LA-UR-05-5164

Allostery in a Coarse-Grained Model of Protein Dynamics

Dengming Ming

Computer and Computational Sciences Division,

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Michael E. Wall*

Computer and Computational Sciences and Bioscience Divisions,

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

(Dated: March 28, 2006)

Abstract

We propose a criterion for optimal parameter selection in coarse-grained models of proteins, and develop a refined elastic network model (ENM) of bovine trypsinogen. The unimodal density-of-states distribution of the trypsinogen ENM disagrees with the bimodal distribution obtained from an all-atom model; however, the bimodal distribution is recovered by strengthening interactions between atoms that are backbone neighbors. We use the backbone-enhanced model to analyze allosteric mechanisms of trypsinogen, and find relatively strong communication between the regulatory and active sites.

A major challenge of molecular biology is to understand regulatory mechanisms in large protein complexes that are abundant in multi-celluluar organisms. To make simulation of such complexes computationally feasible, coarse-grained models have been developed, in which a subset of the atoms in the complex are used to simulate the large-scale motions. However, principled methods to quantify and optimize the accuracy of coarse-grained models are currently lacking.

In one common coarse-graining method, an all-atom model is simplified by considering effective interactions among a subset of the atoms (e.g., just the alpha-carbons). The usual criterion for model accuracy is the ability of a model to reproduce atomic mean-squared displacements (MSDs). However, MSDs are just one aspect of protein dynamics – a stricter criterion for the accuracy of a coarse-grained model is the similarity between the configurational distributions of the selected atoms in the coarse-grained and all-atom models. Such a criterion is also biologically relevant, in part because the conformational distribution is a key determinant of protein activity [1].

One useful measure of the difference between conformational distributions is the Kullback-Leibler divergence $D_{\mathbf{x}}$ (see definition below) [2, 3]. Recently, an analytic expression for $D_{\mathbf{x}}$ was obtained for harmonic vibrations of a protein-ligand complex both with and without a protein-ligand interaction [3]. Here we show how an equivalent expression may be applied to refine a coarse-grained model of protein dynamics. To use the expression for $D_{\mathbf{x}}$ requires the marginal probability distribution of a subset of the atoms in a protein, which we calculate in the harmonic approximation. We then apply the equations to refine an anisotropic elastic network model (ENM) [4] of trypsinogen dynamics with respect to an all-atom model calculated using CHARMM [5]. The unimodal density-of-states distribution of the ENM disagrees with the bimodal distribution obtained from the all-atom model; however, the bimodal distribution is recovered by strengthening interactions between atoms that are backbone neighbors. Finally, the backbone-enhanced elastic network model (BENM) is used to analyze allosteric mechanisms of trypsinogen, revealing relatively strong communication between the regulatory and active sites.

Let $P(\mathbf{x})$ be the probability distribution of the $3N$ atomic coordinates $\mathbf{x} = (x_1, y_1, z_1, \ldots, x_N, y_N, z_N)$ of a molecular model in the harmonic approximation. Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_1$ is the $3N_1$ coordinates of a subset of atoms of interest, and $\mathbf{x}_2$ is the $3N_2$ coordinates of the remaining atoms. We are interested in calculating the marginal

distribution $P(\mathbf{x}_1)$:

$$P(\mathbf{x}_1) = \int d^{3N_2}\mathbf{x}_1 \, P(\mathbf{x}_1, \mathbf{x}_2). \tag{1}$$

We now calculate $P(\mathbf{x}_1)$ in a model of molecular vibrations. Consider a harmonic approximation to the potential energy function $U(\mathbf{x})$, where $\mathbf{x}$ is the deviation from an equilibrium conformation $\mathbf{x}_0$:

$$U(\mathbf{x} + \mathbf{x}_0) \approx U(\mathbf{x}_0) + \frac{1}{2}\mathbf{x}^\dagger \mathbf{H}\, \mathbf{x}. \tag{2}$$

