similar ranges, the average R_g and helix-helix distance values for each case (white stars) show that the inhibitor-bound form has a lower average value compared to the average apo. Also, apo frames are mostly populated in one region on the map having higher R_g and distance values. However, inhibitor-bound frames are populated in two different

regions, the more compact of which was about 1.5 kcal/mol less stable in the apo simulations. The R_g and helix-helix distances of the crystal structures are shown with blue dots, and lie outside the sampled range of MD conformations, implying significant differences in the crystal and solution environments.

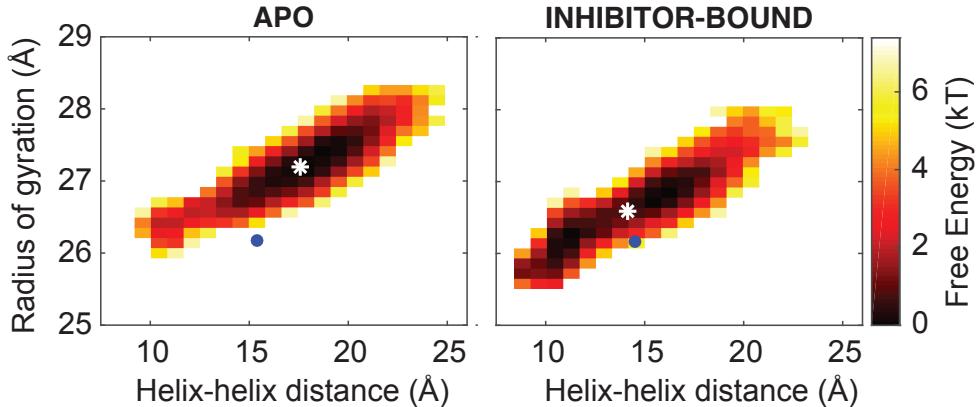


FIGURE 4. Radius of gyration *versus* distances between central helices in domains I and II are shown as a 2D histogram for apo (left) and inhibitor-bound neurolysin (right). The average values for each case are marked on the figures as white stars. The values calculated from the crystal structures are shown with blue dots.

Changes in the contact network upon inhibitor binding

To explore the differences in the bound and apo structural ensembles in greater detail, we next examine the formation and/or loss of amino acid contacts. The closest distance between each residue pair is calculated for a given trajectory frame, and we define a contact between two residues when this distance is less than 4 Å. Contacts are identified over ~800 frames from the 16 MD simulations for both apo and inhibitor-bound forms, and for each contact i we compute the fraction of frames in which that contact is formed (c_i). We then compute the change in c_i upon ligand binding for each contact (Δc_i), which can take values between -1 and 1. To establish a threshold value of significance for the Δc_i values, we divide the 16 apo trajectories into two subsets and compute Δc_i between the subsets; the same is done for the inhibitor bound set. The largest Δc_i observed in either case is 0.35, which we use as a significance threshold. Fig. 5 A shows the apo-IB Δc_i values that have absolute values greater than this cutoff overlaid on the neurolysin molecule, with red indicating that a new contact is observed in the IB ensemble, and blue indicating a contact is lost. Also, these significant residue pairs are listed in Table S1.

Exploring the contact network in more detail, we find that significant contacts are lost in the inhibitor binding site. The helix in the inhibitor binding site (green), which connects domain I and II, loses two contacts with an adjacent loop, also shown in green. The lost contact pairs are **Ser146-Arg491** and Glu138-Thr487. Throughout the text and in Table S1, the residues that have non-bonded interactions with the inhibitor in the crystal structure (data from PDBsum (29, 30)) are shown in bold. **Arg491** and Thr487 in the inhibitor binding site loop also form new contacts with **Ile77** and **Lys648**, respectively. Both contact formation and loss are observed between helices adjacent to the central helix in domain II. The most significant contact loss is

between **Tyr49** and **Glu73**. We should note that **Tyr49** is the only residue that has hydrogen bond with the inhibitor in the crystal structure of neurolysin. These lost contacts appear to accommodate the formation of several intra-domain contacts between central helices, as shown above in Fig. 4 *B*.

Changes are also observed in distant regions from the inhibitor binding site (up to 60 Å) away. Gly477 in the active site helix forms a new contact with Pro505 in the neighboring helix, coinciding with a loss of contact with Ile646. The Gly-Tyr-rich active site loop (residues 599-611 as shown fuchsia in Fig. 5 *A*) has one contact formed upon inhibitor binding between Asp607 and Asp562 in the neighboring helix. However, in the neighborhood of the active site, we do not observe significant changes in contacts involving the active site histidines and the catalytic zinc ion in the contact network.

