QUANTIFYING ALLOSTERIC EFFECTS IN PROTEINS

Dengming Ming[1] and Michael E. Wall[1,2*]

[1] *Computer and Computational Sciences Division, Los Alamos National Laboratory, Los Alamos NM 87545*

[2] *Bioscience Division, Los Alamos National Laboratory, Los Alamos NM 87545*

* Correspondence to: Michael E. Wall, Computer and Computational Sciences Division, Los Alamos National Laboratory, Mail Stop B256, Los Alamos, NM, 87545 USA. E-mail: mewall@lanl.gov

Los Alamos National Laboratory LA-UR-04-4470

Keywords: allosteric regulation, protein dynamics, conformational distribution, design principles, ligand binding, protein activity, control points, Kullback-Leibler divergence, allosteric potential, rate divergence

ABSTRACT

In allosteric regulation, protein activity is altered when ligand binding causes changes in the protein conformational distribution. Little is known about which aspects of protein design lead to effective allosteric regulation, however. To increase understanding of the relation between protein structure and allosteric effects, we have developed theoretical tools to quantify the influence of protein/ligand interactions on probability distributions of reaction rates and protein conformations. We define the rate divergence, $\overline{D}_k$, and the allosteric potential, $\overline{D}_{\mathbf{x}}$, as the Kullback-Leibler divergence between either the reaction-rate distributions or protein conformational distributions with and without ligand bound. We then define $D_{\mathbf{x}}$ as the change in the conformational distribution of the combined protein/ligand system, derive $D_{\mathbf{x}}$ in the harmonic approximation, and identify contributions from three separate terms: one term, $D_{\mathbf{x}}^{\omega}$, results from changes in the eigenvalue spectrum; a second term, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$, results from changes in the mean conformation; and a third term, $D_{\mathbf{x}}^{\mathbf{v}}$, corresponds to changes in the eigenvectors. Using normal modes analysis, we have calculated these terms for a natural interaction between lysozyme and the ligand tri-N-acetyl-D-glucosamine, and compared them with calculations for a large number of simulated random interactions. The comparison shows that interactions in the known binding-site are associated with large values of $D_{\mathbf{x}}^{\mathbf{v}}$. The results motivate using allosteric potential calculations to predict functional binding-sites on proteins, and suggest the possibility that, in Nature, effective ligand interactions occur

at intrinsic control points at which binding induces a relatively large change in the protein conformational distribution.

INTRODUCTION

Modern understanding of allosteric mechanisms in proteins began with the symmetry model,[1] which sought to explain the Hill-equation binding of oxygen to hemoglobin[2] in terms of a structural difference between apo- and oxygenated hemoglobin.[3,4] Other models of allostery, beginning with the sequential model,[5] have led to modification and refinement of the original concepts, and today mechanisms of allosteric regulation are known in much more detail.[6,7] However, the central idea explaining allosteric mechanisms still holds: allosteric interactions cause conformational changes that lead to altered protein activity.

The importance of considering continuous conformational distributions in understanding allosteric effects was recognized by Weber.[8] Neutron scattering experiments later provided evidence for changes in protein dynamics upon ligand binding,[9,10] and it was subsequently realized that ligand binding at an allosteric site can influence binding at a remote site without inducing a mean conformational change, solely through alteration of atomic fluctuations.[11] Since then the role of protein dynamics in allosteric regulation has been the subject of numerous studies, mostly concerned with the effects of allosteric interactions on free energies.[12,13]

The importance of examining changes in the conformational distribution to understand allosteric effects in proteins has recently been emphasized.[14] The quantitative treatment of changes in the conformational distribution, however, demands advances in the theory of allostery. Therefore, in the spirit of progress in rate theories,[15] here we consider the effects of protein/ligand interactions on probability distributions of reaction rates and protein conformations. We define the rate divergence and allosteric potential as the Kullback-Leibler divergence[16] between reaction-rate distributions or conformational distributions before and after ligand binding. We then define $D_{\mathbf{x}}$ as the change in the conformational distribution of the combined protein/ligand system, and derive $D_{\mathbf{x}}$ in the harmonic approximation, resulting in three separate terms for contributions from changes in the eigenvalue spectrum ($D_{\mathbf{x}}^{\omega}$), the mean conformation ($D_{\mathbf{x}}^{\Delta\mathbf{x}}$), and the eigenvectors ($D_{\mathbf{x}}^{\mathbf{v}}$). We describe normal modes analysis of an X-ray structure of lysozyme with and without ligand, and use the results of the simulation to calculate $D_{\mathbf{x}}$ for ligand binding to lysozyme in the harmonic approximation. Finally, we compare calculations of $D_{\mathbf{x}}$ for a large number of random protein/ligand binding configurations, and find that interactions in the natural binding site have a relatively high value of $D_{\mathbf{x}}^{\mathbf{v}}$. The results motivate using calculations of $D_{\mathbf{x}}^{\mathbf{v}}$ to predict functional binding sites on proteins, and suggest that functional sites might correspond to control points at which binding causes a relatively large change to the protein conformational distribution.

MATERIALS AND METHODS

Normal Modes Analysis

An atomic model of lysozyme was taken from a 1.77 Å X-ray structure of turkey egg-white lysozyme in complex with tri-NAG (Protein Data Bank (PDB) entry 1JEF[17]) (Fig. 1). Water molecules were removed and hydrogen atoms were added to the protein using the HBUILD module of CHARMM[18]. The extended-atom model TOPH19 and the polar hydrogen parameter set for proteins, PARAM19, were used for the energy minimization and normal modes analysis.

To calculate the normal modes of apo-lysozyme, the tri-NAG atoms were removed and the energy was minimized *in vacuo* using the Adopted Basis Newton-Raphson (ABNR) minimization algorithm[19] with a distance-dependent dielectric constant of 4 times the distance ($\varepsilon = 4r$), and a final root-mean-square (RMS) energy gradient of 10^{-7} kcal/mol Å. This RMS gradient has been shown to be satisfactory for calculating normal modes and is expected to yield only real-frequency modes.[20] As in a previous normal modes study of lysozyme,[21] the cutoff distance for the non-bonded neighbor list update was 8 Å, and a switching function with an "on" distance of 6.5 Å and an "off" distance of 7.5 Å was used for energy evaluations. The neighbor list was updated every 10 minimization steps.

To facilitate analysis of the lysozyme complex, we used a system composed of the protein and the NAG monomer (residue 132) which is most deeply bound in the lysozyme cleft (Fig. 1), removing the two other NAGs. The extended-atom representation of the NAG molecule and the corresponding energy parameters were taken from a published normal-modes study of lysozyme.[21] As was done in that study, we began the energy minimization with a refinement of the coordinates of the ligand (NAG) in the presence of the protein. Steepest descent minimization was performed for a maximum of 2000 steps or until the RMS energy gradient fell below 1×10^{-7} kcal /mol Å, using harmonic constraints on all atoms. Very large harmonic constraints, k_c=1×10^{7} kcal /(atom mass) mol Å^2, were placed on the protein atoms, and smaller constraints, k_c=10.0 kcal /(atom mass) mol Å^2, were placed on the ligand atoms. This initial minimization improved the ligand coordinates without significantly changing the refined X-ray coordinates of the protein. The entire system was then minimized as for the apo-lysozyme minimization.

To calculate the normal modes of a model, the force constant matrix was calculated using the VIBRAN module of CHARMM, and was diagonalized using the DIAG module. Potentials were evaluated *in vacuo* with $\varepsilon = 4r$, using the same cutoff conditions on the non-bonded energy terms as were used in minimization. We ensured that there were no modes with negative eigenvalues, and six modes with near-zero frequency (smaller than 0.003 cm^{-1}), corresponding to rigid-body motions of the protein.

To derive $D_{\mathbf{x}}$ in the harmonic approximation, the apo-protein model is assumed to include a non-interacting ligand (see below). Normal modes analysis of such a model yields eigenvectors $\mathbf{v}_i = \left(\mathbf{v}_i^{\mathrm{P}}, \mathbf{v}_i^{\mathrm{L}}\right)$ with both protein elements $\mathbf{v}_i^{\mathrm{P}}$ and ligand elements $\mathbf{v}_i^{\mathrm{L}}$. Because there is no interaction between the protein and ligand, each mode is either a protein mode, with non-zero elements in $\mathbf{v}_i^{\mathrm{P}}$ and zero-valued elements in $\mathbf{v}_i^{\mathrm{L}}$, or a ligand mode, with zero-valued elements in $\mathbf{v}_i^{\mathrm{P}}$ and non-zero elements in $\mathbf{v}_i^{\mathrm{L}}$. The apo-protein eigenvectors $\mathbf{v}_i$ have the same dimensionality as those of the protein/ligand complex ($\mathbf{v}'_i$), enabling the equation for $D_{\mathbf{x}}$ to be derived. To concentrate on changes in the protein and not in the ligand, however, we did not include a non-interacting ligand in the CHARMM simulations of apo-lysozyme. In eliminating the ligand from the apo-lysozyme model, we obtained only the protein modes and protein elements $\mathbf{v}_i^{\mathrm{P}}$ of each eigenvector. The ligand modes and ligand elements $\mathbf{v}_i^{\mathrm{L}}$ of the non-interacting NAG were not needed in the calculations below.