The matrix $\mathbf{H}$ is the Hessian of U evaluated at $\mathbf{x}_0$: $H_{ij}|_{\mathbf{x}_0} = \partial^2 U/\partial x_i \partial x_j|_{\mathbf{x}_0}$. We assume a Boltzmann distribution for $P(\mathbf{x})$, and ignore solvent and pressure effects:

$$P(\mathbf{x}) = Z^{-1} e^{\frac{-\mathbf{x}^\dagger \mathbf{H} \mathbf{x}}{2k_B T}} = (2\pi k_B T)^{-3N/2} e^{\frac{-\left|\mathbf{\Omega} \mathbf{V}^\dagger \mathbf{x}\right|^2}{2k_B T}} \prod_{i=1}^{3N} \omega_i, \tag{3}$$

where Z is the partition function, k_B is Boltzmann's constant, T is the temperature, the elements of the matrix $|\mathbf{\Omega}|^2 = \mathrm{diag}(\omega_1^2, \ldots, \omega_{3N}^2)$ are the eigenvalues of $\mathbf{H}$, and the columns of the matrix $\mathbf{V}$ are the eigenvectors of $\mathbf{H}$. To calculate $P(\mathbf{x}_1)$ we define the submatrices $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{G}$ as follows:

$$\mathbf{H}\,\mathbf{x} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{G} \\ \mathbf{G}^\dagger & \mathbf{H}_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \mathbf{x}_1 \; + \; \mathbf{G}\mathbf{x}_2 \\ \mathbf{G}^\dagger \mathbf{x}_1 \; + \; \mathbf{H}_2 \mathbf{x}_2 \end{pmatrix}. \tag{4}$$

$\mathbf{H}_1$ couples coordinates from $\mathbf{x}_1$; $\mathbf{H}_2$ couples coordinates from $\mathbf{x}_2$; and $\mathbf{G}$ couples coordinates between $\mathbf{x}_1$ and $\mathbf{x}_2$. Eq. (3) now can be expressed as

$$P(\mathbf{x}) = Z^{-1} e^{\frac{-\mathbf{x}^\dagger \mathbf{H} \mathbf{x}}{2k_B T}} = (2\pi k_B T)^{-3N/2} e^{\frac{-\left|\bar{\mathbf{\Omega}} \bar{\mathbf{V}}^\dagger \mathbf{x}_1\right|^2 - \left|\mathbf{\Lambda} \mathbf{U}^\dagger \mathbf{x}_2 + \mathbf{\Lambda}^{-1} \mathbf{U}^\dagger \mathbf{G}^\dagger \mathbf{x}_1\right|^2}{2k_B T}} \prod_{i=1}^{3N} \omega_i, \tag{5}$$

where the diagonal elements of the matrix $|\mathbf{\Lambda}|^2 = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{3N_1}^2)$ and the columns of the matrix $\mathbf{U}$ are the eigenvalues and eigenvectors of $\mathbf{H}_2$, and the diagonal elements of the matrix $\left|\bar{\mathbf{\Omega}}\right|^2 = \mathrm{diag}(\bar{\omega}_1^2, \ldots, \bar{\omega}_{3N_1}^2)$ and the columns of the matrix $\bar{\mathbf{V}}$ are the eigenvalues and eigenvectors of a matrix $\bar{\mathbf{H}}$ defined as

$$\bar{\mathbf{H}} = \mathbf{H}_1 - \mathbf{G}\mathbf{H}_2^{-1}\mathbf{G}^\dagger = \bar{\mathbf{V}} \left|\bar{\mathbf{\Omega}}\right|^2 \bar{\mathbf{V}}^\dagger. \tag{6}$$