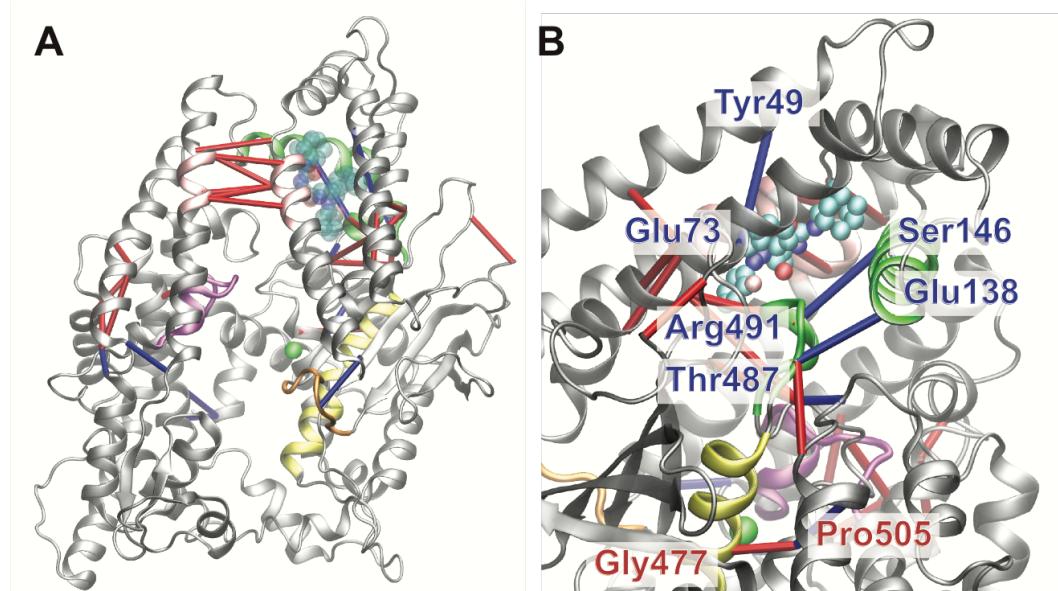


FIGURE 5. (A) Changes in the neurolysin contact network upon inhibitor binding. Formed contacts are shown in red and lost contacts in blue. The inhibitor is shown in transparent sphere representation, and regions are colored as in Fig. 1 *A*. (B) Contact changes in the inhibitor binding site and the active site are shown in detail. The inhibitor is shown in sphere representation, and the zinc ion is shown as a green sphere.

Identifying key residues from Linear Discriminant Analysis

We apply Linear Discriminant Analysis with ITERative procedure (LDA-ITER) (34) to automatically identify the differences between apo and inhibitor-bound neurolysin. All frames from apo and inhibitor-bound trajectories are merged into two sets, and the LDA-ITER method is applied to determine a projection matrix that best discriminates between the two datasets. The columns of this matrix constitute LDA “modes”, the elements of which give displacements for each C α atom. In contrast to principal component analysis (PCA), which identifies collective variables that show the most variance in the entire set, LDA modes are collective variables that show large differences *between* the datasets, and small differences *within* the datasets. The

trajectories are projected on the first LDA mode in Fig. S5, showing that it separates the apo and inhibitor-bound datasets into two non-overlapping distributions. We calculate the displacements in the first LDA mode and show them in Fig. 6, where amino acids that exhibit high displacement are shown as spheres. There are residues having high displacement values in the inhibitor binding site, namely **Tyr76**, **Ile77**, **Met113**, **Arg491**, Asp488 and Thr487, where bold typesetting again signifies residues that interact with the inhibitor. Other significant changes are observed on residues in the active site loops, namely Leu602, Tyr606, Gly608, Gln609, Tyr610, and Glu420. Both LDA and the contact network analysis methods point to similar regions, namely the inhibitor binding site, central helices, active site loops that undergo significant structural changes upon inhibitor binding. **Ile77**, **Thr487** and **Arg491** appear as significant residues from both automated LDA and contact network analysis (Table S1), and are shown in larger spheres in Fig. 6. LDA thus reveals key residues that cause structural differences between the apo and inhibitor-bound neurolysin in an automated way without using any region information, and these results are in consistency with – and also complementary to – our contact network results.

