

Iterative cluster-NMA: A tool for generating conformational transitions in proteins

Adam D. Schuyler,¹ Robert L. Jernigan,² Pradman K. Qasba,³ Boopathy Ramakrishnan,^{3,4} and Gregory S. Chirikjian^{5*}

¹ Department of Neurology, University of Michigan, Ann Arbor, Michigan 48109

² L.H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa 50011

³ Structural Glycobiology Section, CCR Nanobiology Program, Frederick, Maryland 21702

⁴ Basic Science Program, SAIC-Frederick, Inc., CCR, NCI-Frederick, Frederick, Maryland 21702

⁵ Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT

Computational models provide insight into the structure–function relationship in proteins. These approaches, especially those based on normal mode analysis, can identify the accessible motion space around a given equilibrium structure. The large magnitude, collective motions identified by these methods are often well aligned with the general direction of the expected conformational transitions. However, these motions cannot realistically be extrapolated beyond the local neighborhood of the starting conformation. In this article, the iterative cluster-NMA (icNMA) method is presented for traversing the energy landscape from a starting conformation to a desired goal conformation. This is accomplished by allowing the evolving geometry of the intermediate structures to define the local accessible motion space, and thus produce an appropriate displacement. Following the derivation of the icNMA method, a set of sample simulations are performed to probe the robustness of the model. A detailed analysis of β 1,4-galactosyltransferase-T1 is also given, to highlight many of the capabilities of icNMA. Remarkably, during the transition, a helix is seen to be extended by an additional turn, emphasizing a new unknown role for secondary structures to absorb slack during transitions. The transition pathway for adenylate kinase, which has been frequently studied in the literature, is also discussed.

Proteins 2009; 74:760–776.
© 2008 Wiley-Liss, Inc.

Key words: protein mechanics; elastic network; normal mode analysis; cluster-NMA; rigid-body motions; β 1,4-galactosyltransferase-T1; adenylate kinase.

INTRODUCTION

Experimental methods have produced a wealth of high-resolution protein structures. At the time of writing, the Protein Data Bank (PDB¹) contains more than 40,000 structures solved by X-ray crystallography, nearly half of which are at a resolution better than 2 Å. This detailed information has provided a solid foundation for understanding the structure–function relationship, which involves the determination of biological function from structure. This article presents a new method, called iterative cluster normal mode analysis (icNMA), which produces a transition pathway between known conformations by following the local accessible motion space of the evolving conformation. The resulting pathway provides insight into motion-driven biological function.

Numerous experimental methods have been employed in the study of macromolecular structure and dynamics, including fluorescent resonance energy transfer,² nuclear magnetic resonance,^{3,4} hydrogen exchange,⁵ and crystallography. These purely experimental methods are often limited by a tradeoff between spatial and temporal resolution. In response to this obstacle, computational modeling techniques have been applied with great success.

Before icNMA is presented, it will be useful to review previous modeling techniques, especially the cluster normal mode analysis (cNMA^{6,7}) method upon which icNMA is based. The first set of models discussed (including cNMA) are primarily concerned with identifying the accessible motion space *local* to a given equilibrium conformation. This first set serves as a foundation for understanding the second set, which are capable of exploring the *extended* energy landscape.

In one of the earliest computational models, Levitt⁸ uses a simplified structure representation and a potential energy function to study protein folding. On the basis of similar concepts, molecular dynamics (MD) methods use complex potential functions and numerical integration to

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant number: R01GM075310

*Correspondence to: Dr. Gregory S. Chirikjian, Department of Mechanical Engineering, The Johns Hopkins University, 223 Latrobe Hall, 3400 North Charles Street, Baltimore, MD 21218.

E-mail: gregc@jhu.edu.

Received 24 March 2008; Revised 13 June 2008; Accepted 17 June 2008

Published online 19 August 2008 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.22200

produce conformation trajectories.^{9,10} These atomic resolution models and all-inclusive potential functions come at a significant computational cost,¹¹ often restricting simulations to timescales orders of magnitude shorter than relevant biological processes.

Normal mode analysis (NMA), which is well suited for identifying large-scale, cooperative structure motions, was introduced.^{12–16} The success of NMA-based methods can be attributed to the robustness and simplicity of motions around the equilibrium conformation.^{17,18} The first NMA models were still based on complex potential functions and thus presented computational limitations. This was addressed by the elastic network model (ENM¹⁹), which proposed the usage of a single parameter harmonic potential function to model all pairwise atomic interactions. The Gaussian network model^{20,21} used a coarse-grained model relying only upon the α -carbon trace representation of a protein along with the ENM to produce a measure of atomic mobility. This scalar model was then extended into the anisotropic network model (ANM²²) to capture magnitude and direction of atomic fluctuations.

The already simplified ANM representation has been even further reduced through the use of coarse graining, in which multiple atoms are grouped into single representative points. Excellent reviews of coarse-grained models are given by Tozzini²³ and Bahar and Rader.¹⁷ A comparison between varying levels of grain resolution is given by Sen *et al.*²⁴ Although these “bead models” do succeed in reducing the number of degrees of freedom (DOFs) in the structure representation, they do so at the expense of altering atomic interaction geometries (i.e. when multiple atoms are represented by a single point, all of their distinct contact interactions must be collapsed onto that representative point).

In response to the shortcomings of coarse-grained NMA, the authors developed cNMA^{6,7} in which groups of atoms are represented as rigid bodies embedded in the ENM. This approach utilizes a multiscale structure representation which includes all atoms and all pairwise atomic interactions (i.e. no deformed geometries), while at the same time, reduces the total number of DOFs needed in the parameterization. The cNMA method is the foundation for icNMA and is reviewed in Section “Review of cNMA”. The complete derivation and comparison of cNMA to ENM-based C_α -NMA is given in Schuyler and Chirikjian⁶ and its application at atomic resolution to large structures is given in Schuyler and Chirikjian.⁷ The RTB method^{25,26} is similar to cNMA, but several substantial differences in coordinate system definitions and computational procedures are discussed in Schuyler and Chirikjian.⁷ Clustering schemes have also been incorporated into MD methods²⁷ and into NMA methods which explicitly define solvent.²⁸

Proteins are known to sample various conformation states essential for carrying out their functions. These

states range from relatively minor changes in structure to large scale rearrangements, such as those seen in hemoglobin and myosin. There have been many attempts at understanding these transitions.^{29–32} It has been established that relatively few, low-frequency normal modes can identify the direction of global motions required to achieve conformational transitions.³³ This general categorization has been further refined by a database study that relates the degree of collectivity in a transition with the effectiveness of ENM normal modes to capture the transition direction.^{34,35} Motion correlation analysis across the low frequency modes has provided information on cooperative structure motion and domain stability.³⁶

The methods discussed thus far are only relevant to the local motion subspace around an equilibrium conformation. To achieve a full understanding of biologically relevant structure transitions, we now introduce a second class of methods, which are capable of exploring the extended energy landscape.

Cryo-electron microscopy (cryo-EM) has been used to study conformational changes.³⁷ The identification of secondary structure elements has been used to fit atomic resolution subunits onto low-resolution, cryo-EM electron density maps of large complexes.³⁸ There has also been success in coupling cryo-EM with NMA-based methods. The continuous valued density map has been converted into discrete mass locations, which serve as the basis for low-resolution NMA. This approach shows agreement with the lowest modes of atomic resolution NMA.³⁹ There have also been methods for deforming known crystal structures onto low-resolution cryo-EM electron density maps.^{40–42} These methods use cost functions to optimally select a subset of normal modes and assign relative weighting coefficients that best produce the desired conformational changes. These methods start to introduce dynamics into the structure analysis, but it must be noted that the series of conformations converging to the best fit is not intended as (and can not be interpreted as) a pathway, but as a byproduct of the fitting process.

Linear interpolation of Cartesian coordinates between known conformation pairs causes obvious steric violations, but modified interpolation methods are able to avoid this. For example, the elastic network interpolation (ENI) method interpolates between contact maps of two conformations.^{43,44} One particular aspect of this and many of the other models for generating transition pathways is that they lead to a single pathway; this neglects the widely perceived stochastic nature of protein folding. Consequently these single pathway methods likely provide a most probable representative pathway, with variations around these single pathways providing a full ensemble representation. For example, ENI is able to fit a pathway generated for a core central domain of the 16S rRNA onto an MD trajectory using only the lowest 1% of the normal modes.⁴⁵ Normal modes are also able to

guide conformational changes according to a small set of distance constraints⁴⁶ or an X-ray diffraction pattern.⁴⁷ Yang *et al.*³⁴ investigate the transition pathways of 170 pairs of structures and find that the normal modes succeed in specifying the directionalities of the transitions only when the collectivity of the motion is high.