Generation of Structures with Random Protein/Ligand Interactions

New complexes with NAG located at random surface points were generated with the aid of the program MSMS.[22] Surface points were created by running MSMS on apo-lysozyme with a probe radius of 1.4 Å (Fig. 1). A total of 111 points were randomly selected to generate structures with arbitrary protein/ligand interactions. Ten additional points were manually selected to examine complexes in which NAG is shifted just 0.1 Å-2.0 Å from the natural binding site. For a given surface point, the NAG molecule was

moved from its correct binding position to a new position by translation of its center of mass to the surface point. To avoid clashes between the ligand and the protein, the NAG molecule was then moved outwards 4.0 Å along the surface normal as calculated by MSMS (the NAG molecule has the approximate shape of a plane with dimensions 4.8 Å × 8.6 Å). Energy minimization and normal modes analysis were carried out following exactly the same procedure as for the NAG-lysozyme complex with correct binding position. The closest distance between heavy atoms in the ligand and protein varied from 2.7 Å to 3.0 Å, which is consistent with the distance of a hydrogen bond.

RESULTS

Theory

Definition of allosteric potential and rate divergence

Consider a protein that performs a function such as catalysis. The protein function is regulated by binding of a ligand at either the active site or an allosteric site. Let $P(k)$ and $P(\mathbf{x})$ be the equilibrium probability distribution of reaction rates k and configurations $\mathbf{x}$ of the apo-protein, and $P'(k)$ and $P'(\mathbf{x})$ be the same probability distributions when ligand is bound. We propose that a reasonable measure of the magnitude of an allosteric effect is the Kullback-Leibler divergence[16] $\overline{D}_k$ of the reaction rate probability distributions before and after ligand binding, calculated as

$$\overline{D}_k = \int_0^\infty dk \left(\log \frac{P'(k)}{P(k)} \right) P'(k), \tag{1}$$

which is always non-negative (see APPENDIX A). Because the relation between a protein's configuration and its reaction rate is complex, and because of limitations in the detailed knowledge of protein chemistry, there is currently no general method for theoretical calculation of $P(k)$ for proteins. We are therefore motivated to obtain an upper bound on $\overline{D}_k$ using the conformational distribution $P(\mathbf{x})$. Assuming the rate $k(\mathbf{x})$ is a function of the configuration $\mathbf{x}$, the relation between $P(k)$ and the conformational distribution $P(\mathbf{x})$ is

$$P(k) = \int d^{3N}\mathbf{x} P(\mathbf{x}) \delta(k(\mathbf{x}) - k), \tag{2}$$

where the integral is over all conformations $\mathbf{x}$.

The Kullback-Leibler divergence $\overline{D}_\mathbf{x}$ between the conformational probability distributions is

$$\overline{D}_\mathbf{x} = \int d^{3N}\mathbf{x} \left(\log \frac{P'(\mathbf{x})}{P(\mathbf{x})} \right) P'(\mathbf{x}), \tag{3}$$

where the integral is over all conformations **x**.

Assume $k(\mathbf{x})$ remains the same with and without ligand. Although $k(\mathbf{x})$ may exhibit a complex dependence on the conformation, if $k(\mathbf{x})$ is invertible, $\overline{D}_k = \overline{D}_{\mathbf{x}}$. If instead the rate k is the same for all conformations, then $\overline{D}_k = 0$. By general consideration of the assumptions that each rate k may correspond to more than one conformation **x**, that each conformation **x** corresponds to only a single rate k, and that $k(\mathbf{x})$ is the same with and without ligand, Equations (1) – (3) lead to the relation (see APPENDIX A):

$$\overline{D}_{\mathbf{x}} \geq \overline{D}_k . \tag{4}$$

Assuming Boltzmann distributions for $P(\mathbf{x})$ and $P'(\mathbf{x})$, Equation (3) shows that $\overline{D}_{\mathbf{x}}$ is essentially the negative mean relative energy of ligand-bound protein conformations, where each energy is measured relative to the energy of the equivalent apo-protein conformation. Because $\overline{D}_{\mathbf{x}}$ is associated with a mean energy, and because Equation (4) shows that $\overline{D}_{\mathbf{x}}$ is the highest possible value of $\overline{D}_k$, we call $\overline{D}_{\mathbf{x}}$ the allosteric potential.

Harmonic approximation

Here we obtain a simple expression for a quantity closely related to $\overline{D}_{\mathbf{x}}$ in the harmonic approximation of protein dynamics (see APPENDIX A for details not provided here). Let the protein/ligand conformation be $\mathbf{x} = (\mathbf{x}^{\mathrm{P}}, \mathbf{x}^{\mathrm{L}})$, consisting of $3N$ atomic coordinates for N atoms. The protein coordinates are represented by $\mathbf{x}^{\mathrm{P}}$, and the ligand coordinates by $\mathbf{x}^{\mathrm{L}}$.

To obtain a simple expression, instead of considering $P(\mathbf{x}^{\mathrm{P}})$ and calculating the allosteric potential $\overline{D}_{\mathbf{x}}$, we use the full protein/ligand conformational distribution $P(\mathbf{x})$ to calculate a related quantity that we denote as $D_{\mathbf{x}}$. We define the apo-protein system as a system in which there is no interaction between the protein and the ligand. The conformation may be described as a superposition of normal modes of either the apo-protein or the protein/ligand system:

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{3N} a_i(\mathbf{x})\mathbf{v}_i = \mathbf{x}'_0 + \sum_{i=1}^{3N} a'_i(\mathbf{x})\mathbf{v}'_i \,. \tag{5}$$

$\mathbf{x}_0$ is the equilibrium conformation, a_i are the coefficients of the expansion (with units of length L), and $\mathbf{v}_i = \left(\mathbf{v}_i^{\mathrm{P}}, \mathbf{v}_i^{\mathrm{L}}\right)$ are the eigenvectors of the Hessian matrix $\mathbf{H}(\mathbf{x})$ of the potential energy function $U(\mathbf{x})$ evaluated at $\mathbf{x}_0$: $H_{ij}\big|_{\mathbf{x}_0} = \partial^2 U / \partial x_i \partial x_j \big|_{\mathbf{x}_0}$. The protein elements of the eigenvectors are represented by $\mathbf{v}_i^{\mathrm{P}}$, and the ligand elements by $\mathbf{v}_i^{\mathrm{L}}$. Unprimed quantities refer to the apo-protein system, and primed quantities refer to the protein/ligand system. The expression for the potential energy of a conformation $\mathbf{x}$ is

$$U(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{3N} \omega_i^2 |a_i(\mathbf{x})|^2 \tag{6}$$

where ω_i^2 are the eigenvalues of the Hessian matrix and have units MT^{-2}, with M being mass and T being time. We assume a Boltzmann distribution of conformations, ignore the solvent, and ignore changes in volume.

Using Eq. (3), the expression for $D_{\mathbf{x}}$ is

$$D_{\mathbf{x}} = Z^{-1}\int d^{3N}\mathbf{a}'\left\{\log\frac{Z}{Z'} + \frac{1}{2k_BT}\sum_{i=1}^{3N}\left(\omega_i^2\left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i + \sum_{j=1}^{3N}a_j'\mathbf{v}_j'\cdot\mathbf{v}_i\right|^2 - \omega_i'^2\left|a_i'\right|^2\right)\right\}e^{\frac{-1}{2k_BT}\sum_{k=1}^{3N}\omega_k'^2|a_k'|^2} \qquad (7)$$

Where $\Delta\mathbf{x}_0 = \mathbf{x}_0' - \mathbf{x}_0$, the integral is over the range [-∞,∞] in each coefficient a_i, and the partition function Z is given by

$$Z = \int d^{3N}\mathbf{a}e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i^2|a_i|^2} = \left(2\pi k_BT\right)^{3N/2}\prod_{i=1}^{3N}\omega_i^{-1} \qquad (8)$$

with an equivalent expression for Z'. Equation (7) evaluates to

$$D_{\mathbf{x}} = \sum_{i=1}^{3N}D_{\mathbf{x}i} = \sum_{i=1}^{3N}\left(\log\frac{\omega_i'}{\omega_i} + \frac{1}{2k_BT}\omega_i^2\left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i\right|^2 + \frac{1}{2}\sum_{j=1}^{3N}\frac{\omega_j^2}{\omega_i'^2}\left|\mathbf{v}_i'\cdot\mathbf{v}_j\right|^2 - \frac{1}{2}\right), \qquad (9)$$

which can be calculated exactly given the normal modes of the apo-protein and protein/ligand systems. For convenience, indices i and j have been exchanged in the double-summation term. Each mode i contributes an independent term $D_{\mathbf{x}i}$ to the total $D_{\mathbf{x}}$. Each $D_{\mathbf{x}i}$ is in turn the sum of three kinds of term: the leftmost term $D_{\mathbf{x}i}^{\omega} = \log\left(\omega_i'/\omega_i\right)$ is the contribution from the difference in the eigenvalue spectrum; the middle term $D_{\mathbf{x}i}^{\Delta\mathbf{x}} = \omega_i^2\left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i\right|^2/2k_BT$ is the contribution from the difference in the mean

conformation; the remaining terms, $D_{\mathbf{x}i}^{\mathbf{v}} = \sum_j \left(\omega_j^2 / \omega_i'^2\right) \left|\mathbf{v}_i' \cdot \mathbf{v}_j\right|^2 / 2 - 1/2$, make up the contribution from the difference in the eigenvectors. We also define the sums $D_{\mathbf{x}}^{\omega} = \sum_i D_{\mathbf{x}i}^{\omega}$; $D_{\mathbf{x}}^{\Delta\mathbf{x}} = \sum_i D_{\mathbf{x}i}^{\Delta\mathbf{x}}$; and $D_{\mathbf{x}}^{\mathbf{v}} = \sum_i D_{\mathbf{x}i}^{\mathbf{v}}$.