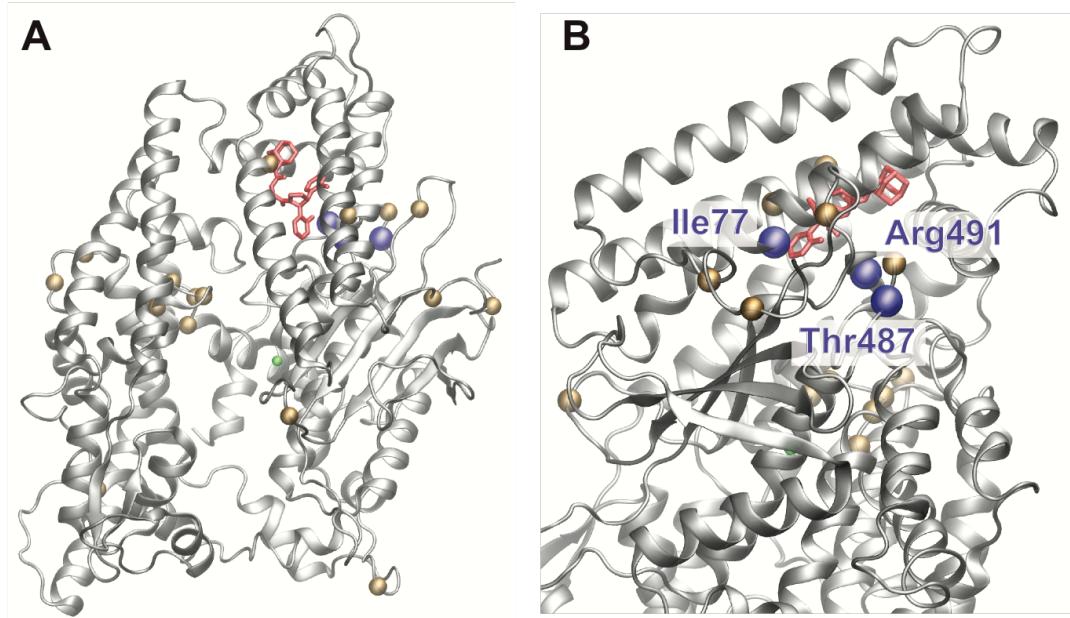


FIGURE 6. Key residues determined from atomic displacements using Linear Discriminant Analysis (LDA) are shown as small spheres (gold). Residues that are common between LDA and the contact network analysis are shown as larger spheres (purple). (A) Whole protein, (B) Binding site detail.

Changes in inter- and intra-domain communication upon inhibitor binding

As the R_g values for the apo and IB frames show different distributions (Fig. 4), we focus on the most populated cluster for apo in Fig. 4 A and the second most populated

cluster for the IB case where the frames have lower Rg as shown in Fig. 4 B. We randomly chose ten frames from each cluster. The chosen apo frames have an RMSD value with the crystal apo structure in a range of 2.6-3.2 Å, and exhibit conformations that are more open compared to the crystal apo structure. On the other hand, the chosen inhibitor-bound frames have lower RMSD values with the crystal apo (1.3-1.9 Å), and have conformations that are more closed.

To investigate how these structural changes can affect inter- and intra-domain communication, we apply ENM to the sets of randomly chosen frames. Surprisingly, the one-to-one correspondence between ENM_{apo} and ENM_{IB} modes (as seen in Fig. 2 A) based on the crystal structures is gone; modes from the new inhibitor-bound frames from MD (ENM_{IB-MD}) have low overlap with the crystal apo structure, showing a significant change in collective motions (Fig. S6). In comparison, apo frames from MD (ENM_{apo-MD}) still show significant overlap with the ENM_{apo} crystal structure modes.

We next compare residue cross-correlations – which are the dot products of eigenvector sets – between the ENM_{apo-MD} modes and the ENM_{IB-MD} modes. Cumulative cross-correlation values from the first ten low-frequency modes are averaged over ten frames for apo and inhibitor-bound cases. For simplicity, only the residues in the eight regions shown in the right panels of Fig. 7 are considered in the cross-correlation calculations. Average regional cross-correlation values are given in the left panel of Fig. 7 A for apo and in Fig. 7 B for the inhibitor-bound. Residue correlations in the apo structure exhibit high intra-domain correlations both in domain II (regions R1 to R4, where R4 has a positive correlation with R5) and in domain I (regions R6, R7 and R8). In the inhibitor-bound case, similar intra-domain correlations are obtained but with lower values (Fig. 7 B). Remarkably, the positive correlation between R2 and R4 disappears in the inhibitor-bound frames, and this is coincident with a new positive inter-domain correlation between R4 and R6. The highly negative correlation between domains I and II also decreases in the inhibitor-bound form (e.g. R1-R6, R3-R6, R4-R7). This implies that allosteric inhibitor binding results in increased inter-domain correlations and decreased intra-domain correlations in neurolysin collective dynamics.