The undesired, but inherent, linearity of the interpolation methods is addressed by a very successful family of methods that use switching functions to allow for the transition pathway to shift from the energy basin of one conformation to that of another.^{48–54} The plastic network model (PNM⁵⁰) uses an ENM to define energy basins around each known state and then solves for the saddle point located at the global minimum of their intersection. The transition pathway connecting the saddle point to the neighboring equilibrium states is produced by the TReK steepest descent method in CHARMM.⁵⁵ Of particular interest are that the PNM assumes the globally optimal transition state is accessible from the equilibrium states and that the connecting pathways from the saddle point down to each equilibrium state are the preferred pathways in the reverse direction. Remarkably, the PNM pathway for adenylate kinase is consistent with intermediate crystal structures. In a similar way, Yang *et al.*⁵⁶ find good agreement between the normal modes and the conformational variations observed across 156 X-ray structures and 28 NMR structures (reported in one set). Notably, these sets include unbound structures as well as ones having different ligands.

The cryo-EM methods and the interpolation methods give insight into the dynamics of large conformational changes, but they are based on either forced motions or artificial sampling of mode space. These issues are addressed by icNMA, which uses the efficient, all atom cNMA method to guide the transition pathway according to the motion space described by each intermediate conformation. The remainder of the article is structured as follows. A Hamiltonian mechanics foundation for normal mode analysis is presented in a preliminary section. This background simplifies the icNMA derivation in the method section and relates the icNMA method to several fundamentals of statistical mechanics allowing for a more powerful interpretation of the results. An analysis of β 1,4-galactosyltransferase-T1 is given so that further details of icNMA may be discussed in context. A Q1 versus Q2 plot is given for the frequently studied adenylate kinase; it shows nonlinearity of the transition path, which is in agreement with other methods.^{50,51,57,58}

NORMAL MODE ANALYSIS DERIVED FROM HAMILTONIAN MECHANICS

The icNMA method is based on a few fundamental theories from statistical mechanics. These concepts are

presented and will allow for a more direct formulation of the icNMA method.

For small motions about an equilibrium, the potential energy of a biomolecular structure is parameterized by its generalized coordinates, \mathbf{q} , and is written as

$$V(\mathbf{q}) = C + \frac{1}{2} \mathbf{q}^T K \mathbf{q} \quad (1)$$

where C is a constant (which can be ignored by appropriate choice of the datum in the definition of potential energy) and K is the stiffness (or Hessian) matrix. This quadratic potential is nothing more than the first few terms in the Taylor-series expansion of the molecular potential, where the linear term drops out from the definition of equilibrium (i.e. $\partial V / \partial q_i = 0$ for all values of i).

The Hamiltonian of the system is defined as

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \mathbf{p}^T [M^{-1}(\mathbf{q})] \mathbf{p} + V(\mathbf{q}) \quad (2)$$

where \mathbf{p} is the vector of all conjugate momenta corresponding to the generalized coordinates and $M(\mathbf{q})$ is the mass matrix with the “-1” exponent denoting its inverse. Depending on the choice of coordinates, the mass matrix can be reduced to a constant, M_0 , for small deformations around an equilibrium. This is demonstrated in Schuyler and Chirikjian⁶ for cluster coordinates, which are used by icNMA.

The foundation for normal mode analysis is more easily derived from the Hamiltonian by performing the following coordinate transforms

$$\tilde{\mathbf{q}} = M_0^{1/2} \mathbf{q} \quad (3)$$

$$\tilde{\mathbf{p}} = M_0^{-1/2} \mathbf{p} \quad (4)$$

where the exponent of “1/2” indicates a matrix square root. The Hamiltonian is expressed in these coordinates as

$$\tilde{H}(\tilde{\mathbf{q}}, \tilde{\mathbf{p}}) = \frac{1}{2} \tilde{\mathbf{p}}^T \tilde{\mathbf{p}} + \frac{1}{2} \tilde{\mathbf{q}}^T \tilde{K} \tilde{\mathbf{q}} \quad (5)$$

where we have defined the mass-weighted Hessian as

$$\tilde{K} = M_0^{-1/2} K M_0^{-1/2} \quad (6)$$

The Boltzmann distribution describes the accessibility of all states in an equilibrium ensemble and is given as the probability density function on phase space

$$f(\tilde{\mathbf{q}}, \tilde{\mathbf{p}}) = \frac{1}{Z(\beta)} \exp\{-\beta \tilde{H}(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})\} \quad (7)$$

where

$$Z(\beta) = \int_{\tilde{\mathbf{q}}} \int_{\tilde{\mathbf{p}}} \exp\{-\beta \tilde{H}(\tilde{\mathbf{q}}, \tilde{\mathbf{p}})\} d\tilde{\mathbf{p}} d\tilde{\mathbf{q}} \quad (8)$$

is the partition function, and $\beta = 1/k_B T$ (k_B is Boltzmann's constant and T is temperature measured in degrees Kelvin).

In the context of generating conformational ensembles, it is more desirable to have a distribution over only the generalized coordinate ($\tilde{\mathbf{q}}$) and not over the conjugate momenta ($\tilde{\mathbf{p}}$). The constant mass matrix has effectively decoupled the $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{p}}$ terms in the Hamiltonian and allows the Boltzmann distribution to be integrated in closed form over $\tilde{\mathbf{p}}$ yielding a Gaussian distribution of conformations

$$\rho(\tilde{\mathbf{q}}) = \int_{\tilde{\mathbf{p}}} f(\tilde{\mathbf{q}}, \tilde{\mathbf{p}}) d\tilde{\mathbf{p}} \quad (9)$$

$$= \hat{Z}(\beta) \exp\{-\beta \tilde{V}(\tilde{\mathbf{q}})\} \quad (10)$$

where the integration over the kinetic energy portion of the Hamiltonian is incorporated into the scaling factor

$$\hat{Z}(\beta) = \frac{1}{Z(\beta)} \int_{\tilde{\mathbf{p}}} \exp\left\{-\frac{\beta}{2} \tilde{\mathbf{p}}^T \tilde{\mathbf{p}}\right\} d\tilde{\mathbf{p}} \quad (11)$$

and the remaining portion of the Hamiltonian is the potential energy in mass-weighted coordinates

$$\tilde{V}(\tilde{\mathbf{q}}) = \frac{1}{2} \tilde{\mathbf{q}}^T \tilde{K} \tilde{\mathbf{q}} \quad (12)$$

The distribution in Eq. (10), indicates that the most populated conformation states are those whose displacements from the equilibrium ($\tilde{\mathbf{q}} = \mathbf{0}$) correspond to the lowest potential energies as defined by Eq. (12). These conformation displacements are more easily identified by projecting the mass-weighted generalized coordinate onto a new basis defined by the solutions to the eigenproblem

$$\tilde{K} \tilde{\mathbf{v}}_i = \tilde{\lambda}_i \tilde{\mathbf{v}}_i \quad (13)$$

The significance of the eigenvectors ($\{\tilde{\mathbf{v}}_i\}$, unit length by convention) and eigenvalues ($\{\tilde{\lambda}_i\}$) becomes apparent when considering the forces and energies associated with conformation displacements along the eigenvector axes.

The restoring force opposing displacements away from the equilibrium state is

$$\mathbf{f}_R(\tilde{\mathbf{q}}) = -\frac{\partial \tilde{V}(\tilde{\mathbf{q}})}{\partial \tilde{\mathbf{q}}} = -\tilde{K} \tilde{\mathbf{q}} \quad (14)$$

Evaluating this force for a unit magnitude displacement along one of the eigenvectors produces

$$\mathbf{f}_R(\tilde{\mathbf{v}}_i) = -\tilde{K} \tilde{\mathbf{v}}_i = -\tilde{\lambda}_i \tilde{\mathbf{v}}_i \quad (15)$$

which indicates that the restoring force acts along the same axis as the displacement, but in the opposite direc-

tion. The eigenvectors define the axes of harmonic oscillations around the equilibrium state and the eigenvalues are the corresponding squared frequencies.

The potential energy associated with a unit magnitude displacement along one of the eigenvectors is

$$\tilde{V}(\tilde{\mathbf{v}}_i) = \frac{\tilde{\lambda}_i}{2} \quad (16)$$

This simple relationship complements Eq. (10), and indicates that the most populated conformation states are reached by displacements along the low index (i.e. low energy) eigenvectors. \tilde{K} is symmetric so the eigenvectors, also referred to as normal modes or mode shapes, are pairwise orthogonal and define a basis for all conformation motions around the equilibrium state. The eigenvector basis and its relationship to the system's potential energy are the foundation for all NMA methods and will be referenced in the icNMA method section.

REVIEW OF cNMA

The reader is referred to the original publications^{6,7} for the full derivation and application of cNMA. The following section restates the interaction model and the generalized coordinates, which are both used during the formulation of icNMA.

A structure of n atoms is represented as N rigid bodies (clusters of atoms). The harmonic potential defined in Eq. (1) is produced by defining an ENM, in which the clusters are interconnected by a network of springs with an atomic cutoff distance of $r_c = 5 \text{ \AA}$. No springs are defined between atoms within a cluster, so this ENM is a subset of the traditional atomistic ENM.