Eq. (6) is equivalent to an equation independently derived to study local vibrations in the nucleotide-binding pockets of myosin and kinesin [6]. Performing the integral in Eq. (1) leads to the desired equation for $P(\mathbf{x}_1)$:

$$P(\mathbf{x}_1) = (2\pi k_B T)^{-3N_1/2} e^{\frac{-\left|\bar{\mathbf{\Omega}} \bar{\mathbf{V}}^\dagger \mathbf{x}_1\right|^2}{2k_B T}} \prod_{i=1}^{3N_1} \bar{\omega}_i. \tag{7}$$

Now consider the problem of optimal selection of the parameters Γ of a coarse-grained model of protein dynamics. Let $\mathbf{x}_\alpha$ be the coordinates of the N_α alpha-carbons in an an all-atom model, and $\mathbf{x}_\alpha^{(\Gamma)}$ be the same coordinates in the coarse-grained model. We define the optimal coarse-grained model as the one for which the Kullback-Leibler divergence between $P^{(\Gamma)}(\mathbf{x}_\alpha)$ and $P(\mathbf{x}_\alpha)$ is minimal, *i.e.*, for which Γ is chosen such that

$$D_{\mathbf{x}_\alpha}^{(\Gamma)} = \int d^{3N_\alpha}\mathbf{x}_\alpha \, P^{(\Gamma)}(\mathbf{x}_\alpha) \ln \frac{P^{(\Gamma)}(\mathbf{x}_\alpha)}{P(\mathbf{x}_\alpha)} \tag{8}$$

is minimal. We previously calculated an analytic expression for equations like Eq. (8) when $P(\mathbf{x}_\alpha)$ and $P^{(\Gamma)}(\mathbf{x}_\alpha)$ are both governed by harmonic vibrations [3]:

$$D_{\mathbf{x}_\alpha}^{(\Gamma)} = \sum_{i=1}^{3N_\alpha} \left(\ln \frac{\omega_i^{(\Gamma)}}{\bar{\omega}_i} + \frac{1}{2k_BT}\bar{\omega}_i^2 \left| \bar{\mathbf{v}}_i^\dagger \Delta\mathbf{x}_{\alpha,0} \right|^2 + \frac{1}{2} \sum_{j=1}^{3N_\alpha} \frac{\bar{\omega}_j^2}{{\omega_i^{(\Gamma)}}^2} \left| {\mathbf{v}_i^{(\Gamma)}}^\dagger \bar{\mathbf{v}}_j \right|^2 - \frac{1}{2} \right). \tag{9}$$

In Eq. (9), ${\omega_i^{(\Gamma)}}^2$ and $\mathbf{v}_i^{(\Gamma)}$ are the eigenvalue and eigenvector of mode i of the coarse-grained model; $\bar{\omega}_i^2$ and $\bar{\mathbf{v}}_i$ are the i^{th} eigenvalue and eigenvector of the matrix $\bar{\mathbf{H}}$ calculated for the alpha-carbon atoms of the all-atom model (Eq. (6)), and $\Delta\mathbf{x}_{\alpha,0} = \mathbf{x}_{\alpha,0}^{(\Gamma)} - \mathbf{x}_{\alpha,0}$ is the difference between the equilibrium coordinates of the coarse-grained and all-atom models. An optimal coarse-grained model of harmonic vibrations is thus one with parameters Γ such that $D_{\mathbf{x}_\alpha}^{(\Gamma)}$ calculated using Eq. (9) is minimal.

In the ENM [4], interacting alpha-carbon atoms are connected by springs aligned with the direction of atomic separation. Following the Tirion model of harmonic vibrations [7], each spring has the same force constant γ. For a given interaction network, the eigenvectors $\mathbf{v}_i^{(\gamma)}$ are independent of γ, and each eigenvalue ${\omega_i^{(\gamma)}}^2$ is proportional to γ. The value of γ at which $D_{\mathbf{x}_\alpha}^{(\gamma)}$ is minimal may be calculated using Eq. (9):

$$\gamma = \frac{1}{3N_\alpha} \sum_{i=1}^{3N_\alpha} \sum_{j=1}^{3N_\alpha} \frac{\bar{\omega}_j^2}{a_i^2} \left| {\mathbf{v}_i^{(\gamma)}}^\dagger \bar{\mathbf{v}}_j \right|^2. \tag{10}$$