The normal modes may also be defined in terms of mass-weighted coordinates $\mathbf{x} \to \mathbf{M}^{1/2}\mathbf{x}$, where $\mathbf{M} = \mathrm{diag}\left(m_1, m_1, m_1, ..., m_N, m_N, m_N\right)$, m_k being the mass of atom k, as is done in CHARMM's VIBRAN module[18]. In this case, $D_{\mathbf{x}}$ is still given by Eq. (9), but the coefficients a_i have units $M^{1/2}L$, and the eigenvalues ω_i^2 are squared frequencies with units T^{-2}. The frequency values ω_i may also be divided by the speed of light to yield units L^{-1}, as is done in CHARMM, and as is the convention used to report results below.

Analysis of a Natural Lysozyme/Ligand Interaction

Lysozyme is a protein that has been well-studied by normal modes simulations;[21,23,24] its thermal vibrations have been studied experimentally using inelastic neutron scattering.[10] Because the active site is located in a cleft between two large lobes of the protein, we expect ligand binding at the active site to significantly change the conformational distribution of lysozyme. To quantify this effect, as an application of our theory, we have calculated the normal modes of an X-ray crystallography model of lysozyme both with and without its interaction with tri-N-acetyl-D-glucosamine (tri-NAG), and have estimated $D_{\mathbf{x}}$ in the harmonic approximation.

Normal modes

For the apo-protein, the root-mean-square deviation (RMSD) of atomic positions before and after energy minimization was 1.36 Å for all atoms, and 1.01 Å for backbone atoms. For the lysozyme/NAG complex, the RMSD of all protein atoms was 1.30 Å, 1.03 Å for the backbone atoms, and 0.43 Å for the NAG molecule. The RMSD between the minimized apo-protein and the minimized complex was 0.51 Å for the backbone atoms.

Frequency spectra for the apo-protein and the lysozyme/NAG complex are shown in Fig. 2. In each case, there are about 250 modes with frequencies below 50 cm^{-1}; such low-frequency modes are known to describe most of the displacements of the backbone atoms, which are most important for describing large-scale conformational changes.[25] The calculated spectra are very similar, and have the same overall frequency-dependence: in the low frequency region (< 250 cm^{-1}), a maximum density of 7 cm appears near the frequency 50 cm^{-1}, after which the density of the modes drops very sharply until the frequency 120 cm^{-1}; it then tapers from 3cm at 120 cm^{-1} to 2cm at 500 cm^{-1} with a standard deviation of about 1 cm.

We found significant changes in the low-frequency eigenvectors between the apo-protein and the complex. The very lowest-frequency mode (mode 7, 3.39 cm^{-1}) of the apo-protein corresponds to a torsion between the two major lobes of the protein, which bracket the binding cleft. The dot-product squared of the protein-element component of the eigenvector of this mode with that of mode 7 of natural lysozyme/NAG complex

$\left|\mathbf{v}_7^{\mathrm{P}} \cdot \mathbf{v}_7'^{\mathrm{P}}\right|^2$ has a value of just 0.01. We could not locate a clear analogue of mode 7 of the apo-protein among the modes in the complex, but the first 10 modes of the complex account for much of this mode ($\sum_{i=7,16}\left|\mathbf{v}_7^{\mathrm{P}} \cdot \mathbf{v}_i'^{\mathrm{P}}\right|^2 = 0.85$). The second-lowest-frequency mode (mode 8, 3.45 cm^{-1}) of the apo-protein, which is a bending mode and corresponds to an opening and closing of the cleft, appears to correspond to the very lowest-frequency mode (mode 7, 3.13 cm^{-1}) of the natural lysozyme/NAG complex, but $\left|\mathbf{v}_8^{\mathrm{P}} \cdot \mathbf{v}_7'^{\mathrm{P}}\right|^2$ is only 0.56. By inspection of visualized motions, the difference appears to come from an approximately 30º rotation of the bending plane of mode 7 in the complex with respect to the plane of mode 8 in the apo-protein.

Estimation of $\mathbf{D}_x$

Differences in the eigenvalue spectrum, mean conformation and eigenvectors were quantified using the three contributions to $D_{\mathbf{x}}$ in the harmonic approximation: $D_{\mathbf{x}}^{\omega}$, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ and $D_{\mathbf{x}}^{\mathbf{v}}$. To concentrate on changes in the protein, we only used the protein elements of the eigenvectors in calculation of Eq. (9), and to focus on the influence of large-scale conformational changes, we only used low-frequency modes to calculate $D_{\mathbf{x}}^{\omega}$, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$, and $D_{\mathbf{x}}^{\mathbf{v}}$.

The value of $D_{\mathbf{x}}^{\omega}$ was calculated to be -5.75 using the first 250 modes. The dependence of $D_{\mathbf{x}i}^{\omega}$ on i exhibits high variation at low frequencies, with the amplitude decreasing as the

mode index increases (Fig. 3). At higher frequencies, the value of $D_{\mathbf{x}i}^{\omega}$ exhibits variations about a mean of approximately -0.05.

For calculation of $D_{\mathbf{x}}^{\Delta\mathbf{x}}$, we first made a mass-weighted least-squares superposition of the apo-protein coordinates and the coordinates from the protein/ligand complex using the McLachlan algorithm,[26] in an effort to minimize the contribution due to rigid-body protein displacements. Each term $D_{\mathbf{x}i}^{\Delta\mathbf{x}}$ involves a factor that is the squared inner product $\left|\Delta\mathbf{x}_0 \cdot \mathbf{v}_i\right|^2$ between the mean coordinate change upon ligand binding and an eigenvector of the apo-protein. This factor corresponds to the fraction of the total mean squared displacement (MSD) $\left|\Delta\mathbf{x}_0\right|^2$ that is accounted for by a displacement along eigenvector $\mathbf{v}_i$ in configuration space. To determine how well the low-frequency normal modes can express the conformational change upon ligand binding, we calculated $\left|\Delta\mathbf{x}_0 \cdot \mathbf{v}_i^{\mathrm{P}}\right|^2$ for each eigenvector using just the subspace of the protein Cα coordinates (Fig. 4). We found that 80% of the total Cα MSD was explained by a sum of the contributions from the 100 lowest-frequency eigenvectors (Fig. 4). The value of $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ was calculated to be 0.324 using these 100 eigenvectors, assuming a temperature of 300 K.

A value for $D_{\mathbf{x}}^{\mathbf{v}}$ was calculated using the first 250 modes of the complex, and the first 300 modes of the apo-protein. We used more modes for the apo-protein than for the complex because we expect each eigenvector of the complex to be mostly accounted for by a

linear combination of apo-protein eigenvectors that have neighboring frequencies, including higher frequencies. The value obtained was 33.0. The distribution of $D_{\mathbf{x}i}^{\mathbf{v}}$ versus mode index i (Fig. 5) indicates that the main contribution to $D_{\mathbf{x}}^{\mathbf{v}}$ comes from the low-frequency modes.

Analysis of Random Lysozyme/Ligand Interactions

To investigate the possibility of using the allosteric potential to predict ligand-binding sites, the three terms of the harmonic approximation of $D_{\mathbf{x}}$ were estimated for a large number of lysozyme complexes with random ligand interactions (see MATERIALS AND METHODS). We also examined the effect of different ligand interactions on plots of the frequency spectrum and on the magnitude of the conformational change as measured by the RMSD of atomic positions.

The frequency distributions of new complexes were very similar to that of the apo-protein (Fig. 2). Most of the complexes have their lowest frequency in the range 2.0 cm^{-1} to 3.5 cm^{-1}, and 40 percent of them have a higher lowest-frequency than that of natural lysozyme/NAG complex (Fig. 6). About 55 percent of the points in Fig. 1 have a higher $D_{\mathbf{x}}^{\omega}$ than that of the natural binding site (Fig. 7). The results indicate that $D_{\mathbf{x}}^{\omega}$ for the X-ray structure with correct binding site lies in the middle of those for complexes whose binding sites are at random positions on the protein surface.