The clustering can be defined on a per residue basis, which is the highest resolution application of cNMA, but most costly at $\mathcal{O}(n^3)$; it can be defined according to domain, chain or subunit, which break the dependence of N on n , allowing the cNMA method to achieve $\mathcal{O}(n)$ computational complexity; or it can be defined by some combination of these options as a hybrid multiscale model. As an alternative to these hierarchically-based clustering schemes, studies of protein flexibility and rigidity have been conducted^{59,60} and incorporated into coarse-grained models.⁶¹ The Vishveshwara group developed a clustering algorithm based on graph spectral analysis⁶² and have used other graph theoretic methods to study structure connectivity.⁶³

Regardless of the clustering algorithm used with cNMA, a structure's conformation is defined by the position and orientation of the N embedded rigid bodies. The translational motion of cluster c is measured by the displacement of its center of mass, \mathbf{x}_c , with respect to its location in the initial equilibrium state, \mathbf{x}_c^I , as defined by

$$\chi_c = \mathbf{x}_c - \mathbf{x}_c^I \quad (17)$$

The rotational displacement of cluster c is parameterized by the axis-angle vector $\boldsymbol{\gamma}_c \in \mathbb{R}^3$. This vector can be expressed as $\boldsymbol{\gamma}_c = \theta_c \cdot \hat{\mathbf{a}}_c$, where $\hat{\mathbf{a}}_c$ is the normalized direction of $\boldsymbol{\gamma}_c$ and θ_c is the magnitude of $\boldsymbol{\gamma}_c$. The rotation matrix corresponding to $\boldsymbol{\gamma}_c$ is expressed with Rodrigues' formula as

$$R(\boldsymbol{\gamma}_c) \triangleq \exp \{J(\boldsymbol{\gamma}_c)\} \quad (18)$$

$$= \mathbb{I}_3 + \sin(\theta_c)J(\hat{\mathbf{a}}_c) + (1 - \cos(\theta_c))J(\hat{\mathbf{a}}_c)^2 \quad (19)$$

where \mathbb{I}_3 is the 3×3 identity matrix and the skew symmetric matrix function, $J: \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$, is defined by

$$J \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \quad (20)$$

Each cluster's generalized coordinates are given by

$$\boldsymbol{\delta}_c = [\boldsymbol{\chi}_c^T, \boldsymbol{\gamma}_c^T]^T \in \mathbb{R}^6 \quad (21)$$

and the whole structure's generalized coordinates are the stacked vector

$$\boldsymbol{\delta} = [\boldsymbol{\delta}_1^T, \dots, \boldsymbol{\delta}_N^T]^T \in \mathbb{R}^{6N} \quad (22)$$

METHOD

Given the initial, \mathcal{C}_I , and final, \mathcal{C}_F , conformations, the transition pathway is a sequence of connecting intermediate conformations, $\{\mathcal{C}_i\}$. The first conformation in this sequence is defined to be the initial conformation: $\mathcal{C}_1 = \mathcal{C}_I$. The remaining pathway progresses towards \mathcal{C}_F and is generated by the following iterative procedure:

1. Perform cNMA on \mathcal{C}_i .
2. Compute the reference direction, $\hat{\boldsymbol{\delta}}$, from \mathcal{C}_i to \mathcal{C}_F .
3. Construct the global motion, \mathbf{g} , from cNMA modes with guidance from reference direction.
4. Generate the next conformation in the pathway by displacing the current conformation according to the global motion. This can be conveniently expressed as $\mathcal{C}_{i+1} = \mathcal{C}_i + \mathbf{g}$.

The cNMA computations of the first step are based on the derivations in Section "Normal Mode Analysis Derived from Hamiltonian Mechanics", and are computed in the cluster coordinates defined in Section "Review of cNMA". The reference direction in the second step and the global motion construction in the third step, are defined in the following sections.

Reference direction

The reference direction points from the current conformation, \mathcal{C}_i , to the desired goal conformation, \mathcal{C}_F . The reference direction is only used to identify whether candidate motions move towards or away from the final conformation.

The translational component of cluster c of the reference direction is calculated as

$$\hat{\boldsymbol{\chi}}_c = \mathbf{x}_c^{\mathcal{F}} - \mathbf{x}_c \quad (23)$$

where $\mathbf{x}_c^{\mathcal{F}}$ and \mathbf{x}_c are the center of mass positions of cluster c in the goal and current intermediate conformations, respectively.

The rotational motion of cluster c of the reference direction is calculated by solving for the rotation that optimally aligns cluster c of the current conformation with cluster c of the final conformation. Each atom's Cartesian coordinates are placed in sequence as a column vector in the matrix $A \in \mathbb{R}^{3 \times n}$ for the current conformation and in matrix $B \in \mathbb{R}^{3 \times n}$ for the final conformation. The rotation matrix

$$\hat{R}_c = [BA^T AB^T]^{\frac{1}{2}} [AB^T]^{-1} \quad (24)$$

is applied to the positions in B and the new atomic positions in the columns of $\hat{R}_c B$ are optimally alignment in RMSD with the corresponding column positions in A . Inverting the relationship defined in Eq. (18) provides a closed form solution for extracting the axis-angle vector, $\hat{\boldsymbol{\gamma}}_c$, from \hat{R}_c .

The cluster's reference direction is defined as

$$\hat{\boldsymbol{\delta}}_c = [\hat{\boldsymbol{\chi}}_c^T, \hat{\boldsymbol{\gamma}}_c^T]^T \in \mathbb{R}^6 \quad (25)$$

and the conformation's reference direction is given by the stacked vector

$$\hat{\boldsymbol{\delta}} = [\hat{\boldsymbol{\delta}}_1^T, \dots, \hat{\boldsymbol{\delta}}_N^T]^T \in \mathbb{R}^{6N} \quad (26)$$

Global motion

The Hamiltonian mechanics analysis in Section "Normal Mode Analysis Derived from Hamiltonian Mechanics" identifies a set of basis motions for the space of conformation displacements. The lowest energy motions from the basis lead to the most populated conformation states in the ensemble representation. This section develops a method for constructing a single conformation displacement from the set of "building block" basis motions. In general, a set of mode shapes are independent oscillations with distinct frequencies, thus precluding their superposition. However, in the case of icNMA, there are two major factors which support the

representation of the mode set by the subspace that it spans rather than as a set of independent motions.

Consider the extreme case where multiple modes have the same eigenvalue. The modes degenerate into an *arbitrary*, pairwise orthogonal set. In this situation, the specific mode shapes are no longer important, they only serve as basis vectors of a subspace. This degeneration is not an all or none property. As eigenvalues become increasingly close, individual mode directions become increasingly more arbitrary. In the context of coarse-grained NMA, the first few low modes, at best, may be sufficiently separated in the frequency domain allowing for analysis of their specific motions.⁶⁴ However, the ensemble of lowest modes is usually quite dense in the frequency domain and is accordingly subject to mode degeneration.

The second major factor contributing to the use of a global motion results from the ENM atomic interaction model. The ENM is a harmonic, “smooth” approximation of the energy surface. Distinguishing between mode shapes in this approximate subspace would be artificial and an over interpretation of the model. Van Wynsberghe and Cui⁶⁵ present an example illustrating the importance of motion ensembles over individual modes. They use a motion correlation analysis and demonstrate that a pair of structural components in the voltage gated ion channel, KvAP, are correlated under the lowest mode, anticorrelated under the second lowest mode and show no correlation under the ensemble of the lowest 76 modes.

The conformation space between equilibrium states is unstable and supportive of an ensemble of trajectories.⁴⁸ In agreement with this, the subspace of low frequency modes is multidimensional and identifies a distribution of conformations. The question that now remains is how to appropriately construct a global motion within the low-mode subspace that advances a trajectory across the energy landscape to the goal conformation.

The derivations in Section “Normal Mode Analysis Derived from Hamiltonian Mechanics” deal with the properties of individual modes. The global motion aims to combine multiple modes into a single representative motion and requires additional consideration. The probability density function in Eq. (10) identifies the most populated conformation states by relating them to their associated potential energies in Eq. (12). This result supports the construction of a global motion from the normal modes derived in Eq. (13). The first six eigenvalues are zero valued and correspond to rigid translation and rotation of the structure; these motions are not included in the global motion. The remaining mode shapes, indexed $\{7, \dots, d\}$, are nonrigid deformations and are included in the global motion. The normal modes are derived in a mass-weighted coordinate system, but it is more appropriate to express the global motion in the non-mass-weighted cluster coordinates. This is accom-

plished by reversing the coordinate change in Eq. (3) with the transform

$$\mathbf{v}_i = M_0^{-1/2} \tilde{\mathbf{v}}_i \quad (27)$$

The equipartition theorem states that the potential energy of a system is distributed equally, on average, across each of the system’s degrees of freedom. The normal mode solutions to Eq. (13) define a basis for which each mode shape represents one degree of freedom. Conformation displacements in the direction of each mode shape must produce a constant valued energy, when averaged over the pathway. This is achieved by defining the global motion as

$$\mathbf{g} = \sum_{i=7}^d \frac{w_i}{\sqrt{\lambda_i}} \mathbf{v}_i \quad (28)$$

where each mode is scaled by two terms.

The $1/\sqrt{\lambda_i}$ scaling is a foundation that produces *exactly* equal energies for all modes [Eq. (16)]. The inverse frequency scaling corresponds to the physical interpretation of “soft” lower frequency modes moving along shallower slopes of the energy basin than the “stiff” higher frequency modes. The lower frequency modes achieve a greater magnitude of displacement than the higher frequency modes under the same fixed energy.