The proportionality constants $a_i^2 = {\omega_i^{(\gamma)}}^2/\gamma$ are determined from the eigenvalue spectrum calculated using an arbitrary value of γ (because the eigenvalues $\omega_i^{(\gamma)2}$ are proportional to γ, the constants a_i^2 are independent of γ). It is easily shown that the third and fourth terms of Eq. (9) cancel when γ assumes the value given by Eq. (10).

The interaction network in an elastic network model is generated by enabling interactions only between pairs of atoms separated by a distance less than or equal to a cutoff distance r_c.

To optimize the model, the value of r_c for which $D^{(\gamma)}_{\mathbf{x}_\alpha}$ is minimal is numerically estimated, using values of γ from Eq. (10).

As a test case for optimization, we developed a coarse-grained model of bovine trypsinogen from an all-atom model (223 amino acids obtained from PDB entry 4TPI [8]). CHARMM was used for all-atom simulations using the CHARMM22 force field with default parameter values. HBUILD was used to generate hydrogen positions, and the energy was initially minimized using 2000 steps of relaxation by the adopted basis Newton-Raphson method, gradually reducing the weight of a harmonic restraint to the crystal-structure coordinates. The final minimized structure was obtained through vacuum minimization until a gradient of 10^{-7} Kcal/mol Å was achieved, and the Hessian $\mathbf{H}$ was calculated in CHARMM. The coordinates of the elastic network model were taken from the alpha-carbon coordinates of the minimized all-atom model.

The alpha-carbon vibrations of the all-atom model were calculated by diagonalizing $\bar{\mathbf{H}}$ from Eq. (6). Interestingly, the distribution of the density-of-states for the vibrations is bimodal (Fig. 1) with 2/3 of the frequencies in the low-frequency spectrum and 1/3 of the frequencies in the high-frequency spectrum. Calculation of the density-of-states distribution from other globular proteins yields bimodal patterns with a similar 2:1 ratio between the numbers of low- and high-frequency modes (unpublished results).

The best elastic network model of trypsinogen was obtained using a cutoff distance r_c of approximately 7.75 Å, for which the optimal value of γ is 53.4 Kcal/mol Å^2, yielding a value of $D_{\mathbf{x}_\alpha} = 312.9$ in a sharp minimum with respect to r_c. The density-of-states distribution for the elastic network model is unimodal, unlike that for the all-atom model (Fig. 1).

Although the ENM treats all alpha-carbon pairs equally, the distribution of distances between successive alpha-carbons along the protein backbone is known to be tightly centered about 3.8 Å. In addition, two of the six alpha-carbons nearest to a typical alpha-carbon are backbone neighbors, which might explain why 1/3 of the CHARMM-derived modes have significantly higher frequencies than the others. We therefore wondered whether the ENM might be improved by enhancing interactions between backbone neighbors.

Indeed, a more accurate coarse-grained model is obtained by using a force constant enhanced by a factor of ϵ for interactions between alpha-carbons that are neighbors on the backbone. Minimization of $D_{\mathbf{x}_\alpha}$ for such a backbone-enhanced elastic network model (BENM) with respect to ϵ and r_c subject to Eq. (10) yields a model with $\epsilon = 42$, $r_c = 10.5$ Å,