Minimization of the structures resulted in RMSDs of backbone atoms with respect to the X-ray structure in the range 0.75 Å to 1.25 Å (Fig. 8). We found that 50% of the complexes have larger RMSDs than that of the natural lysozyme/NAG complex. Eighty percent of the complexes have RMSD values within 0.1 Å of that of the natural lysozyme/NAG complex. When the RMSD is calculated instead with respect to the structure of the fully minimized apo-protein, 60% of the complexes were found to have a larger backbone RMSD than that of the natural lysozyme/NAG complex (Fig. 9).

Values of the mean-conformation contribution, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$, for the complexes span from almost 0 to 0.6 (Fig. 10). As was done for the natural complex, only the 100 lowest-frequency modes were used to calculate $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ for new complexes (the contribution of individual modes to the total RMSD for new complexes was similar to the case illustrated in Fig. 4). About 55% of the binding sites generate a higher $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ than that of the natural binding site and about 60% of all points are concentrated in the range 0.25 to 0.4. Even for the 10 complexes with NAG located in the neighborhood of the natural binding site, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ can vary from 0.25 to 0.4.

Results of calculations of the contribution to $D_{\mathbf{x}}$ from the eigenvectors, $D_{\mathbf{x}}^{\mathbf{v}}$, are shown in Fig. 11. Values were calculated as for the natural protein/ligand complex, and are in the range -5 to 55. Just ten (less-than 10%) of the test points resulted in a higher value of $D_{\mathbf{x}}^{\mathbf{v}}$ than that of the natural protein/ligand complex. Near the natural binding site, 7 of 10

manually placed points were found to have high D_x^v values. One of these is the highest-value point, located about 1.3 Å from the natural binding position.

Of the 10 automatically generated points with high values of D_x^v, four are near residues that are important for substrate interactions and catalysis (Fig. 12). The two green points in the cleft are near residues 35, 52, 56-58, and 107-109. Of these residues, Glu35 and Asp52 are the catalytic residues of the Phillips mechanism,[27] and Gln57, Ala107 and Trp108 are involved in interactions with the substrate.[28] The orange point also lies within the cleft, contacting residues 62, 98, and 107; in addition to Ala107, Trp62 is involved in ligand interactions.[28] The magenta point is near Arg114, which is involved in ligand interactions.[28] Of the other six points, the two cyan points are near residues 79-85, which overlap the EF-hand calcium-binding motif in a structure of the homologous baboon α-lactalbumin (PDB code 1ALC[29]); calcium also binds to this motif in some lysozymes. The four red points are located away from known functional sites: one point is near the N-terminus (residues 1-4); another is near residue 13, at the end of the N-terminal helix, and residue 18, at a turn after the N-terminal helix; the final two are near residues 16 and 96. Because terminal residues tend to exhibit large fluctuations, the interactions near the N-terminus might be expected to cause large changes in the conformational distribution. It would be interesting to learn whether lysozyme residues 13, 16, 18, and 96 are located at functionally important sites.

CONCLUSIONS

In the traditional view of allosteric mechanisms, ligand binding at a regulatory site induces a change in the mean protein conformation. In the perspective supported here, ligand binding at a regulatory site induces a change in the entire conformational distribution $P(\mathbf{x})$ of a protein. We have proposed the allosteric potential $\overline{D}_{\mathbf{x}}$ as a measure of the change in $P(\mathbf{x})$ upon ligand binding, and developed $D_{\mathbf{x}}$ as a measure of the change in the combined protein/ligand conformational distribution. It will be interesting to further develop methods for calculating the allosteric potential, not only because it measures the global change in $P(\mathbf{x})$, but also because it serves as an upper limit on the change in $P(\mathbf{x})$ local to an active site, and thus provides a quantitative limit of an allosteric effect on an active site. To address specific allosteric mechanisms, it will be interesting to calculate local allosteric potentials, such as the change in the conformational distribution of an active site upon ligand binding to a regulatory site.

Calculation of $D_{\mathbf{x}}$ in the harmonic approximation yields some general insights into the nature of allosteric effects in proteins. The harmonic approximation yields three terms that contribute to the total $D_{\mathbf{x}}$ (Eq. (9)): one term, $D_{\mathbf{x}}^{\omega}$, corresponds to changes in the frequency spectrum; a second term, $D_{\mathbf{x}}^{\Delta\mathbf{x}}$, corresponds to changes in the mean conformation; and a third term, $D_{\mathbf{x}}^{\mathbf{v}}$, corresponds to changes in the eigenvectors. Through this separation of terms, it is apparent that there need not be any difference at all in the mean conformation upon ligand binding to result in a large value of $D_{\mathbf{x}}$, supporting the

possibility of allostery without conformational change.[11] In the analysis of different lysozyme/NAG interactions, the values of $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ were usually small compared to those of either $D_{\mathbf{x}}^{\omega}$ or $D_{\mathbf{x}}^{\mathbf{v}}$, indicating that, compared to changes in the eigenvalues or eigenvectors, changes in the mean conformation contributed relatively little to the value of $D_{\mathbf{x}}$, and therefore had a relatively small effect on the conformational distribution. The value of $D_{\mathbf{x}}$ may be large even if there is no change in the eigenvalues, indicative of the fact that it is sensitive to the detailed changes in the conformational distribution that are captured by the eigenvectors.

Although changes in the eigenvectors upon ligand binding are measured by the term $D_{\mathbf{x}}^{\mathbf{v}}$, analysis of allosteric effects using $D_{\mathbf{x}}^{\mathbf{v}}$ differs from a simple comparison of eigenvectors before and after ligand binding. An important difference is that a large change in the eigenvectors of a system does not in itself imply a large change in the conformational distribution, whereas a large change in $D_{\mathbf{x}}^{\mathbf{v}}$ does. For example, consider a system in which two modes A and B are degenerate, *i.e.*, the eigenvalues of the modes are the same. Given a solution with eigenvectors $\mathbf{v}_A$ and $\mathbf{v}_B$, any two eigenvectors that are orthonormal, linear combinations of $\mathbf{v}_A$ and $\mathbf{v}_B$ yield a physically equivalent solution. By only comparing the eigenvectors, one might incorrectly conclude that alternative solutions are different. Calculation of $D_{\mathbf{x}}^{\mathbf{v}}$ involves both the eigenvalues and the eigenvectors, however, and for this case yields a value of zero, indicating that the solutions are physically equivalent (see APPENDIX A). Our theoretical framework

provides a principled way to consider the eigenvalues when comparing eigenvectors, enhancing normal modes analysis.

Looking beyond estimation of the allosteric potential using the harmonic approximation, methods to estimate free-energy distributions using molecular dynamics simulations[30] will allow estimation of the allosteric potential using more reasonable molecular-mechanics Hamiltonians. Such simulations will enable a more realistic sampling of the energy landscape, and will enable the effects of the solvent and volume changes to be modeled. In addition, experimental characterization of the conformational distribution is currently possible using inelastic neutron scattering experiments,[9,10] which can provide measurements of the frequency spectrum, or diffuse X-ray scattering, which may be compared with models of the conformational distribution.[31-35] Such experiments might allow experimental estimation of the allosteric potential. Measurement of the rate divergence D_k requires measurement of the density of states $g(k)$, which has already been experimentally demonstrated for carbon monoxide binding to myoglobin.[36] Theoretical calculation of the rate divergence will rely on development of methods to calculate protein activity given a protein structure, which is a major goal of theoretical molecular biology.

In examining a large number of complexes with NAG at different binding positions, we found that most positions yielding a large value of $D_{\mathbf{x}}^{\mathbf{v}}$ are associated with functional sites of lysozyme. These results suggest that calculating $D_{\mathbf{x}}^{\mathbf{v}}$ for many simulated interactions

between proteins and ligands might help to predict functionally important binding sites on proteins. As was seen with lysozyme, such sites might be either active sites or regulatory sites. (Just as binding at a regulatory site may change protein activity, ligand binding to the active site might change protein activity, e.g., by altering the configuration of key catalytic residues.) Because factors other than the change in the protein conformational distribution might be important to the effect of any given protein/ligand interaction, we do not expect the predictions of functional binding sites to be perfectly accurate. Nevertheless, it would be interesting to test this idea through prediction of ligand-binding sites using a database of known protein/ligand structures. Methods that make use of simplified models of protein dynamics will be useful for this task.[37]

The development of the present theoretical tools for quantifying allosteric effects advances a theory of allosteric regulation which explicitly considers changes in the protein conformational distribution. The first application of these tools has suggested the possibility that, in Nature, effective ligand interactions are selected in a way that causes a relatively large change in the conformational distribution. Speculating about what might give rise to such a design principle, it is plausible that natural selection might favor biological systems in which protein activity is controllable by molecular signals. Such control can be facilitated by ligand interactions that cause large changes in the conformational distribution. Weber once offered a well-known description of a protein as a "kicking and screaming stochastic molecule."[38] Perhaps Nature favors ligand interactions that greatly perturb the stochastic dance of proteins.