The w_i terms are undetermined weighting factors which allow for variability of each mode’s relative contribution, subject to the following: (i) The equipartition theorem requires the average energy contribution from each mode to be uniform (i.e. $\langle |w_i| \rangle = \text{const}$), and (ii) Each of the steps in the transition pathway must be equal in energy so that no part of the pathway is biased by unequal sampling (i.e. $V(\mathbf{g}) = \text{const}$). Both of these constraints are addressed by evaluating the potential energy of an arbitrary global motion.

$$V(\mathbf{g}) = \frac{1}{2} \left(\sum_{i=7}^d \frac{w_i}{\sqrt{\lambda_i}} \mathbf{v}_i \right)^T K \left(\sum_{j=7}^d \frac{w_j}{\sqrt{\lambda_j}} \mathbf{v}_j \right) \quad (29)$$

$$= \frac{1}{2} \left(\sum_{i=7}^d \frac{w_i}{\sqrt{\lambda_i}} \tilde{\mathbf{v}}_i^T \right) \tilde{K} \left(\sum_{j=7}^d \frac{w_j}{\sqrt{\lambda_j}} \tilde{\mathbf{v}}_j \right) \quad (30)$$

$$= \frac{1}{2} \left(\sum_{i=7}^d \frac{w_i}{\sqrt{\lambda_i}} \tilde{\mathbf{v}}_i \right) \cdot \left(\sum_{j=7}^d \frac{w_j}{\sqrt{\lambda_j}} \tilde{K} \tilde{\mathbf{v}}_j \right) \quad (31)$$

$$= \frac{1}{2} \left(\sum_{i=7}^d \frac{w_i}{\sqrt{\tilde{\lambda}_i}} \tilde{\mathbf{v}}_i \right) \cdot \left(\sum_{j=7}^d w_j \sqrt{\tilde{\lambda}_j} \tilde{\mathbf{v}}_j \right) \quad (32)$$

$$= \frac{1}{2} \sum_{i=7}^d \sum_{j=7}^d w_i w_j \frac{\sqrt{\tilde{\lambda}_j}}{\sqrt{\tilde{\lambda}_i}} (\tilde{\mathbf{v}}_i \cdot \tilde{\mathbf{v}}_j) \quad (33)$$

$$= \frac{1}{2} \sum_{i=7}^d w_i^2 \quad (34)$$

The orthogonality of the eigenvectors reduces the dot product in the last parentheses to a delta function, which eliminates the summation over j .

Global motions of equal energy (and a weighting factor normalization constraint) are produced by setting the result of Eq. (34), to a constant

$$\frac{1}{2} \sum_{i=7}^d w_i^2 = s^2 \quad (35)$$

This is an equation for a hypersphere whose radius, $s\sqrt{2}$, sets a constant value for the overall energy of the mode ensemble. An optimization over the w_i values on the hypersphere can be used to construct the global motion that most directly moves to the goal conformation. However, this optimization is too costly for iterative application and does not necessarily satisfy the constraint: $\langle |w_i| \rangle = \text{const}$.

It is difficult to allow the modes to vary in energy along the pathway and still guarantee the final distribution is equal. This is resolved by setting all w_i values to the same constant magnitude. This computational convenience, which also reduces the dimension of the optimization space, is expressed as

$$w_i = s_i \cdot c \quad \begin{cases} s_i = \pm 1 \\ c = s\sqrt{\frac{2}{d-6}} \end{cases} \quad (36)$$

where s_i is the sign of the weighting factor and c is the constant magnitude, which is determined directly from the constant energy constraint [Eq. (35)]. The set of all possible w_i combinations define the 2^{d-6} vertices of a hypercube inscribed within the original hypersphere. This is a tremendous reduction in search space dimensionality, but is still too computationally costly. The alternative employed here is to independently set each s_i value rather than simultaneously optimize over the hypercube vertices. This simplification reduces the computational complexity to a series of $d-6$ comparisons.

The constant magnitude, c , defined in Eq. (36), is factored out of the global motion summation producing the expression

$$\mathbf{g} = c \sum_{i=7}^d \frac{s_i}{\sqrt{\tilde{\lambda}_i}} \mathbf{v}_i \quad (37)$$

The remainder of this section is a discussion of the weighting factor magnitude (c), the summation upper bound (d), the mode shape sign choice (s_i), and the implications of mode shape “addition”.

The weighting factor magnitude, c , is defined in terms of a constant energy level, s , which is distributed equally over the $d-6$ nonrigid modes [Eq. (36)]. Selecting an appropriate energy level is not trivial or even necessarily uniform across structures. A qualitatively equivalent result is achieved by scaling the global motion to produce a pre-determined root-mean-square displacement (RMSD), μ . The reason for this approach is that the interaction model for cNMA is an ENM defined by an atomic separation distance ($r_c = 5\text{\AA}$). The ENM remains valid as long as the conformation displacement is small, ensuring that the local geometries do not change significantly. Restricting the global motion to a “local neighborhood” is more naturally stated with an RMSD constraint than an energy constraint.

The summation upper bound, d , can be reduced from its maximum value (i.e. the total number of modes: $d = 6N$) to isolate the contribution from the lower modes. Before the global motion is scaled with the coefficient c , it has a total magnitude given by: $m(d) = \sum_{i=7}^d \tilde{\lambda}_i^{-1/2}$. Therefore, to achieve a particular percentage, p , of this total magnitude, a new upper bound, \hat{d} , on the summation is defined such that

$$m(\hat{d}) = p \cdot m(d) \quad (38)$$

The distribution of the frequency spectrum results in $\hat{d} < pd$ and often times $\hat{d} \ll pd$.

The mode shapes of cNMA (and all NMA-based techniques) are indications of principal axes of motion, not directions. The general equation for an eigenproblem $A\mathbf{v} = \lambda\mathbf{v}$ indicates that the sign of the mode shape is arbitrary: if $\{\mathbf{v}, \lambda\}$ is a solution, then so is $\{-\mathbf{v}, \lambda\}$. In the context of analyzing a single conformation with an NMA-based method, the sign ambiguity is irrelevant because each mode shape can be visualized in the “plus” and “minus” direction to observe the entire oscillation around the equilibrium state. However, in the context of transition pathway generation, a sign choice must be made. The desired result is a pathway which continues to minimize RMSD from the goal conformation, $\mathcal{C}_{\mathcal{F}}$.

A direct implementation of the RMSD minimization involves displacing the current intermediate conformation in the “plus” and “minus” directions along each mode shape, projecting the conformations from cluster coordinates into full Cartesian coordinates, computing RMSD values to $\mathcal{C}_{\mathcal{F}}$ and choosing the minimum. A significantly more efficient (and equivalent) method is to

compute the reference direction according to Eq. (26), and then a simple dot product: $s_i = \text{sign}(\mathbf{v}_i \cdot \hat{\mathbf{d}})$ indicates which direction to follow.

All that remains of the global mode construction in Eq. (37) is the addition operation. In Cartesian coordinate-based NMA, mode shapes can be unambiguously combined to produce the global motion (i.e. addition of translational displacements is commutative). The cNMA method is based on cluster coordinates which include translational and rotational components. The translational displacements are Cartesian and can be summed. The rotational components are given by axis–angle vectors which represent rotation matrices. After each mode’s rotational displacements have been put into rotation matrix form, the cumulative rotational displacement on each cluster is computed by matrix multiplication, which is not a commutative operation. Because of the computational expense of converting each axis–angle vector and the arbitrary order of multiplication, this process is not desirable.

The small motion assumption which the cNMA method is predicated on, allows the rotational displacement of each mode to be represented by the first-order approximation of Rodrigues’ expression for the rotation matrix [Eq. (18)]: $R(\boldsymbol{\gamma}) \approx \mathbb{I}_3 + J(\boldsymbol{\gamma})$. The composition of two rotation matrices can be simplified as

$$R(\boldsymbol{\gamma}_1)R(\boldsymbol{\gamma}_2) \approx \mathbb{I}_3 + J(\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2) \quad (39)$$

by making use of the facts that (1) the addition of skew symmetric matrices is equivalent to summing the corresponding axis–angle vectors; and (2) the second-order term $J(\boldsymbol{\gamma}_1)J(\boldsymbol{\gamma}_2)$ can be discarded in this first-order calculation. The matrix multiplication is reduced to the addition of axis–angle vectors, which is a commutative process and computationally cheap. It is therefore a straightforward process to create a global motion from a set of cNMA mode shapes and frequencies, as defined by Eq. (37).

Transition pathway discussion

The ENM is redefined for each intermediate structure enabling the cNMA mode set to reflect the current accessible motion space. The ENM implicitly assumes that each intermediate conformation is at equilibrium. This condition cannot be strictly observed over the course of the transition. However, the core transition motion that we aim to capture involves large coordinated structure motions, which are primarily dependent on the overall shape and density of the ENM and are substantially less sensitive to local rearrangements.^{18,66} It is for this reason that we can continue to apply the ENM as a traveling energy basin across the transition pathway. As an alternative, there are switching methods for including multiple energy basins in various network based models,^{48,49} as well as MD methods.⁶⁷

The cumulative effect of cNMA recalculation provides icNMA with the flexibility to produce a nonlinear pathway and to allow temporary localized unfolding. These two characteristics are critical in generating a functionally relevant transition pathway,⁴⁸ and are also both inherently excluded by linear interpolation methods.