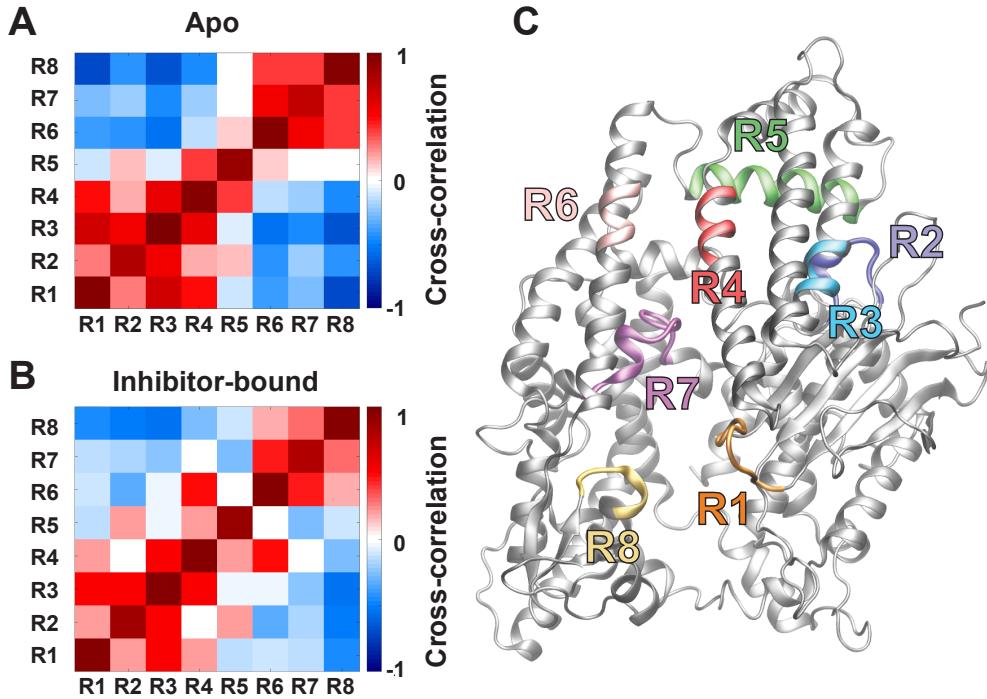


FIGURE 7. Average regional cross-correlation maps between significant neurolysin regions (shown on the right) for ten randomly chosen MD frames from (A) apo, and (B) inhibitor-bound. Residues forming the regions are shown in (C) and are defined as follows, R1: 418-425, R2: 487-494, R3: 76-80, R4: 108-113, R5: 142-149, R6: 167-171, R7: 601-613, and R8: 220-224.

DISCUSSION

Our results from both coarse-grained ENM and PCA of all-atom MD simulations show that inhibitor-bound neurolysin exhibits a hinge-bending type motion that is observed in other neuropeptidases (42). There are also two other common low-frequency modes that are observed in apo and inhibitor-bound forms. This shows how steady the global dynamics of neurolysin are even in the presence of an inhibitor. Although there is very high overlap between the top three ENM_{apo} and PCA_{apo} modes (Fig. 3 B), the mode shapes show some slight differences. This is partly due to the low resolution of Ca -based ENM, and therefore ENM modes may not be a sufficient description of neurolysin dynamics. We also observe significant mixing and shifting of ENM modes when comparing between different simulation conditions, which underscores the importance of using cumulative overlap values to compare mode sets; this was also shown previously in comparisons of PCA and ENM modes for multiple HIV-1 protease (43) and HIV-1 reverse protease (44). Moreover, ENM is structure dependent. As shown in the last section, ENM analysis starting from alternative structures can give very different results. Grant *et al.* (45) have used this feature of ENM as a motivation in their ensemble normal mode analysis method for the exploration of allosteric mechanisms in proteins upon binding. On the other hand, PCA is limited by sampling. The MD simulations conducted here are long enough to observe the principal motions of neurolysin, but are by no means exhaustive (e.g. Fig. S4). ENM, in contrast, can solve for all modes exactly for a given covariance matrix.