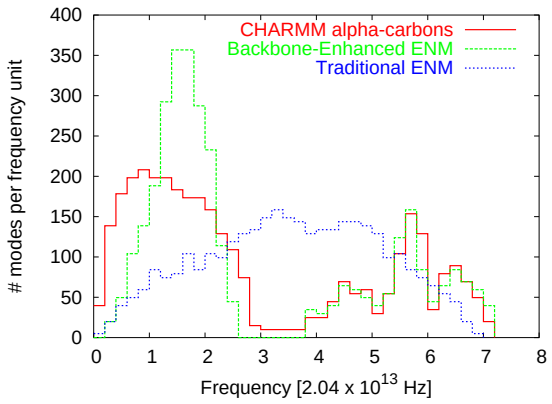

FIG. 1: Density-of-states distribution for all-atom and elastic network models of trypsinogen. Frequency units are $(\text{Kcal/mol}\,\text{Å}^2\,m_p)^{1/2} = 2.04 \times 10^{13}$ Hz, where m_p is the proton mass. Densities were estimated by counting the number of modes in bins of width 0.2, and normalizing the integral to 663, which is the total number of non-zero modes. The ENM (*dotted blue*) does not reproduce the bimodal distribution from the all-atom model (*solid red*); however, the BENM recovers the bimodal distribution (*dashed green*).

and $\gamma = 4.26$ Kcal/mol Å^2, resulting in a much lower value $D_{\mathbf{x}_\alpha} = 102.3$. The density-of-states distribution for this model agrees quite well with that of the all-atom model (Fig. 1), especially considering that the model is optimized with respect to $D_{\mathbf{x}_\alpha}$, which does not directly involve the density-of-states distribution. The agreement is especially good for the high-frequency modes, suggesting that a uniform force constant is a reasonable approximation for interactions between alpha-carbons that are backbone neighbors. Furthermore, the overlap $\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{v}_i^{(\gamma)\dagger}\bar{\mathbf{v}}_j|^2/N$ for the 223 highest-frequency modes is 0.99, indicating that the spaces of the high-frequency eigenvectors are nearly identical between the BENM and all-atom models. In contrast, the low-frequency distribution of BENM states is narrower than that of the all-atom model, indicating that a uniform force constant is a poorer approximation for interactions between alpha-carbons that are not backbone neighbors.

Both the BENM and the ENM yield patterns of alpha-carbon MSDs that are similar to that of the all-atom model (Fig. 2). Because there are fewer low-frequency BENM modes than low-frequency CHARMM modes (Fig. 1), the BENM MSDs are consistently smaller than the CHARMM MSDs; however, the BENM MSDs may be improved by selecting $\gamma = 1.2$ Kcal/mol Å^2 (Fig. 2). These improved MSDs come at the cost of a higher value of $D_{\mathbf{x}_\alpha} = 528.4$, and a change in the frequency scale by a factor $(1.2/4.3)^{1/2} = 0.53$,

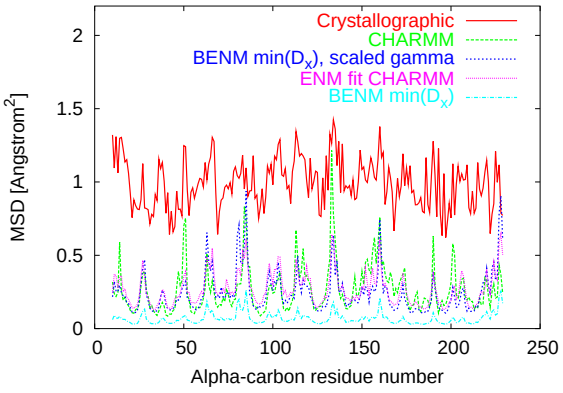

FIG. 2: Mean-squared displacements of alpha-carbon positions for trypsinogen residues 10–229 obtained from normal-modes simulations using CHARMM (*dashed green*), a BENM with parameters that minimize $D_{\mathbf{x}_\alpha}$ with respect to CHARMM (*dotted blue*), the same BENM but with γ adjusted to better agree with CHARMM MSDs (*fine-dotted magenta*), and an ENM with parameters adjusted to agree with CHARMM MSDs (*dash-dotted cyan*). Values were calculated at $T = 300$ K using the Equipartition Theorem. Harmonic vibrations at thermal equilibrium are known to inadequately model crystallographic MSDs, which include other sources of disorder (*solid red*) [9].

resulting in a poor model of the density-of-states distribution. The ENM with parameters that minimize $D_{\mathbf{x}_\alpha}$ exhibits poor MSDs (not shown); however, an ENM with $r_c = 15.4$ Å and $\gamma = 0.4$ Kcal/mol Å^2 yields MSDs that agree well with those of the CHARMM model (Fig. 2). In agreement with previous results using the ENM [4], we confirmed that the parameters of both the ENM and BENM may be adjusted to yield a reasonable model of crystallographic MSDs (not shown).