As discussed in the Introduction, there are interpolation-based methods that use interatomic distances to drive one conformation to another. These approaches guarantee a transition pathway that reaches the goal conformation and maintains atomic contact distances within the range defined by the starting and goal conformations. However, these forced motions do not take into account the local energy landscape. In contrast, the icNMA method is not artificially constrained and is able to produce conformations with undesirable local geometries (i.e. bonded atomic pairs that stretch too far apart or nonbonded atoms that approach too closely).

The icNMA method is designed to capture the core motion of the transition pathway, but if tighter geometric control over the intermediate conformations is preferred, any of the following modifications can be applied if a global motion causes an undesirable conformation: (i) Perform an energy minimization to relax the intermediate conformation; or (ii) Revert the pathway back to the previous conformation and stiffen the harmonic potential between each pair of atoms that cause an undesirable contact distance; or (iii) Revert the pathway back to the previous conformation and merge the clusters containing each pair of atoms that cause an undesirable contact distance. In Section “Results and Discussion”, a comparison is made between an unconstrained icNMA pathway and one using the merged cluster modification.

Conformation comparison by RMSD and bRMSD

Consider a pair of conformations that are based on the same crystal structure. By the definition of clustering, the relative atomic positions within each cluster remain fixed, so the RMSD between conformations is entirely due to differences in cluster locations. In this situation it is possible to achieve an RMSD value of zero. Now consider a pair of conformations for the same protein, but derived from different crystal structures (e.g. the open and closed conformation states). The RMSD between conformations has a contribution from differences in cluster positions *and* a contribution from differences in relative atomic positions within each cluster. The former varies as the clusters move and can be used to evaluate the global progress of a conformational transition, whereas the latter remains constant and is an indication of local structure geometries that have been locked into place as a result of clustering. To make use of these global and local quantities, it is necessary to decouple the pair of contributions. This is accomplished by

introducing a new, cluster-based, metric for conformation comparison.

Given a pair of conformations and a clustering scheme, the RMSD is calculated independently for each cluster: isolate cluster c from each conformation, center and optimally align the pair, compute the RMSD between the clusters, and define the quantity as $RMSD_c$. Once this procedure is performed for each cluster, the $RMSD_c$ quantities are used to define the new metric, referred to as background RMSD (bRMSD)

$$\text{bRMSD} = \sqrt{\frac{1}{n} \sum_{c=1}^N [A(c) \cdot RMSD_c^2]} \quad (40)$$

where $A(c)$ is the number of atoms in cluster c . This quantity represents the lowest possible RMSD that can be achieved by optimal cluster positioning.

Crystal structure resolution is not equivalent to resolution of atomic position. The PDB states as a guideline that resolution of atomic position is one-tenth to one-fifth of the crystal resolution for structures with an R -value (a quantification of the model's agreement with the crystallographic data) less than 0.2. Accordingly, "good" crystal structures with resolutions of ~ 2.0 Å, give atomic positions with accuracies of ~ 0.2 – 0.4 Å. The observed bRMSD quantity for clustering by residue is ~ 0.3 Å, which is well within this range. The DOFs locked into place when clustering by residue impose a restriction on the atoms in the structure, which corresponds to an RMSD value less than the atomic resolution of the model. Therefore, any (more aggressive) candidate clustering may be validated by performing a bRMSD calculation prior to the icNMA computations.

The progress of an icNMA transition pathway is monitored by a combination of the RMSD and bRMSD metrics. The quantities discussed below are depicted in Figure 1. Consider an icNMA intermediate conformation, C_i , at an RMSD $d = RMSD(C_i, C_F)$ from the goal conformation. The space of all possible global motion steps around C_i defines a hyper-sphere, S_i , of radius μ . A second hyper-sphere, S_F , is defined around C_F with a radius d . The surface of S_i that falls within S_F represents global motions that lead to potential C_{i+1} conformations which reduce the RMSD to C_F ; the surface of S_i outside of S_F represents motions that increase RMSD. The intersection of S_i and S_F is defined by an angle relative to the reference direction, $\hat{\delta}$, as

$$\theta = \text{acos}\left(\frac{\mu}{2d}\right) \quad (41)$$

This solution is obtained with the law of cosines and is based on the geometry of the configuration shown in Figure 1.

As the icNMA pathway gets closer to the goal conformation, θ becomes smaller, indicating a restriction on

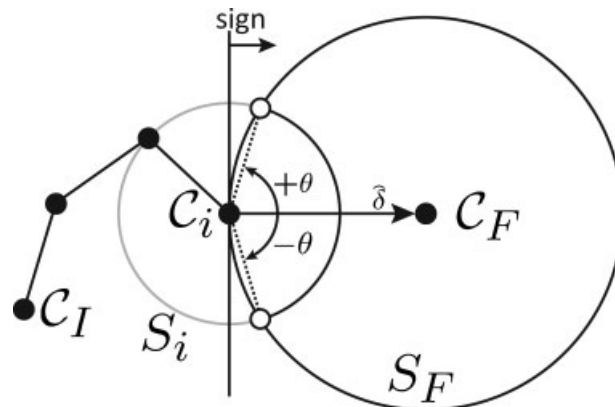


Figure 1

An arbitrary icNMA pathway is depicted as a series of conformations (filled circles) connected by global motions (solid lines). The current intermediate conformation (C_i) is surrounded by the hypersphere S_i . The sign choice for the direction of the global motion construction restricts the space of all possible global motions to the S_i hemisphere to the right of the vertical line labeled "sign". The reference direction ($\hat{\delta}$) is the labeled arrow connecting C_i to the goal conformation (C_F). The hypersphere S_F surrounds C_F and contains all conformations that are closer to C_F than the current intermediate. The portion of S_i within S_F is shown as a solid arc and the portion outside of S_F is grayed. The θ angle [Eq. (41)] quantifies the portion of the global motion space that reduces the RMSD of the pathway.

the available space of RMSD reducing global motions. In the limit of $\theta(d \rightarrow \infty) = \pi/2$ the entire hemisphere of S_i is RMSD minimizing. The θ function falls off very quickly as d decreases: $\theta(d = \mu) = \pi/3$ and $\theta(d = \mu/2) = 0$. It is not productive to continue an icNMA simulation that diverges, so a value of $d = \mu$ is used as the termination criteria. The progress of the simulation as it approaches the termination criteria is quantified by the function

$$\pi(i) = \frac{RMSD(C_I, C_F) - RMSD(C_i, C_F)}{RMSD(C_I, C_F) - (\text{bRMSD}(C_I, C_F) + \mu)} \quad (42)$$

where the input parameters for the functions $RMSD$ and $bRMSD$ indicate the conformation pair to which they are applied. The numerator quantifies how much RMSD has been traversed. The denominator quantifies the total RMSD the pathway is expected to traverse; excluding the RMSD which cannot be reduced due to clustering (i.e. bRMSD) and excluding the RMSD associated with the step size (i.e. μ). It is possible for $\pi(i)$ to reach a value slightly greater than 1, if the trajectory gets within μ RMSD of the goal conformation, but in practice, the function will approach a value of 1.

RESULTS AND DISCUSSION

Example structure

The β 1,4-galactosyltransferase-T1 structure (open: PDB = 1FR8⁶⁸ and closed: PDB = 1NKH,⁶⁹ commonly written

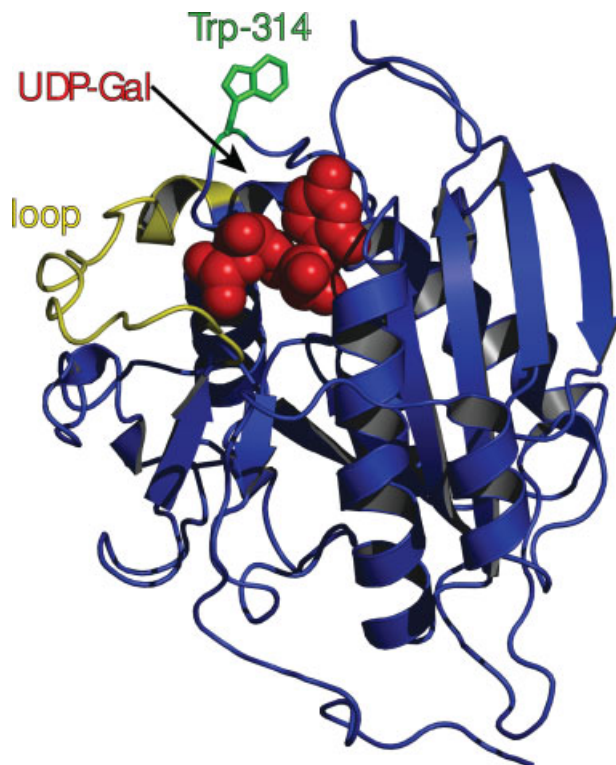


Figure 2

Cartoon representation of the open conformation of $\beta 4\text{Gal-T1}$. The loop residues are shown in yellow, the UDP-Gal ligand is space filled shown in red, and Trp³¹⁴ is shown in green sticks. (All structure representations are created with PyMOL.⁷⁰)

with the abbreviation $\beta 4\text{Gal-T1}$, is composed of 2209 atoms in 271 residues and is the catalytic component of the lactose synthase enzyme (Fig. 2). During biological function, $\beta 4\text{Gal-T1}$ binds uridine diphosphogalactose (UDP-Gal), thus causing a large conformational change in the loop region comprising residues 345-365. The mean RMSD value between conformations for the alpha-carbons on the loop is 9.8 Å, and Lys³⁵² achieves a displacement of 20 Å. In comparison, the rest of the structure experiences a mere 0.6 Å displacement.⁶⁹ Trp³¹⁴ is opposite the large loop and swings down over the pocket assisting in the ligand binding.⁷¹