We have observed large-scale changes in the distances between central helices in domains I and II together with a significant change in radius of gyration in solution, although only small differences were observed between apo and inhibitor bound crystal structures. The largest conformational change between the crystal structures (20) is in the last three turns of central helix in domain II, which undergoes a small shift about 1 Å. Our simulations suggest the solvated environment yields more closed structures that are observed both in the apo and inhibitor-bound forms (Fig. 4). This underscores the limitations of single structures from crystallography in explaining the effects of inhibitor binding.

The contact network analysis reveals long-range allosteric changes in neurolysin and pinpoints some key residues especially in the inhibitor binding site. In our simulations with the inhibitor, we observe a lost contact between **Tyr49** in the inhibitor binding site and **Glu73** on an adjacent helix in domain II. In the inhibitor-bound crystal structure, **Tyr49** is in a coordinated motion with **Ile77**, and this provides enough space for **Glu73** to alter its side chain conformation (20). Therefore, the contact loss between **Glu73** and **Tyr49** might be an expected result in the inhibitor-bound form. Loss of contacts in the inhibitor binding site is in line with previous results on the trypsin-benzamidine system (46). Although there is not any experimental data on the effect of Pro505 on neurolysin activity, our contact network analysis shows that Gly477 in the active site helix makes a new contact with Pro505, which loses its contact with Ile646 in the domain II. Thus, Pro505 might be a key residue in allosteric regulation of neurolysin.

Similar long-range changes upon binding in neurolysin were observed around the inhibitor binding site and active site loop using the Linear Discriminant Analysis (LDA) machine learning method. LDA-ITER is a variant of LDA first applied to biomolecular systems by Sakuraba and Kono (34). LDA is typically used to classify data into categories after it is trained with a labeled dataset, but LDA collective variables are also useful to identify differences between structural ensembles. To our knowledge, our work is the second study where this method is applied, and we find that the method is able to reveal key molecular differences that are consistent with the contact network analysis. However, there are significant residues on the active site loop (Leu602, Tyr606, Gly608, Gln609 and Tyr610) that are revealed as key residues from LDA but not in contact network analysis. The role of the active site loop for peptide selection has been shown both experimentally and computationally (47) for a closely related neuropeptidase thimet oligopeptidase (TOP), which can also recognize and cleave neuropeptides and also has a very similar active site loop compared to neurolysin, where Tyr606 is the only different residue in the sequence (48). They have also observed that the hinge-bending mode in TOP assists the motion of the flexible loop towards the peptide in order to form a hydrogen bond. Mutations on another residue (Gly608) on the active site loop of neurolysin have been shown to affect loop flexibility as well as hydrolytic activity (49). Moreover, Tyr610 in neurolysin (33) and its corresponding residue in another closely related neuropeptidase (Tyr611 in dipeptidyl carboxypeptidase) (38), were experimentally shown to be important residues in peptide fixation. That we predict residues of known importance gives us confidence in the significance of the set of key residues predicted by LDA. This motivates experimental testing of new key residues (starred in Table S1) to determine their possible role in neurolysin function.

The persistence of large scale protein vibrational modes implies that neurolysin exhibits a “violin” model allosteric mechanism (5). The violin model is based on the long-range effect of molecular binding on a protein’s vibration with an analogy to the effects on a violin’s pitch when there is a slight finger movement of a violin player on the fingerboard. Contrary to the “domino” model, which is the propagation of the signal from an allosteric site to an active site that occurs in a sequential set of events and results a “domino” effect, the “violin” allosteric effect does not exhibit a chain-like network of amino acid interactions as observed in PDZ domain family proteins and G protein-coupled receptors (50, 51). Thus, in the “violin” model, the persistent vibration of the protein is the essential requirement (5, 52). We observe that although the low-frequency modes both in apo and inhibitor-bound forms are well preserved, there are changes to higher-frequency modes as well as the communication between protein domains during collective motion.

Similar observation in the communication networks of heterotrimeric G proteins, hemoglobin and caspase 7 structures upon binding has been revealed from ensemble normal mode analysis using ENM by Yao *et al.* (45). They calculated residue cross-correlations from several ENM methods with different resolutions and applied PCA to the cross-correlation data. They showed that they can obtain a better prediction of residue couplings that explain the allosteric mechanism observed in MD simulations when a all-atom description of intra-protein interactions are considered in ENM. In our study, application of C α -based ENM to distinct inhibitor-bound conformations with lower R_g values yielded different collective dynamics than the apo frames from the most populated cluster. This revealed changes in residue cross-correlations and in the inter- and intra-domain communications in inhibitor-bound neurolysin. It is possible that higher resolution ENM such as an all-heavy-atom-based potential (AAENM, (45)) and mixed coarse-grained ENM (53) could better explain the dynamical differences between apo and inhibitor-bound neurolysin.