ACKNOWLEDGEMENTS

We are grateful to the members of the Protein Function Inference Group at Los Alamos National Laboratory for discussions, and to Hans Frauenfelder, Kevin Sanbonmatsu, and Thomas Terwilliger for reading the manuscript. This work was supported by the Department of Energy under contract to the University of California.

APPENDIX A

The Kullback-Leibler divergence is non-negative

Non-negativity of the Kullback-Leibler divergence follows from the inequality $\log x \le x - 1$:

$$-D_{KL} = \int dx \left(\log \frac{P(x)}{P'(x)} \right) P'(x) \le \int dx \left(\frac{P(x)}{P'(x)} - 1 \right) P'(x) = 0 \Rightarrow D_{KL} \ge 0 \tag{A1}$$

The allosteric potential is an upper bound for the rate divergence

Consider the allosteric potential as a discrete sum:

$$\overline{D}_{\mathbf{x}} = \sum_{\mathbf{x}} \left(\log \frac{P'_{\mathbf{x}}}{P_{\mathbf{x}}} \right) P'_{\mathbf{x}} \tag{A2}$$

Now consider calculation of $\overline{D}_k$ when $k(\mathbf{x})$ is invertible except for just two conformations, named **a** and **b**, which yield the same rate. By induction, if $\overline{D}_{\mathbf{x}} \ge \overline{D}_k$ for this case, then it is also true for higher degeneracies. All of the terms in the sum are unchanged save those for the two conformations, which are combined into one term in $\overline{D}_k$:

$$d_k = \log \left(\frac{P'_{\mathbf{a}} + P'_{\mathbf{b}}}{P_{\mathbf{a}} + P_{\mathbf{b}}} \right) \left(P'_{\mathbf{a}} + P'_{\mathbf{b}} \right) \tag{A3}$$

The equivalent contribution in $\overline{D}_{\mathbf{x}}$ is

$$d_{\mathbf{x}} = \log\left(\frac{P'_{\mathbf{a}}}{P_{\mathbf{a}}}\right)P'_{\mathbf{a}} + \log\left(\frac{P'_{\mathbf{b}}}{P_{\mathbf{b}}}\right)P'_{\mathbf{b}}\,. \tag{A4}$$

We now show that $d_{\mathbf{x}} \geq d_k$, beginning with an expression for $d_{\mathbf{x}} - d_k$:

$$\Delta_z(z) = d_{\mathbf{x}} - d_k = P'_{\mathbf{a}}\left(z'\log\frac{z'}{z} - (1+z')\log\frac{1+z'}{1+z}\right) = P'_{\mathbf{a}}\delta_{z'}(z)\,. \tag{A5}$$

where $z = P_{\mathbf{b}}/P_{\mathbf{a}}$ and $z' = P'_{\mathbf{b}}/P'_{\mathbf{a}}$.

Differentiation of $\delta_{z'}(z)$ yields an extremum at $z = z'$, where $\delta_{z'}(z) = 0$:

$$\frac{d}{dz}\delta_{z'}(z) = -\frac{z'}{z} + \frac{1+z'}{1+z} = 0 \Rightarrow z = z'\,. \tag{A6}$$

The second derivative at $z = z'$ is positive:

$$\frac{d^2}{dz^2}\delta_{z'}(z)\Big|_{z=z'} = \frac{1}{z'} - \frac{1}{1+z'} > 0\,. \tag{A7}$$

The factor $\delta_{z'}(z)$ therefore has a minimum value of zero and is always positive. Because $P'_{\mathbf{a}}$ is positive, Eq. (A5) then implies $d_{\mathbf{x}} \geq d_k$, and therefore $\overline{D}_{\mathbf{x}} \geq \overline{D}_k$. By induction, and by extension to the integral formulas, $\overline{D}_{\mathbf{x}} \geq \overline{D}_k$ if $k = k(\mathbf{x})$.

$D_{\mathbf{x}}$ in the harmonic approximation

Here we give details of the calculation of $D_{\mathbf{x}}$ in the harmonic approximation. Assume a Boltzmann distribution of conformations, using the potential energy of Eq. (6):

$$P(\mathbf{x}) = Z^{-1} e^{\frac{-1}{2k_B T}\sum_{i=1}^{3N}\omega_i^2 |a_i(\mathbf{x})|^2} . \tag{A8}$$

Eq. (3) then becomes

$$D_{\mathbf{x}} = Z'^{-1}\int d^{3N}\mathbf{a}' \left\{ \log\left(\frac{Z}{Z'}\right) + \frac{1}{2k_B T}\left(\sum_{i=1}^{3N}\omega_i^2 |a_i|^2 - \sum_{i=1}^{3N}\omega_i'^2 |a_i'|^2\right)\right\} e^{\frac{-1}{2k_B T}\sum_{i=1}^{3N}\omega_i'^2 |a_i'|^2} . \tag{A9}$$

By Eq. (5) we obtain the relation

$$\mathbf{x}\cdot\mathbf{v}_j = \mathbf{x}_0\cdot\mathbf{v}_j + a_j = \mathbf{x}_0'\cdot\mathbf{v}_j + \sum_{i=1}^{3N} a_i'\mathbf{v}_i'\cdot\mathbf{v}_j , \tag{A10}$$

which we derive using the orthonormality of the $\mathbf{v}_j$'s. From Eq. (A10) we obtain

$$a_i = \Delta\mathbf{x}_0\cdot\mathbf{v}_i + \sum_{j=1}^{3N} a_j'\mathbf{v}_j'\cdot\mathbf{v}_i , \tag{A11}$$

leading to Eq. (7) in the text. Evaluation of the integrals resulting from the expansion of Eq. (7) are performed as follows. The first integral is trivial:

$$Z'^{-1}\int d^{3N}\mathbf{a}' \log\left(\frac{Z}{Z'}\right) e^{\frac{-1}{2k_B T}\sum_{i=1}^{3N}\omega_i'^2 |a_i'|^2} = \log\left(\frac{Z}{Z'}\right) \tag{A12}$$

Next, the integral

$$Z'^{-1}\int d^{3N}\mathbf{a}'\frac{1}{2k_BT}\sum_{j=1}^{3N}\omega_j'^2\left|a_j'\right|^2 e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i'^2|a_i'|^2} \tag{A13}$$

is a sum of Gaussian integrals of the type

$$Z'^{-1}\int d^{3N}\mathbf{a}'\frac{1}{2k_BT}\omega_j'^2\left|a_j'\right|^2 e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i'^2|a_i'|^2} = \frac{1}{2}. \tag{A14}$$

The remaining integrals involve the squared term

$$\left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i+\sum_{j=1}^{3N}a_j'\mathbf{v}_j'\cdot\mathbf{v}_i\right|^2 = \left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i\right|^2+\left|\sum_{j=1}^{3N}a_j'\mathbf{v}_j'\cdot\mathbf{v}_i\right|^2+2\Delta\mathbf{x}_0\cdot\mathbf{v}_i\sum_{j=1}^{3N}a_j'\mathbf{v}_j'\cdot\mathbf{v}_i \tag{A15}$$

The integral over the first term of Eq. (A15) yields $\left|\Delta\mathbf{x}_0\cdot\mathbf{v}_i\right|^2$. When the second term of Eq. (A15) is expanded, the integrals over the cross terms vanish because they are odd functions of a_j'. Other terms of this expansion involve integrals of the type

$$Z'^{-1}\int d^{3N}\mathbf{a}'\frac{1}{2k_BT}\omega_i^2\left|a_i'\right|^2 e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i'^2|a_i'|^2} = \frac{1}{2}\frac{\omega_i^2}{\omega_i'^2}. \tag{A16}$$

The integral over the third term in Eq. (A15) vanishes because it is a sum of odd functions in a_j'. Substituting these expressions for the integrals in the expansion of Eq. (7) yields Eq. (9) in the manuscript.

Comparison of equivalent normal modes solutions

We expect the value of $D_{\mathbf{x}}^{\mathbf{v}}$ (and $D_{\mathbf{x}}$) to be zero for equivalent systems with different eigenvectors, as we illustrate here for a special case. Consider a harmonic system in

which two modes A and B are degenerate, *i.e.*, they both have the same eigenvalue ω^2. For any normal modes solution with eigenvectors $\mathbf{v}_A$ and $\mathbf{v}_B$, there exists a family of equivalent solutions with eigenvectors $\mathbf{v}'_A$ and $\mathbf{v}'_B$ defined as follows:

$$\begin{aligned}\mathbf{v}'_A &= a_{AA}\mathbf{v}_A + a_{AB}\mathbf{v}_B \\ \mathbf{v}'_B &= a_{BA}\mathbf{v}_A + a_{BB}\mathbf{v}_B\end{aligned}, \tag{A17}$$

where

$$a_{AA}^2 + a_{AB}^2 = a_{BA}^2 + a_{BB}^2 = 1, \tag{A18}$$

and

$$a_{AA}a_{BA} + a_{AB}a_{BB} = 0. \tag{A19}$$

Although the eigenvectors of these solutions may be significantly different, because ω'^2_A and ω'^2_B both equal ω^2, and because of Eqs. A18 and A19, $D^{\mathbf{v}}_{\mathbf{x}A}$ and $D^{\mathbf{v}}_{\mathbf{x}B}$ are both zero:

$$D^{\mathbf{v}}_{\mathbf{x}A} = \frac{1}{2}\left(\sum_{j=1}^{3N}\frac{\omega_j^2}{\omega'^2_A}\left|\mathbf{v}'_A\cdot\mathbf{v}_j\right|^2 - 1\right) = \frac{1}{2}\left(\frac{\omega^2}{\omega^2}\left|\mathbf{v}'_A\cdot\mathbf{v}_A\right|^2 + \frac{\omega^2}{\omega^2}\left|\mathbf{v}'_A\cdot\mathbf{v}_B\right|^2 - 1\right) = 0, \tag{A20}$$

and similarly for $D^{\mathbf{v}}_{\mathbf{x}B}$. Because all modes except A and B are identical, all other terms of $D^{\mathbf{v}}_{\mathbf{x}}$ are also zero, and the value of $D^{\mathbf{v}}_{\mathbf{x}}$ is therefore zero, as expected. Because the eigenvalues and mean conformations are also equivalent, by Eq. (9), the value of $D_{\mathbf{x}}$ is also zero.