$\beta 4\text{Gal-T1}$ belongs to a superfamily of enzymes called glycosyltransferases that are involved in the synthesis of sugar moieties of glycoproteins and glycolipids. Crystal structures of many of these enzymes are available and, similar to $\beta 4\text{Gal-T1}$, they exhibit conformational changes involving at least one flexible loop.⁷² Analyses of the other glycosyltransferases is possible and is expected to provide insight into the conformational dynamics of these enzymes.

Simulations

The icNMA simulations are performed on the open conformation without the ligand; clustering is by residue and the global motion step size is $\mu = 0.1$ Å, producing an expected RMSD of: $(\text{bRMSD} + \mu) = 0.60$ Å. The control simulation is run with $p = 100\%$ [Eq. (38)] and no contact distance constraints. A second simulation is run with $p = 50\%$ and no contact constraints, to probe the relevance of the higher frequency modes on the pathway construction. A third simulation is run with $p = 100\%$ and a minimum allowable contact distance of 1.1 Å and a maximum allowable peptide bond length of 2.5 Å, to test the effects of constrained geometries. The three simulations, now referred to as \mathcal{S}_{100} , \mathcal{S}_{50} , and $\hat{\mathcal{S}}_{100}$, respectively, are listed in Table I, along with their DOF usage and π values. The following sections discuss pathway RMSD, mode decompositions, local geometry, pathway features, and energy.

Pathway RMSD

The RMSD evolutions (Fig. 3) show that all simulations experience an initial steep descent in RMSD, followed by asymptotic approaches to their final values. The early success of \mathcal{S}_{50} is explained by the fact that the lowest modes are contributing the majority of the desired transition and the extra modes utilized by \mathcal{S}_{100} , which come from the higher frequency range, are extraneous. The early success of $\hat{\mathcal{S}}_{100}$ indicates that, initially, the pathway naturally stays within the range of the constraints, thus performing like the control simulation.

Mode decompositions

The mode space dimension of \mathcal{S}_{50} is nearly three times smaller than that of $\hat{\mathcal{S}}_{100}$, but yet it achieves a significantly better transition result ($\pi = 86.5\%$ vs. $\pi = 59.1\%$). This comparison indicates that the dimensionality of the motion subspace is less important than which region of the full motion space it represents. This concept is quantified by computing mode space decompositions.

Table I
 $\beta 4\text{Gal-T1}$ icNMA Simulations

	p (%)	Modes (#)	DOF (%)	min (Å)	max (Å)	π (%)
\mathcal{S}_{100}	100	1620	24.5	—	—	96.4
\mathcal{S}_{50}	50	~440	6.6	—	—	86.5
$\hat{\mathcal{S}}_{100}$	100	~1275	19.3	1.1	2.5	59.1

Simulation parameters and π values for three icNMA computations of 80 steps each. The number of modes at each step varies during the simulations of the second and third rows and their listed values are approximations to their averages over the course of the simulations. The DOF % is the number of icNMA modes as a percent of the number of nonrigid DOFs in an all atom Cartesian representation (i.e. $3 \times 2209 - 6 = 6621$). The π values [Eq. (42)] indicate how much of the RMSD between the starting and goal conformations has been traversed.

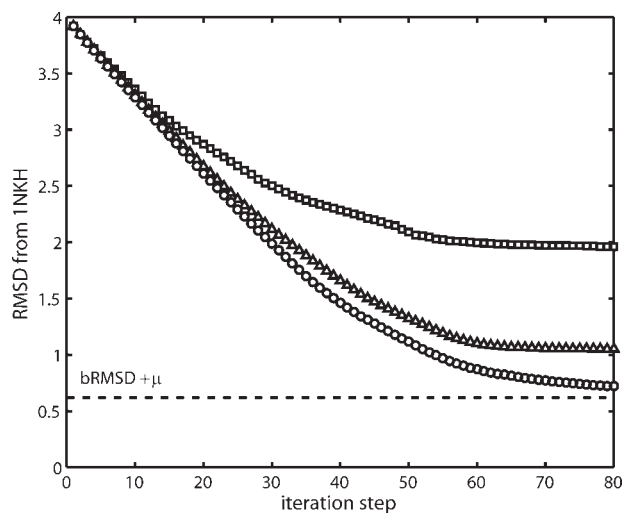


Figure 3

Comparison of RMSD changes for the icNMA pathway simulations defined in Table I. The following symbols are used: $\circ = \mathcal{S}_{100}$, $\triangle = \mathcal{S}_{50}$, $\square = \hat{\mathcal{S}}_{100}$.

A pair of icNMA simulations for the same initial structure, but based on different parameters, will produce pathways that explore different regions of conformation space. At any given intermediate step, the accessible motion space of one simulation is compared to that of the other by using a decomposition. Let the normalized modes of one simulation be given by $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ and the normalized modes of a second simulation be given by $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_q\}$, then the decomposition of the first simulation's motion subspace over that of the second is given by the $p \times q$ matrix whose elements are defined by

$$D_{i,j} = |\mathbf{a}_i \cdot \mathbf{b}_j| \quad (43)$$

The decomposition matrix is populated with values on the interval $[0,1]$ where 0 indicates linear independence and 1 indicates exact alignment of the corresponding mode pair.

The decomposition data is evaluated in two ways. First, a concentration of high values along the diagonal indicates that modes from one set are highly aligned with modes with the same index of the other set (a solid line is superimposed on the decomposition plots for reference). This feature gives a general indication of how the motion subspace of \mathcal{A} compares to that of \mathcal{B} . Second, the vector norm of column j , is a measure of how well mode \mathbf{b}_j is represented by all of \mathcal{A} . Computing this norm for all columns in D gives a precise measure of how well particular dimensions of \mathcal{B} are being captured by the entire space of \mathcal{A} .

The discrepancy in π values is now addressed by mode decompositions with respect to the control simulation. The decompositions are based on the intermediate con-

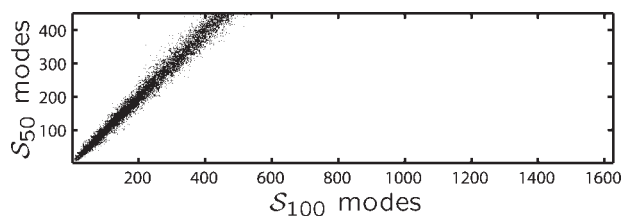


Figure 4

The decomposition of the mode shapes from \mathcal{S}_{50} when compared with the mode shapes of \mathcal{S}_{100} , at intermediate conformation 40. The data occupies the diagonal, thus indicating that \mathcal{S}_{50} has, as expected, accurately and completely reproduced the low frequency motion space of \mathcal{S}_{100} .

formations at step 40, which is halfway through each simulation and also approximately where the RMSD plots start to diverge the fastest.

The decomposition of the \mathcal{S}_{50} modes (Fig. 4) covers the reference line (i.e. the decomposition spans the same number of modes on each axis). This indicates that each of the 450 modes of \mathcal{S}_{50} is highly aligned with the \mathcal{S}_{100} mode of the same index. Therefore, the low-frequency motion space has been completely captured by \mathcal{S}_{50} . This is evident by the fact that \mathcal{S}_{50} achieves almost the same π value as the control. In contrast, the $\hat{\mathcal{S}}_{100}$ decomposition (Fig. 5) deviates from the reference diagonal. As a result, the 1266 modes from $\hat{\mathcal{S}}_{100}$ do not correspond to an equal number of modes in \mathcal{S}_{100} , but rather are evenly distributed across all 1626 \mathcal{S}_{100} modes. For the lower dimension

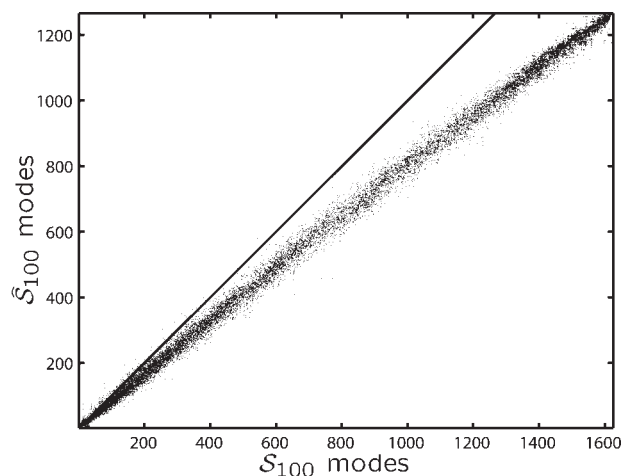


Figure 5

The decomposition of the mode shapes from $\hat{\mathcal{S}}_{100}$ when compared with the mode shapes of \mathcal{S}_{100} , at intermediate conformation 40. The data is shifted to the right of the reference line, thus indicating that modes of $\hat{\mathcal{S}}_{100}$ are uniformly shifted into the higher frequency spectrum. Consequently, parts of the lower frequency mode space are not as densely populated and are not as well captured.