CONCLUSIONS

The absence of conformational change between the apo and inhibitor-bound neurolysin crystal structures raised many questions regarding collective dynamics: Is there a significant difference between the structural ensemble observed in solution versus the crystal structure? Does the inhibitor specifically block the hinge-bending type motion that provides the opening/closing of neurolysin deep channel for ligand entrance? Do apo and inhibitor-bound neurolysin exhibit similar or very distinct collective dynamics? How does the allosteric inhibitor affect the overall enzyme activity? In this work, we looked for answers to these questions using multi-scale computational methods namely CHARMM MD simulations and Elastic Network Model calculations.

Our work provides an extensive analysis of apo and allosteric inhibitor bound neurolysin via molecular simulations, elastic network modeling and two machine-learning methods (PCA and LDA). Elastic network model calculations, which are in conformity with the PCA results, showed the dynamical similarities between two different neurolysin forms, which show a functionally important hinge-like opening/closing motion, a scissor-like motion, and a bending motion including the coupled motion of the active site loops in domains I and II. Changes in

communication between important neurolysin regions were observed using cross-correlation maps from ENM of characteristic frames from apo and inhibitor bound simulations. The allosteric inhibitor binds to the top right of the neurolysin deep channel and blocks the communication between central helix and the loop at the inhibitor binding site in domain II. At the same time, the communication between central helix in domain II and its complimentary helix in domain I is increased. We find good agreement between the results from contact network analysis and the LDA machine learning method. Remarkably, LDA is very successful to define the key regions that undergo significant changes upon inhibitor binding, and we use these to highlight the changing dynamical coupling in neurolysin using ENM cross correlations. While this study has shed light on the collective dynamics of neurolysin in the absence and presence of the allosteric inhibitor, the effect of neuropeptide binding on the overall dynamics of neurolysin still needs to be investigated. This would help to illuminate the effects of known inhibitors, and help to design new compounds to modulate neurolysin function.

Acknowledgment:

This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research.

REFERENCES

1. Monod, J., J. Wyman, and J.P. Changeux. 1965. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* 12: 88–118.
2. Fonnaneck, M.S., L. Ma, and Q. Cui. 2006. Reconciling the “old” and “new” views of protein allostery: A molecular simulation study of chemotaxis Y protein (CheY). *Proteins Struct. Funct. Genet.* 63: 846–867.
3. Popovych, N., S. Sun, R.H. Ebright, and C.G. Kalodimos. 2006. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13: 831–838.
4. McLeish, T.C.B., M.J. Cann, and T.L. Rodgers. 2015. Dynamic Transmission of Protein Allostery without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* 109: 1240–1250.
5. Kornev, A.P., and S.S. Taylor. 2015. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* 40: 628–647.
6. Cooper, a, and D.T.F. Dryden. 1984. European Biophysics A plausible model. *Eur Biophys J.* 11: 103–109.
7. Swain, J.F., and L.M. Giersch. 2006. The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* 16: 102–108.
8. Kern, D., and E.R.P. Zuiderweg. 2003. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* 13: 748–757.
9. Tsai, C.J., A. del Sol, and R. Nussinov. 2008. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J. Mol. Biol.* 378: 1–11.
10. Rawlings, N.D., and A.J. Barrett. 1995. Evolutionary families of metallopeptidases. *Methods Enzymol.* 248: 183–228.
11. Brown, C.K., K. Madauss, W. Lian, M.R. Beck, W.D. Tolbert, and D.W. Rodgers. 2001. Structure of neurolysin reveals a deep channel that limits substrate access. *PNAS.* 98: 3127–3132.
12. Rashid M, Wangler NJ, Yang L, Shah K, Arumugam TV, Abbruscato TJ, K.V. 2014. Functional up-regulation of endopeptidase neurolysin during post-acute and early recovery phases of experimental stroke in mouse brain. *Neurochem. J. O F.* 129(1): 179–189.
13. Checler, F. 2014. Experimental stroke: neurolysin back on stage. *J. Neurochem.* 129: 1–3.
14. Wangler, N.J., S. Jayaraman, R. Zhu, Y. Mechref, T.J. Abbruscato, U. Bickel, and V.T. Karamyan. 2016. Preparation and preliminary characterization of recombinant neurolysin for in vivo studies. *J. Biotech.* 234: 105–115.
15. Jiracek, J., A. Yiotakis, B. Vincent, A. Lecoq, A. Nicolaou, F. Checler, and V. Dive. 1995. Development of highly potent and selective phosphinic peptide inhibitors of zinc endopeptidase 24-15 using combinatorial chemistry. *J. Biol. Chem.* 270: 21701–21706.
16. Yiotakis, A., B. Vincent, and V. Dive. 1996. Development of the First Potent and Selective Inhibitor of the Zinc Endopeptidase Neurolysin Using a Systematic Approach Based on Combinatorial Chemistry of Phosphinic Peptides. *J. Biol. Chem.* 271: 19606–19611.
17. Vincent, B., V. Dive, A. Yiotakis, C. Smadja, R. Maldonado, J.-P. Vincent, and F.F. Checler. 1995. Phosphorus-containing peptides as mixed inhibitors of endopeptidase 3.4.24.15 and 3.4.24.16: effect on neuropeptidyl degradation in vitro and in vivo. *Br. J. Pharmacol.* 115: 1053–1063.
18. Meier, P., M. Maillard, and M. Burnier. 2005. The future of angiotensin II inhibition in cardiovascular medicine. *Curr. Drug Targets - Cardiovasc. Haematol. Disord.* 5: 15–30.