Next consider the problem of quantifying allosteric effects in proteins [3]. In allosteric regulation, molecular interactions cause changes in protein activity through changes in protein conformation. Although the importance of considering continuous conformational distributions in understanding allosteric effects was recognized by Weber [10], theories of allosteric regulation that consider continuous conformational distributions have been lacking. We began to develop such a theory by defining the allosteric potential as the Kullback-Leibler divergence $\bar{D}_{\mathbf{x}}$ between protein conformational distributions before and after ligand binding, and by calculating changes in the conformational distribution of the full protein-ligand complex in the harmonic approximation [3]. Here we use the expression for the marginal distribution in Eq. (7) to calculate an equation for the allosteric potential in the harmonic

approximation, and apply it to analyze allosteric mechanisms in trypsinogen.

Let $\mathbf{x}_p$ be the protein coordinates selected from the coordinates $\mathbf{x}$ of a protein-ligand complex. $P'(\mathbf{x}_p)$ and $P(\mathbf{x}_p)$ are the protein conformational distributions with and without a ligand interaction. Eq. (7) enables $P'(\mathbf{x}_p)$ to be calculated from the full conformational distribution $P'(\mathbf{x})$ of the protein-ligand complex. The equation for the allosteric potential in the harmonic approximation follows from the theory developed in ref. [3]:

$$\bar{D}_{\mathbf{x}} = \sum_{i=1}^{3N_p} \left(\ln \frac{\bar{\omega}_i'}{\omega_i} + \frac{1}{2k_BT}\omega_i^2 \left| \mathbf{v}_i^\dagger \Delta \mathbf{x}_{p,0} \right|^2 + \frac{1}{2} \sum_{j=1}^{3N_p} \frac{\omega_j^2}{\bar{\omega}_i'^2} \left| \bar{\mathbf{v}}_i'^\dagger \mathbf{v}_j \right|^2 - \frac{1}{2} \right). \tag{11}$$

In Eq. (11), $\bar{\omega}'^2$ and $\bar{\mathbf{v}}_i'$ are the i^{th} eigenvalue and eigenvector of the matrix $\bar{\mathbf{H}}$ calculated for the protein atoms of the protein-ligand complex, ω_i^2 and $\mathbf{v}_i$ are the eigenvalue and eigenvector of mode i of the apo-protein, and $\Delta\mathbf{x}_{p,0} = \mathbf{x}'_{p,0} - \mathbf{x}_{p,0}$ is the difference between the equilibrium coordinates of the protein with and without the ligand interaction. The term $\sum_{i=1}^{3N_p} \ln \bar{\omega}_i'/\omega_i$ is proportional to the change in configurational entropy of the protein releasing the ligand, and the term $\sum_{i=1}^{3N_p} \omega_i^2 \left| \mathbf{v}_i^\dagger \Delta \mathbf{x}_{p,0} \right|^2 / 2k_BT$ is proportional to the potential energy required to deform the apo-protein into its equilibrium conformation in the protein-ligand complex.