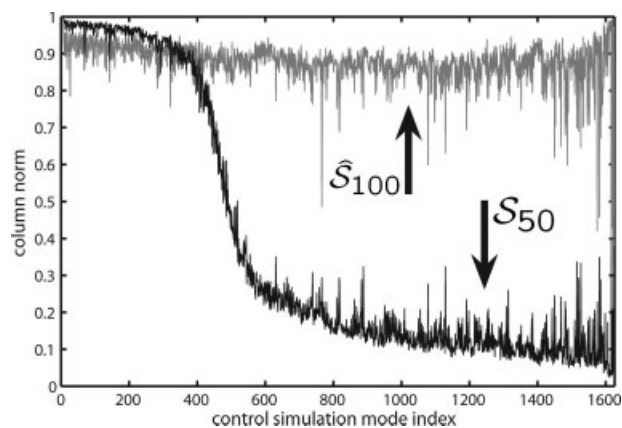


Figure 6

The column norms of the S_{50} and \hat{S}_{100} mode decompositions.

subspace of \hat{S}_{100} to be evenly distributed across the higher dimension subspace of S_{100} , there must necessarily be motions of S_{100} which are not well represented by \hat{S}_{100} .

The plot of column vector norms (Fig. 6) confirms this by clearly showing that even though the \hat{S}_{100} modes capture the entire frequency spectrum fairly well, the S_{50} modes do a noticeably better job of capturing the low frequency modes, which are necessary for constructing the transition pathway. Further, because these modes are more highly weighted in their relative contribution to the global motion than the higher frequency modes, the differential seen in Figure 6 is much more significant.

The preceding decomposition analysis explains why the lower DOF S_{50} actually produces a better π value than the higher DOF \hat{S}_{100} . Intermediate conformation geometries and pathway energies are analyzed in the following sections to ascertain the necessity and effects of constrained simulations.

Local geometry

Figure 7 shows the evolution of the distances between the five pairs of atoms that reach the shortest contact distances during S_{100} . Figure 8 shows the evolution of the lengths of the five peptide bonds that stretch the most during S_{100} . A pair of atoms reach a contact distance of 0.25 \AA and a peptide bond is stretched to a length of 6.5 \AA – both of these situations are not possible. However, these are the extremes of the simulation and are only temporary. In fact, the bulk of the structure is acceptable. In the open conformation of $\beta 4\text{Gal-T1}$, there are 26,340 atom pairs within a cutoff distance of 5 \AA , and of these, 18,840 are between atoms of different residues and are free to move relative to each other. Dur-

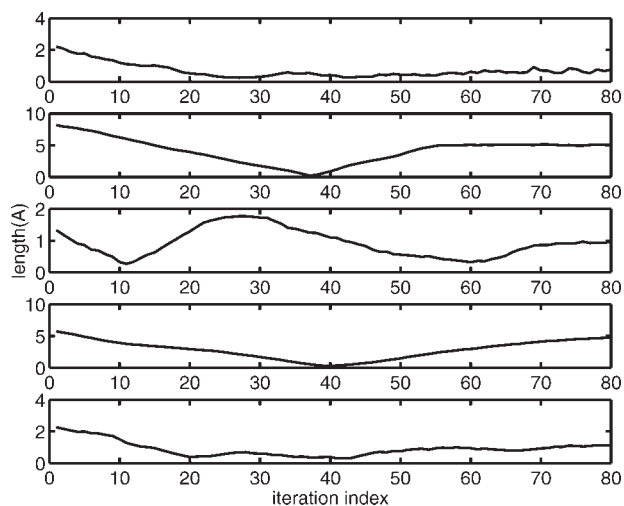


Figure 7

Evolution of the separation distance between atomic pairs that reach one of the five shortest separation distances during the S_{100} simulation. The atom pairs in each plot, from the top down, are: (Ala²²¹:O–Gly²²²:N), (Asp³⁵⁰:N–Lys³⁵¹:CE), (Asn³⁵³:C–Glu³⁵⁴:N), (Arg³⁵⁹:C–Phe³⁶⁰:CE2), (Asn³⁵³:O–Glu³⁵⁴:N). Only 0.3% of all contacts between clusters ever get within 1.1 \AA (the minimum distance constraint enforced by \hat{S}_{100}).

ing S_{100} , 52 of these atomic pairs (0.3%) come within 1.1 \AA of each other and 17 of the peptide bonds (6.3%) stretch beyond 2.5 \AA . The S_{50} results are similar (0.1% and 5.2%, respectively) and are not plotted.

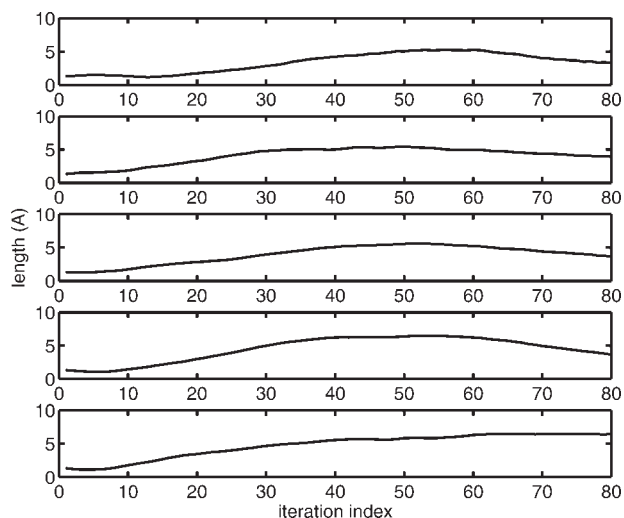


Figure 8

Evolution of the peptide bond lengths for those bonds that become one of the five longest bonds during the S_{100} simulation. The atom pairs in each plot, from the top down, are: (GLY³¹³:C–TRP³¹⁴:N), (PHE³⁶⁰:C–ASP³⁶¹:N), (ARG³⁵⁹:C–PHE³⁶⁰:N), (GLN³⁵⁸:C–ARG³⁵⁹:N), (ASP³⁵⁰:C–LYS³⁵¹:N). Only 6.3% of all peptide bonds ever stretch beyond 2.5 \AA (the maximum distance constraint enforced by \hat{S}_{100}).

Comparing the three icNMA simulations has shown the effects of reducing the dimension of the global motion subspace and of imposing geometric constraints. In particular, the unconstrained simulation, \mathcal{S}_{100} , produces a near complete ($\pi = 96.4\%$) transition pathway while allowing minimal local geometry violations. The constrained simulation, $\hat{\mathcal{S}}_{100}$, removes the undesirable contacts, but at the expense of causing a mode set frequency shift, which limits the progression of the transition pathway ($\pi = 59.1\%$).

Pathway features

There are several features of the $\beta 4\text{Gal-T1}$ transition pathway that make it an excellent example structure. There are many possibilities for icNMA modification, but as demonstrated in the previous sections, \mathcal{S}_{100} successfully produces the core motions of a transition pathway from the open to the closed conformation. The discussion which follows is based on the motions of this simulation and highlight the capabilities of the icNMA method. Several AVI movie clips showing features of the pathway discussed later are available as Supplemental Material on the journal's website.

The main loop and Trp^{314} are involved in a cooperative motion in which they move towards each other and overlap as they come down onto the ligand, holding it in the binding pocket (Fig. 9). A direct interpolation of this motion would cause severe steric clash, but icNMA produces a pathway in which Trp^{314} undergoes a rotation about the backbone so that it can pass underneath the main loop and reach the closed conformation. The very fact that the cNMA parameterization is designed to represent rotational motions allows icNMA to describe this complex transitional motion. Purely Cartesian-based models are not able to characterize this motion without a full atomic model, but the computational expense is prohibitive.

The biological function of $\beta 4\text{Gal-T1}$ is accomplished by the large motions of relatively few residues (27 main loop residues and 6 residues around Trp^{314}), whereas the bulk of the structure remains stationary. This type of motion is common to globular proteins and the analysis greatly benefits from modeling techniques that allow for multiscale representations. The scope of this article does not allow for the detailed inclusion of such a simulation, but the authors performed an icNMA simulation in which the mobile residues mentioned earlier are individually clustered and the remaining structure is defined as a single cluster. This model uses 198 nonrigid DOFs, which is 3% of the all-atom, nonrigid DOFs, and has $\text{bRMSD} = 0.75 \text{ \AA}$. The computation time is cut by an order of magnitude with respect to \mathcal{S}_{100} , but yet the conformational transition pathway still achieves $\pi = 93.5\%$ and never gets more than 0.9 \AA RMSD away from \mathcal{S}_{100} (which is closer than either \mathcal{S}_{50} or $\hat{\mathcal{S}}_{100}$).