REFERENCES

1. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. J Mol Biol 1965;12:88-118.
2. Hill AV. The possible effects of the aggregation of the molecules of haemoglobin on its oxygen dissociation. J Physiol 1910;40:iv-vii.
3. Perutz MF. X-ray analysis of hemoglobin. Science 1963;140:863-869.
4. Perutz MF, Bolton W, Diamond R, Muirhead H, Watson HC. Structure of haemoglobin. An X-ray examination of reduced horse haemoglobin. Nature 1964;203:687-690.
5. Koshland DE, Jr., Nemethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 1966;5:365-385.
6. Perutz MF. Mechanisms of cooperativity and allosteric regulation in proteins. Q Rev Biophys 1989;22:139-237.
7. Strater N, Schnappauf G, Braus G, Lipscomb WN. Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures. Structure 1997;5:1437-1452.
8. Weber G. Ligand binding and internal equilibria in proteins. Biochemistry 1972;11:864-878.
9. Jacrot B, Cusack S, Dianoux AJ, Engelman DM. Inelastic neutron scattering analysis of hexokinase dynamics and its modification on binding of glucose. Nature 1982;300:84-86.

10. Middendorf HD. Biophysical applications of quasi-elastic and inelastic neutron scattering. Annu Rev Biophys Bioeng 1984;13:425-451.

11. Cooper A, Dryden DT. Allostery without conformational change. A plausible model. Eur Biophys J 1984;11:103-109.

12. Luque I, Leavitt SA, Freire E. The linkage between protein folding and functional cooperativity: two sides of the same coin? Annu Rev Biophys Biomol Struct 2002;31:235-256.

13. Kern D, Zuiderweg ER. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748-757.

14. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of *all* dynamic proteins? Proteins 2004;57:433-443.

15. Frauenfelder H, Wolynes PG. Rate theories and puzzles of hemeprotein kinetics. Science 1985;229:337-345.

16. Kullback S, Leibler RA. On information and sufficiency. Annals of Math Stats 1951;22:79-86.

17. Harata K, Muraki M. X-ray structure of turkey-egg lysozyme complex with tri-N-acetylchitotriose. Lack of binding ability at subsite A. Acta Crystallogr D Biol Crystallogr 1997;53:650-657.

18. Brooks BR, Bruccoleri R, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J Comput Chem 1983;4:187-217.

19. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 1983;80:6571-6575.

20. Brooks BR, Janežic D, Karplus M. Harmonic analysis of large systems. I. Methodology. J Comput Chem 1995;16:1543-1553.

21. Bruccoleri RE, Karplus M, McCammon JA. The hinge-bending mode of a lysozyme-inhibitor complex. Biopolymers 1986;25:1767-1802.

22. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38:305-320.

23. Brooks B, Karplus M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc Natl Acad Sci U S A 1985;82:4995-4999.

24. Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 1985;181:423-447.

25. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci U S A 1983;80:3696-3700.

26. McLachlan AD. Gene duplications in the structural evolution of chymotrypsin. J Mol Biol 1979;128:49-79.

27. Phillips DC. The three-dimensional structure of an enzyme molecule. Sci Am 1966;215:78-90.

28. Chipman DM, Sharon N. Mechanism of lysozyme action. Science 1969;165:454-465.

29. Acharya KR, Stuart DI, Walker NP, Lewis M, Phillips DC. Refined structure of baboon alpha-lactalbumin at 1.7 A resolution. Comparison with C-type lysozyme. J Mol Biol 1989;208:99-127.

30. Garcia AE, Onuchic JN. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci U S A 2003;100:13898-13903.

31. Caspar DL, Clarage J, Salunke DM, Clarage M. Liquid-like movements in crystalline insulin. Nature 1988;332:659-662.

32. Clarage JB, Phillips GN, Jr. Analysis of diffuse scattering and relation to molecular motion. Methods Enzymol 1997;277:407-432.

33. Mizuguchi K, Kidera A, Go N. Collective motions in proteins investigated by X-ray diffuse scattering. Proteins 1994;18:34-48.

34. Wall ME, Clarage JB, Phillips GN. Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. Structure 1997;5:1599-1612.

35. Wall ME, Ealick SE, Gruner SM. Three-dimensional diffuse X-ray scattering from crystals of *Staphylococcal* nuclease. Proc Natl Acad Sci U S A 1997;94:6180-6184.

36. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC. Dynamics of ligand binding to myoglobin. Biochemistry 1975;14:5355-5373.

37. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates. Proc Natl Acad Sci U S A 2002;99:8620-8625.

38. Weber G. Energetics of ligand binding to proteins. Adv Protein Chem 1975;29:1-83.

39. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci 1995;20:374.

FIGURES

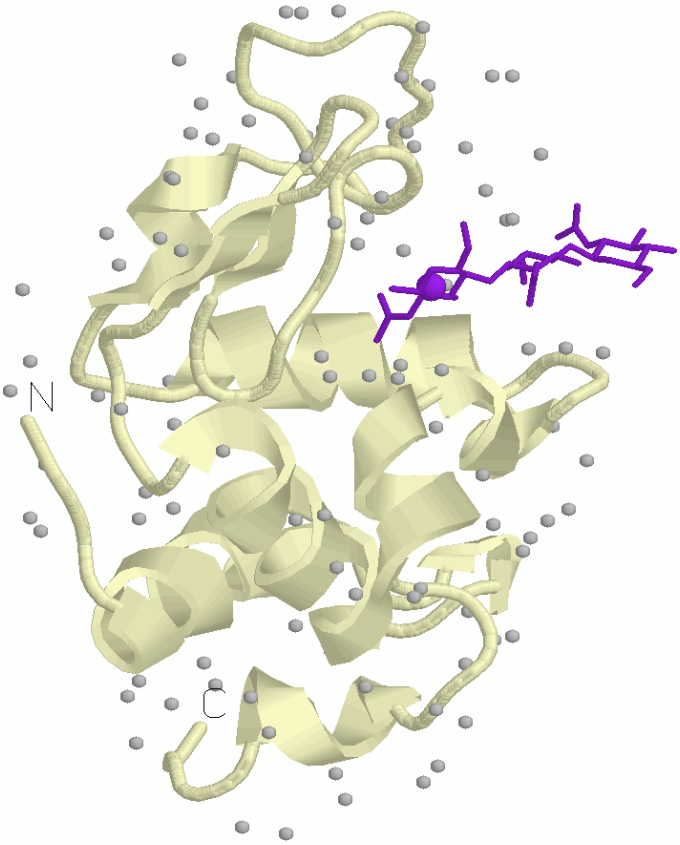

Figure 1: Illustration of the lysozyme/tri-NAG complex (PDB code 1JEF). Lysozyme is depicted as a yellow backbone ribbon, and the tri-NAG is depicted using a purple wire-frame; the center of the deepest-buried NAG monomer is indicated by a purple sphere. The protein is decorated with gray spheres located at 111 automatically generated surface points at which NAG was placed for calculation of the allosteric potential for random ligand interactions. The figure was created using RASMOL.[39]

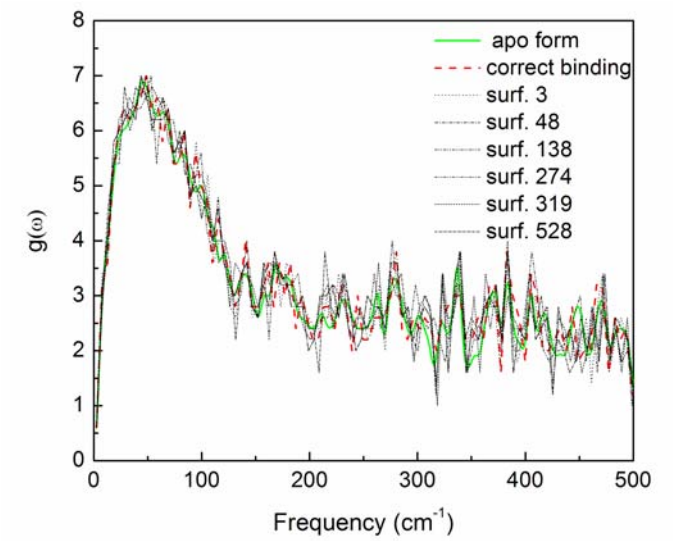

Figure 2: The frequency spectra of apo-lysozyme, the natural lysozyme/NAG complex, and some complexes with arbitrary lysozyme/NAG interactions. Values are densities determined by counting the number of frequencies in a window of 5 cm^{-1} width, centered on the frequency value of the x-axis, and normalizing by the width (division by 5 cm^{-1}).