We used Eq. (11) to calculate changes in the configurational distribution of local regions of trypsinogen upon binding bovine pancreatic trypsinogen inhibitor (BPTI). BPTI binds in the active site and exerts an allosteric effect, enhancing the affinity of trypsinogen for Val-Val [11]. Alpha-carbon coordinates for 223 residues were obtained from a crystal structure of trypsinogen in complex with BPTI (residues 7–229 from PDB entry 4TPI [8], including theoretically modeled residues 7–9), and were used directly to construct backbone-enhanced elastic network models of apo-trypsinogen and the trypsinogen-BPTI complex. As suggested by the refined trypsinogen model above, both models used $r_c = 10.5$ Å, $\gamma = 4.26$ Kcal/mol Å^2, and $\epsilon = 42$.

Local changes in the conformational distribution of trypsinogen were analyzed by considering changes in the neighborhood of each alpha-carbon atom. A neighborhood was defined by selecting the atom of interest plus its five nearest neighbors, and the matrix $\bar{\mathbf{H}}$ was calculated for these six atoms in the models both with (yielding $\bar{\mathbf{H}}'$) and without (yielding $\bar{\mathbf{H}}$) the BPTI interaction. A local value of $\bar{D}_{\mathbf{x}}$ was obtained using the eigenvalues and eigenvectors of $\bar{\mathbf{H}}'$ and $\bar{\mathbf{H}}$ in a suitably modified version of Eq. (11).

Not surprisingly, we found that the local values of $\bar{D}_{\mathbf{x}}$ were relatively large in the neighborhood of the BPTI-binding site (Fig. 3, *left panel*). Values of $\bar{D}_{\mathbf{x}}$ elsewhere on the surface

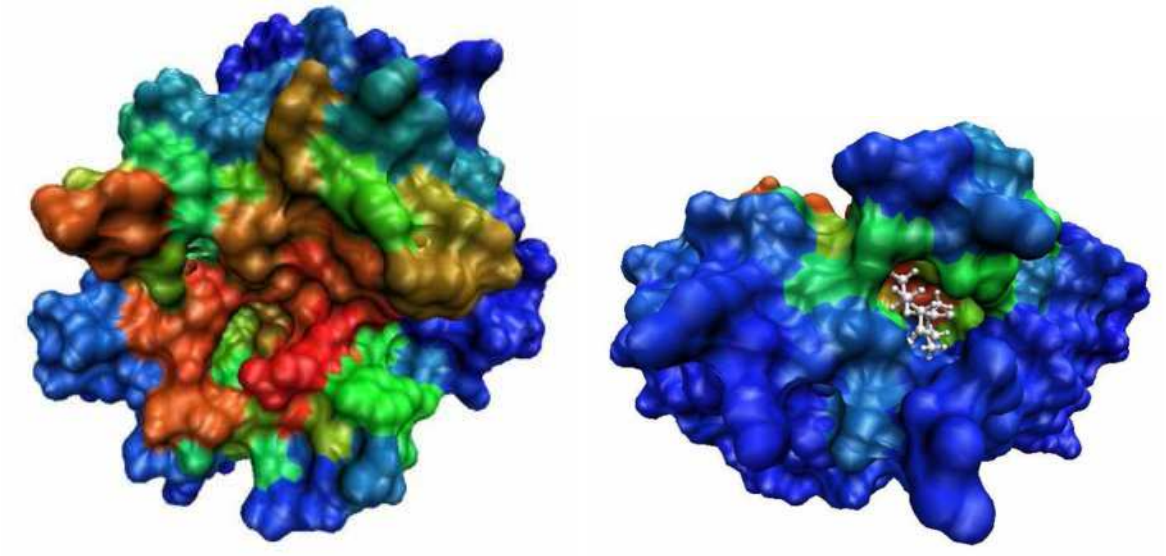

FIG. 3: Visualization of local sites on the surface of trypsinogen that exhibit a large change in the conformational distribution upon binding BPTI. Values of $\bar{D}_{\mathbf{x}}$ are mapped to a logarithmic temperature scale, with red coloring indicating large values. Changes are large both in the BPTI-binding site (*left*) and in the Val-Val binding site (*right*). There is a 90° rotation about the x-axis between the left and right panels.

were smaller, with one interesting exception: values in the Val-Val binding site were comparable to those in the BPTI-binding site (Fig. 3, *right panel*).