19. Adekoya, O.A., S. Sjøli, Y. Wuxiuer, I. Bilto, S.M. Marques, M.A. Santos, E. Nuti, G. Cercignani, A. Rossello, J.O. Winberg, and I. Sylte. 2015. Inhibition of pseudolysin and thermolysin by hydroxamate-based MMP inhibitors. *Eur. J. Med. Chem.* 89: 340–348.
20. Hines, C.S., K. Ray, J.J. Schmidt, F. Xiong, R.W. Feenstra, M. Pras-raves, J.P. De Moes, J.H.M. Lange, M. Melikishvili, M.G. Fried, P. Mortenson, M. Charlton, Y. Patel, S.M. Courtney, C.G. Kruse, and D.W. Rodgers. 2014. Allosteric Inhibition of the Neuropeptidase Neurolysin. *J. Biol. Chem.* 289: 35605–35619.
21. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4: 187–217.
22. MacKerel Jr., A.D., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*. Schleyer P v. R , Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF, III, Schreiner PR., editor. John Wiley & Sons, Chichester, pp. 271–277.
23. Brooks, B.R., C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus. 2009. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30: 1545–1614.
24. Doruker, P., A.R. Atilgan, and I. Bahar. 2000. Dynamics of proteins predicted by molecular simulations and analytical approaches: Application to alpha-amylase inhibitor. *Proteins Struct. Funct. Genet.* 40: 512–524.
25. Atilgan, A.R., S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80: 505–515.
26. Huang, J., and A. MacKerell. 2013. CHARMM36 all atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* 34(25): 2135–2145.
27. Vanommeslaeghe, K., and A.D. MacKerell. 2012. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* 52(12): 3144–3154.
28. Vanommeslaeghe, K., E.P. Raman, and A.D. MacKerell. 2012. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* 52(12): 3155–3168.
29. Laskowski, R.A., E.G. Hutchinson, A.D. Michie, A.C. Wallace, M.L. Jones, and J.M. Thornton. 1997. PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* 22(12): 488–490.
30. Laskowski, R.A., V. V. Chistyakov, and J.M. Thornton. 2005. PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* 33: D266–D268.
31. Eastman, P., J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.-P. Wang, A.C. Simmonett, M.P. Harrigan, B.R. Brooks, and V.S. Pande. 2016. OpenMM 7: Rapid Development of High Performance