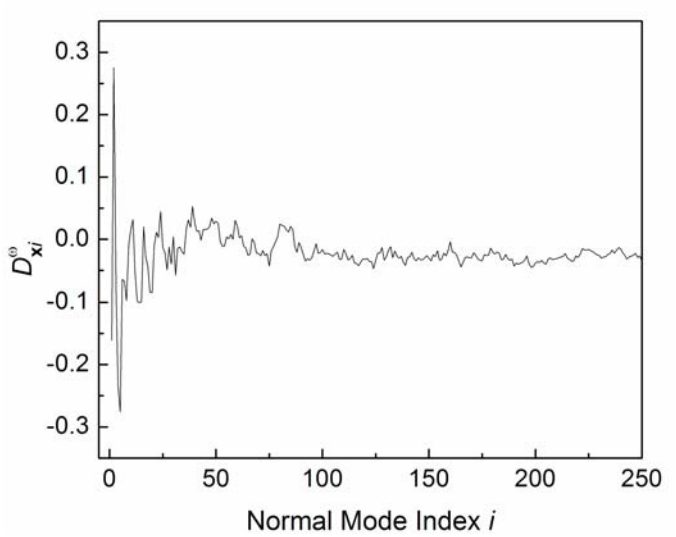

Figure 3. The single-mode contribution to the frequency-spectrum component of the allosteric potential, D^{ω}_{xi}, *vs*. mode index *i* for the natural lysozyme/NAG complex.

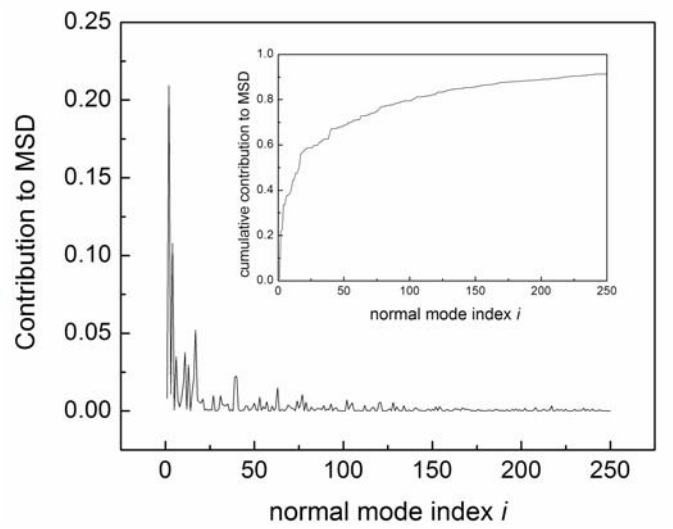

Figure 4. Contributions to the mean-squared displacement (MSD) of Cα atoms between the apo-protein and the natural lysozyme/NAG complex for the 250 lowest-frequency eigenvectors of the apo-protein. Each contribution is calculated as $\left|\Delta\mathbf{x}_0^{(\alpha)} \cdot \mathbf{v}_i^{(\alpha)}\right|^2$, where $\Delta\mathbf{x}_0^{(\alpha)}$ is the displacement of the Cα atoms, and $\mathbf{v}_i^{(\alpha)}$ is the Cα subspace of eigenvector *i*. The total MSD is equal to $\left|\Delta\mathbf{x}_0^{(\alpha)}\right|^2$. The inset shows the cumulative sum of the contributions. Most of the displacement is accounted for by space of the first 50 eigenvectors, and the cumulative contribution after mode 100 (about 3700 modes) is less than 20%. Results are similar for complexes with random ligand interactions. Only the first 100 modes were used to calculate $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ for each complex.

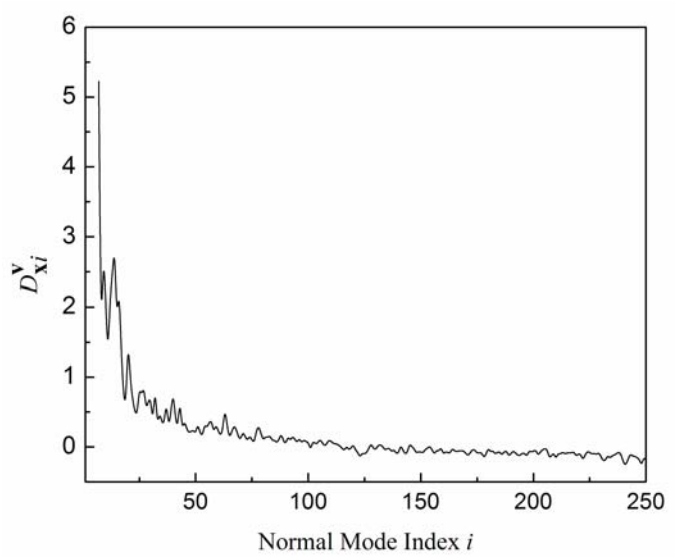

Figure 5: The single-mode contribution to the eigenvector contribution, D^{v}_{xi}, vs. mode index *i* for the natural lysozyme/NAG complex.

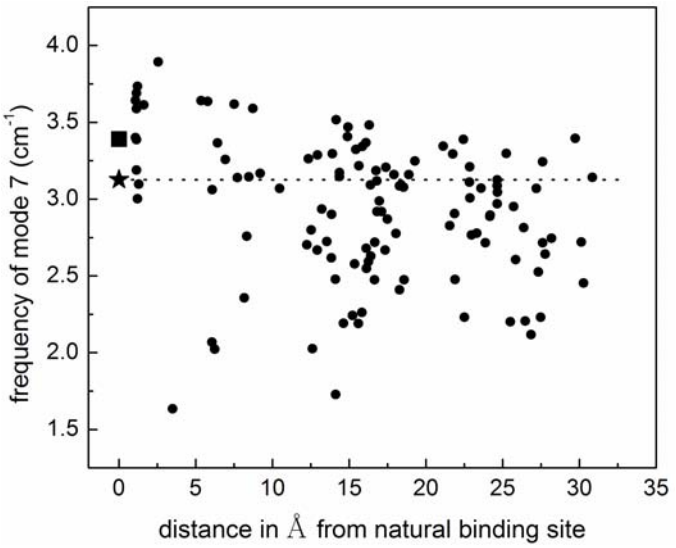

Figure 6. Distribution of the frequency of very lowest-frequency mode (mode 7) of apo-form lysozyme, the natural lysozyme/NAG complex, and complexes with random binding-sites. The box point represents the apo-protein (3.39 cm^{-1}), the star point represents the natural lysozyme/NAG complex (3.13 cm^{-1}), and the dots represent complexes with random NAG binding sites.

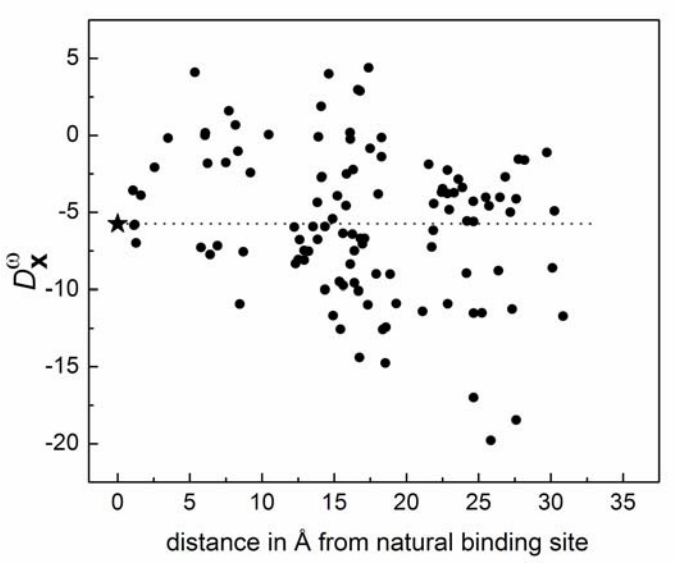

Figure 7: Values of the frequency contribution, $D_{\mathbf{x}}^{\omega}$, vs. the distance to the natural binding site. The star represents the natural lysozyme/NAG complex, and the dots represent complexes with random NAG binding sites.