We also calculated local values of $\bar{D}_{\mathbf{x}}$ for the Val-Val interaction, which causes the crystal structure of trypsinogen to resemble that of active trypsin [8, 12]. We found that values were relatively large in the neighborhood of Ser 195, which is the key catalytic residue for trypsin and other serine proteases: the value of $\bar{D}_{\mathbf{x}}$ in this neighborhood was 40^{th} highest of 223 residues in the crystal structure; 11^{th} of all residues not directly interacting with the Val-Val in the model; the highest of all residues located at least as far as Ser 195 is from the Val-Val ligand; and greater than that for 20 of 60 residues located closer to the ligand. Calculations for both the BPTI interaction and the Val-Val interaction therefore indicate that there is a relatively strong communication between the regulatory and active sites of trypsinogen.

Considering models beyond the ENM and BENM (and even models beyond proteins), the theory presented here leads to a general prescription for modeling harmonic vibrations using coarse-grained models of materials. To optimally model the all-atom conformational

distribution, always use an energy scale for interactions that eliminates the discrepancy due to differences in the eigenvectors (Eq. (10)), and select the coarse-grained model for which the entropy of the conformational distribution is the largest (first term of Eq. (9)).

Although traditional elastic network models can explain characteristics of the functions and dynamics of proteins [13], the present study shows that they provide a poor approximation to the conformational distribution calculated from all-atom models of harmonic vibrations of proteins. Model accuracy is significantly improved by using a backbone-enhanced elastic network model, which strengthens interactions between atoms that are nearby in terms of covalent linkage. Although the backbone-enhanced model appears to accurately capture the high-frequency alpha-carbon vibrations of an all-atom model, the model less accurately captures the slower, large-scale harmonic vibrations (which in turn are known to poorly approximate the full spectrum of highly nonlinear, large-scale protein motions).

We also find that the allosteric potential is a useful tool for computational analysis of allosteric mechanisms in proteins. Using calculations of the allosteric potential, communication between the regulatory and active sites of trypsinogen was observed in a purely mechanical, coarse-grained model of protein harmonic vibrations that does not consider mean conformational changes or amino-acid identities, supporting prior arguments for the possibility of allostery without a mean conformational change [14]. It will be interesting to perform similar analyses on a wide range of all-atom and coarse-grained models of protein vibrations, and to use more realistic calculations of free-energy landscapes [15] to more accurately model changes in protein conformational distributions.

This work was supported by the US Department of Energy.

---

* Correspondence: mewall@lanl.gov

[1] H. Frauenfelder and P. G. Wolynes, Science **229**, 337 (1985).

[2] S. Kullback and R. Leibler, Annals of Math. Stats. **22**, 79 (1951).

[3] D. Ming and M. E. Wall, Proteins **59**, 697 (2005).

[4] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).

[5] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput.

Chem. **4**, 187 (1983).

[6] W. Zheng and B. Brooks, Biophys. J. **89**, 167 (2005).

[7] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).

[8] W. Bode, J. Walter, R. Huber, H. R. Wenzel, and H. Tschesche, Eur. J. Biochem. **144**, 185 (1984).

[9] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. USA **80**, 3696 (1983).

[10] G. Weber, Biochemistry **11**, 864 (1972).

[11] W. Bode, J. Mol. Biol. **127**, 357 (1979).

[12] W. Bode, P. Schwager, and R. Huber, J. Mol. Biol. **118**, 99 (1978).

[13] L. W. Yang, X. Liu, C. J. Jursa, M. Holliman, A. J. Rader, H. A. Karimi, and I. Bahar, Bioinformatics **21**, 2978 (2005).

[14] A. Cooper and D. T. Dryden, Eur. Biophys. J. **11**, 103 (1984).

[15] A. E. Garcia and K. Y. Sanbonmatsu, Proteins **42**, 345 (2001).