- Algorithms for Molecular Dynamics. *bioRxiv*. 091801; doi: <https://doi.org/10.1101/091801>.
- 32. Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 7: 179–188.
 - 33. Rao, C. 1948. The Utilization of Multiple Measurements in Problems of Biological Classification. *J. R. Stat. Soc. Ser. B*. 10: 159–203.
 - 34. Sakuraba, S., and H. Kono. 2016. Spotting the difference in molecular dynamics simulations of biomolecules. *J. Chem. Phys.* 145: 074116; doi: <http://dx.doi.org/10.1063/1.4961227>.
 - 35. Jongeneel, C.V., J. Bouvier, and A. Bairoch. 1989. A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett.* 242: 211–214.
 - 36. Kadonosono, T., M. Kato-Murai, and M. Ueda. 2008. Alteration of substrate specificity of rat neurolysin from matrix metalloproteinase-2/9-type to -3-type specificity by comprehensive mutation. *Protein Eng. Des. Sel.* 21: 507–513.
 - 37. Natesh, R., S.L.U. Schwager, E.D. Sturrock, and K.R. Acharya. 2003. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature*. 421: 551–554.
 - 38. Comellas-Bigler, M., R. Lang, W. Bode, and K. Maskos. 2005. Crystal structure of the *E. coli* dipeptidyl carboxypeptidase Dcp: Further indication of a ligand-dependant hinge movement mechanism. *J. Mol. Biol.* 349: 99–112.
 - 39. Natesh, R., S.L.U. Schwager, H.R. Evans, E.D. Sturrock, and K.R. Acharya. 2004. Structural Details on the Binding of Antihypertensive Drugs Captopril and Enalaprilat to Human Testicular Angiotensin I-Converting Enzyme. *Biochemistry*. 43(27): 8718–8724.
 - 40. Towler, P., B. Staker, S.G. Prasad, S. Menon, J. Tang, T. Parsons, D. Ryan, M. Fisher, D. Williams, N.A. Dales, M.A. Patane, and M.W. Pantoliano. 2004. Ace2 X-rays structure reveals a large hinge-bending motion important for inhibitor binding and catalysis. *J. Biol. Chem.* 279: 17996–18007.
 - 41. Amadei, A., A.B. Linssen, and H.J. Berendsen. 1993. Essential dynamics of proteins. *Proteins*. 17: 412–425.
 - 42. Carvalho, H.F., A.C.A. Roque, O. Iranzo, and R.J.F. Branco. 2015. Comparison of the Internal Dynamics of Metalloproteases Provides New Insights on Their Function and Evolution. 10(9):e013: 1–22.
 - 43. Yang, L., G. Song, A. Carriquiry, and R.L. Jernigan. 2008. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure*. 16: 321–30.
 - 44. Bahar, I., M.H. Cheng, J.Y. Lee, C. Kaya, and S. Zhang. 2015. Biophysical Perspective Substrate Interactions. *Biophys. J.* 109: 1101–1109.
 - 45. Yao, X.Q., L. Skjærven, and B.J. Grant. 2016. Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis. *J. Phys. Chem. B*. 120: 8276–8288.
 - 46. Dickson, A., and S.D. Lotz. 2017. Multiple Ligand Unbinding Pathways and Ligand- Induced Destabilization Revealed by WExplore. *Biophys. J.* 112: 620–629.
 - 47. Bruce, L.A., J.A. Sigman, D. Randall, S. Rodriguez, M.M. Song, Y. Dai, D.E. Elmore, A. Pabon, M.J. Glucksman, and A.J. Wolfson. 2008. Hydrogen bond residue positioning in the 599–611 loop of thimet oligopeptidase is required for substrate selection. *FEBS J.* 275: 5607–5617.

48. Lim, E.J., S. Sampath, J. Coll-rodriguez, J. Schmidt, K. Ray, and D.W. Rodgers. 2007. Swapping the Substrate Specificities of the Neuropeptidases Neurolysin and Thimet Oligopeptidase. *J. Biol. Chem.* 282: 9722–9732.
49. Machado, M.F., V. Rioli, F.M. Dalio, L.M. Castro, M.A. Juliano, I.L. Tersariol, E.S. Ferro, L. Juliano, and V. Oliveira. 2007. The role of Tyr605 and Ala607 of thimet oligopeptidase and Tyr606 and Gly608 of neurolysin in substrate hydrolysis and inhibitor binding. *Biochem. J.* 404: 279–288.
50. Lockless, S.W., and R. Ranganathan. 1999. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science*. 286: 295–299.
51. Süel, G.M., S.W. Lockless, M.A. Wall, and R. Ranganathan. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10: 59–69.
52. Gourinchas, G., S. Etzl, C. Göbl, U. Vide, T. Madl, and A. Winkler. 2017. Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Sci. Adv.* 3(3): e1602498; DOI: 10.1126/sciadv.1602498.
53. Kurkcuoglu, O., R.L. Jernigan, and P. Doruker. 2005. Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model. *QSAR Comb. Sci.* 24: 443–448.