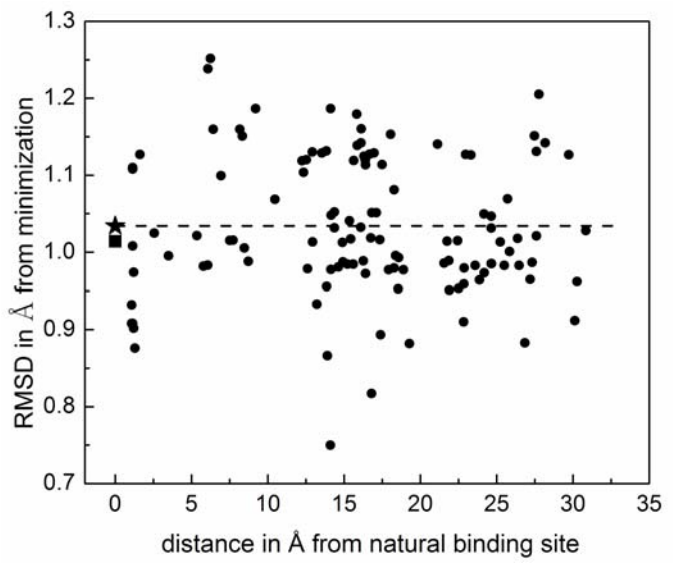

Figure 8. Protein backbone RMSDs between structures before and after minimization. The box point represents the apo-protein (1.01Å), the star represents the natural lysozyme/NAG complex (1.03 Å), and the dots represent complexes with random NAG binding sites.

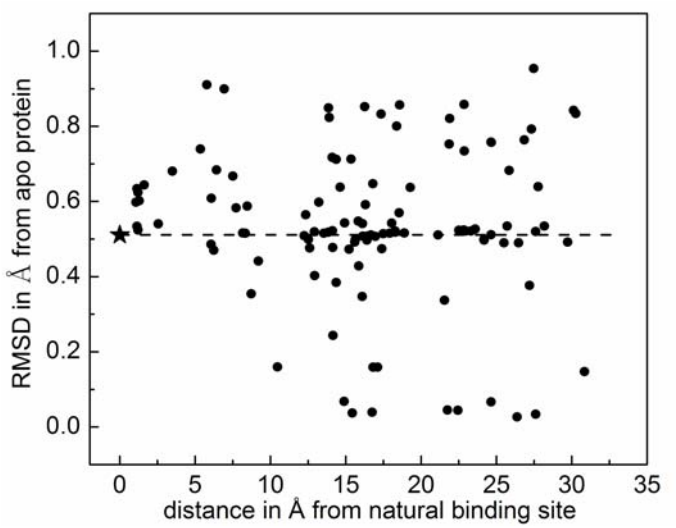

Figure 9. Protein backbone RMSDs for the minimized natural lysozyme/NAG complex and complexes with random binding sites, using the apo-protein as a reference. The star represents the natural lysozyme/NAG complex (0.51 Å), and the dots represent complexes with random NAG binding sites.

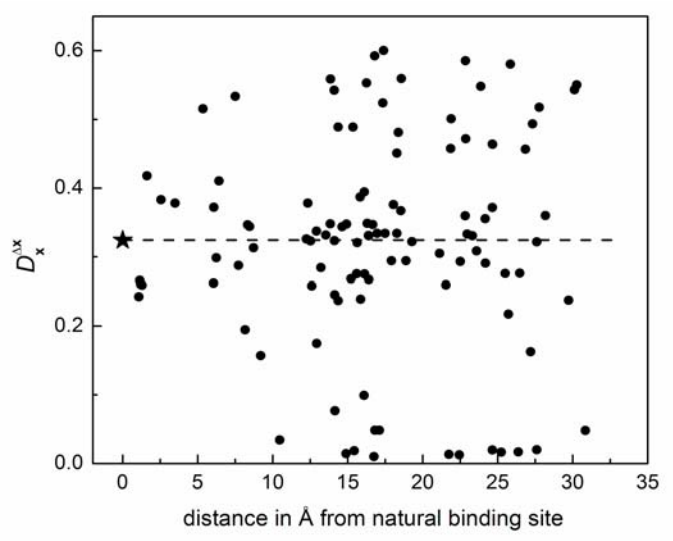

Figure 10: Values of the mean conformation contribution, $D_{\mathbf{x}}^{\Delta \mathbf{x}}$, vs. the distance to the natural binding site. The star represents the natural lysozyme/NAG complex, and dots represent complexes with random NAG binding sites.

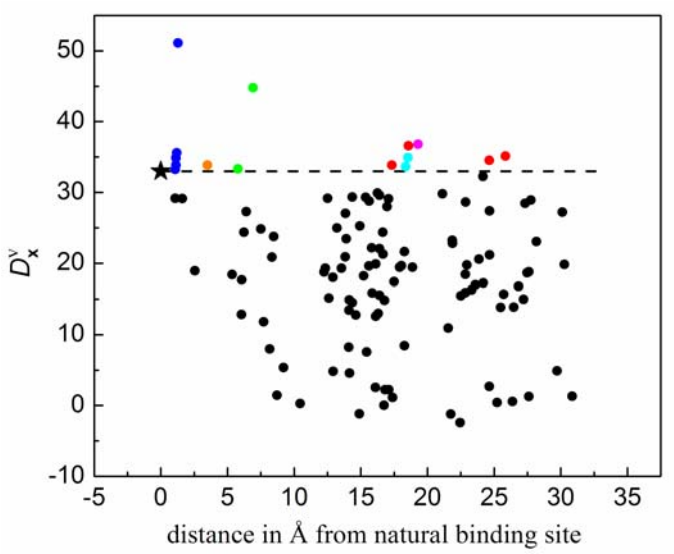

Figure 11: The distribution of the eigenvector contribution, D_x^v, vs. the distance between the test point and the natural binding site. The star indicates the natural lysozyme/NAG complex. Colored points have values at least as large as that of the natural lysozyme/NAG complex. Blue points (7 in total) are in the immediate neighborhood of the natural ligand-binding site – all of the blue points were manually selected. Locations of all other points were generated automatically as described in the text.

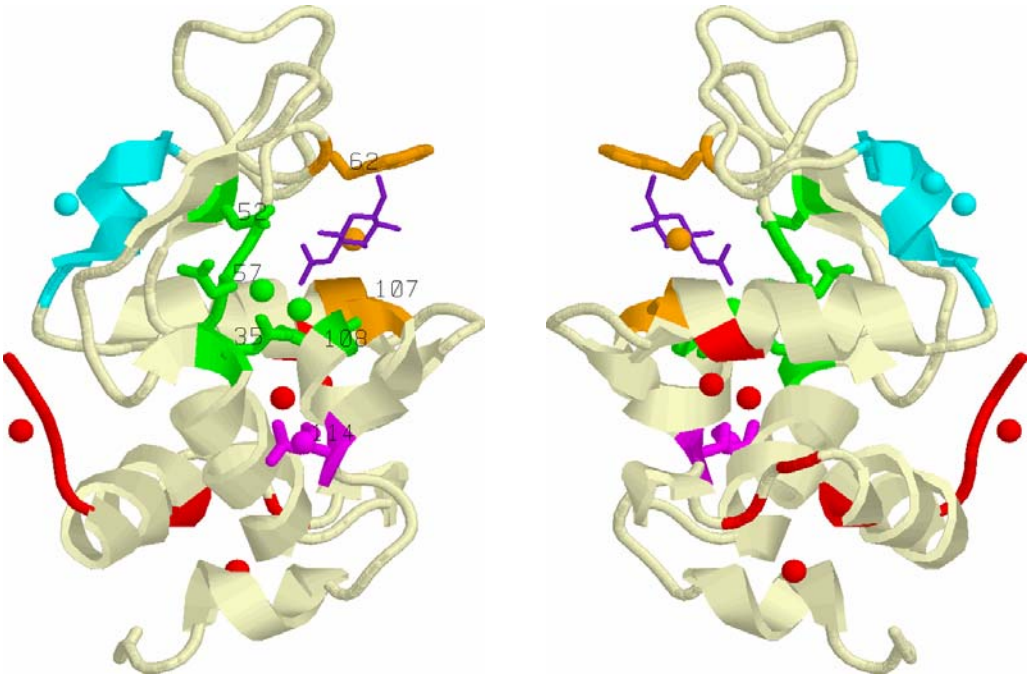

Figure 12. Visualization of sites near automatically generated test points that have a relatively large value of D_x^v. The image at the right is rotated 180° about the y-axis with respect to the image at the left. The deepest-buried NAG monomer is depicted using a purple wire-frame. Automatically generated test points from Fig. 11 are indicated as colored spheres. Residues in PDB entry 1JEF for which the Cα atom is within 6 Å of an automatically generated test point are painted using the color of the test point: green (35, 52, 56-58, 107-109), orange (62, 98, 107), cyan (79-85), magenta (114), and red (1-4, 13, 16, 18, 96). Residue 107 is near both a green and orange point and is painted orange. Residues among these that are known to be important to the function of lysozyme (see text) are labeled and rendered using wire-frames. The figure was created using RASMOL.[39]