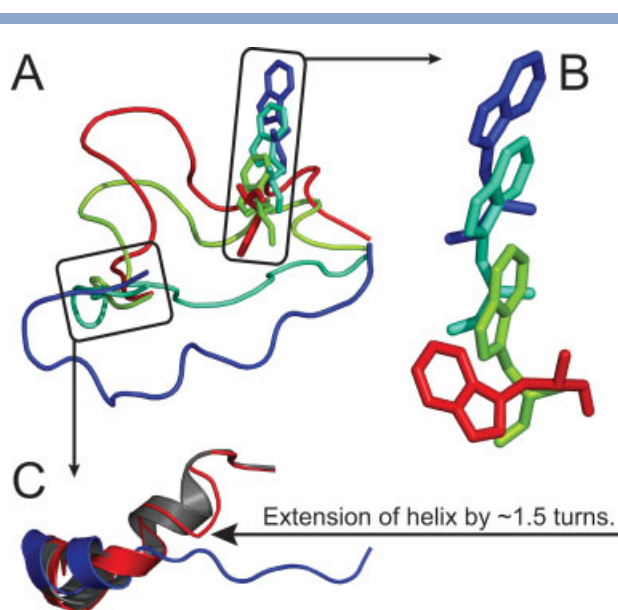


Figure 9

The superimposed intermediate conformations show the icNMA transition (blue \rightarrow red) of $\beta 4\text{Gal-T1}$. The viewpoint in (A) is from the binding pocket looking outward and shows Trp^{314} pass inside of the main loop as they approach each other. The isolated view in (B) shows Trp^{314} rotating about its backbone connection (the side chain starts directly above the C_α and finishes horizontal from it), thus allowing it to fit under the main loop and close onto the binding pocket, which is below and to the right. The isolated view in (C) shows slack in the main loop of the open conformation (blue) get incorporated into a neighboring helix. The final icNMA conformation (red) overlays the crystal structure of the closed conformation (gray).

The final interesting characteristic of the $\beta 4\text{Gal-T1}$ transition pathway included in this discussion is the extension of an alpha-helix. As the main loop undergoes its large motion, it creates slack (perhaps this is what allows the clearance for Trp^{314} to pass underneath it). This slack is converted into an additional turn (residues 358–363) on a neighboring alpha-helix as the path approaches the goal conformation (Fig. 9).

Pathway energies

An effective way to monitor pathway energy is by utilizing the ENM framework upon which icNMA is already based. The distances between all pairs of atoms in each end state are computed and serve as reference values (i.e. equilibrium states). During a transition, displacements from these reference values result in deformation of the corresponding springs in the ENM. The energy function for atom a in intermediate conformation C_i is defined as

$$V_a(i) = \frac{1}{2} \sum_{b=1}^n \min(\delta d_{a,b}(i, \mathcal{I}), \delta d_{a,b}(i, \mathcal{F}))^2 \quad (44)$$

where $\delta d_{a,b}(i, \mathcal{I})$ is the magnitude of the change in distance between atoms a and b as compared between

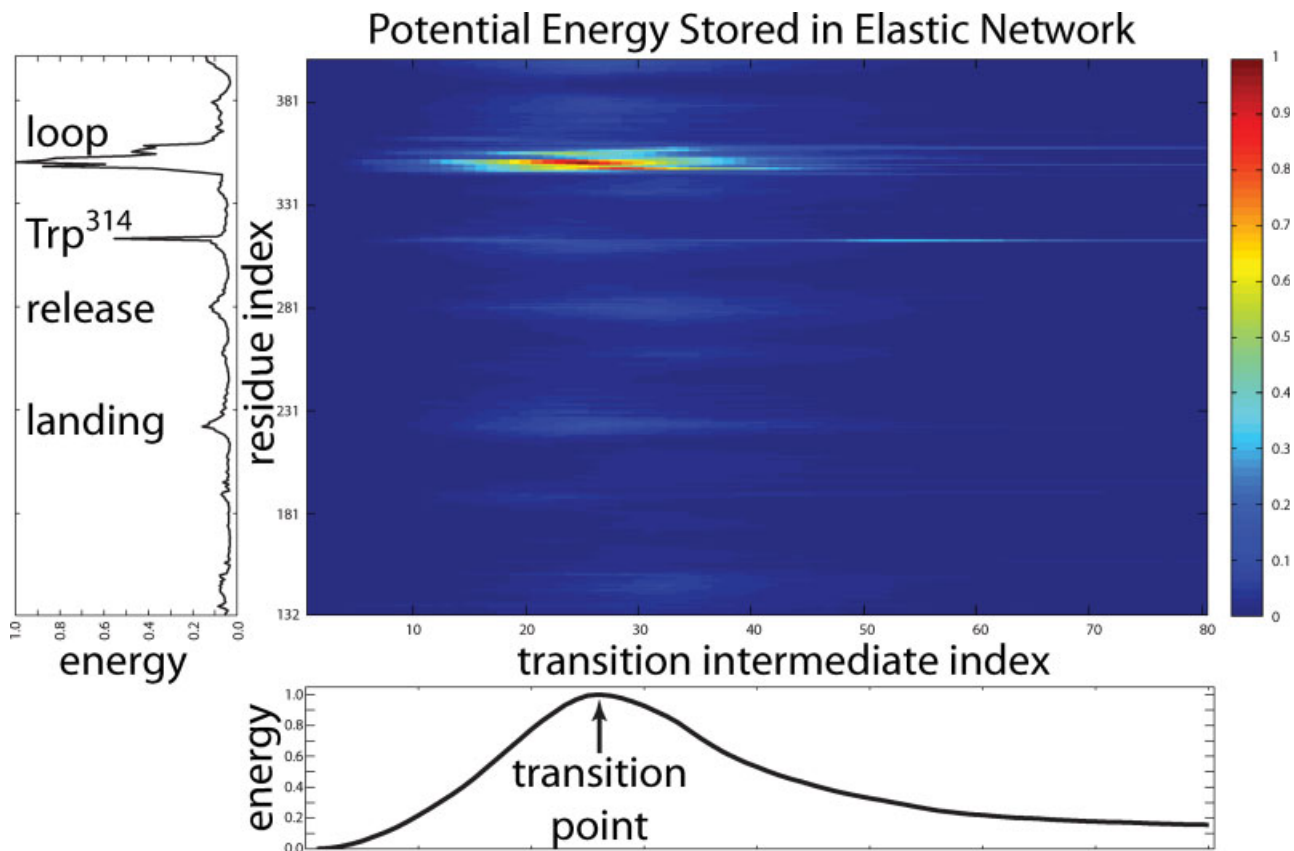


Figure 10

The potential energy contained in the springs radiating from each C_i as a function of pathway intermediate index (colormap in center). Summing across the rows (i.e. over the pathway) produces the vertical plot at the left, which shows the total normalized energies produced by each atom's motions. Summing down the columns (i.e. over the set of atoms) produces the horizontal plot at the bottom, which shows the evolution of the structure's total normalized energy over the transition.

conformations C_i and C_f . The “min” function acts as a switch allowing each atomic pair to be evaluated relative to the end state to which it is closest. The potential function is shown in Figure 10 for the \mathcal{S}_{100} simulation.

The potential function varies with atom index (i.e. spatial location) and conformation index (i.e. “temporal” location). Fixing one of these variables and summing over the other produces an energy distribution over either space or “time”. These plots are shown alongside the potential function in Figure 10.

Summing over conformation indices produces an energy distribution over atomic indices (vertical plot in Fig. 10). This plot clearly identifies the structural components involved in the transition. The main loop and Trp³¹⁴ are the largest contributors to system energy, and the next two locations are where the main loop anchors to the structure prior to the transition (labeled “release”) and after the transition (labeled “landing”).

Summing over atomic indices produces an energy distribution over conformation indices (horizontal plot in Fig. 10). This distribution reveals the energy barrier in

the transition. Rather than summing over all atomic indices, summations can be taken independently on each of the four locations identified earlier. The peaks in each of these energy distributions (not shown, but readable from the potential function) indicates when in the transition the corresponding structure element is active.

Combining all of these findings results in the following pathway sequence: (1) Trp³¹⁴ completes the first phase of its motion, (2) the main loop reaches an intermediate position and the landing site rearranges, both coinciding with the transition barrier, (3) immediately after the barrier is crossed, the release site rearranges, and (4) Trp³¹⁴ completes its second phase of motion. The two stage motion of Trp³¹⁴ is evident in Figure 9, by the sliding motion of the side chain, followed by its rotation around the backbone.

Adenylate kinase comparison

β 4Gal-T1 was chosen as the primary example structure for this article because its crystal structures suggest a

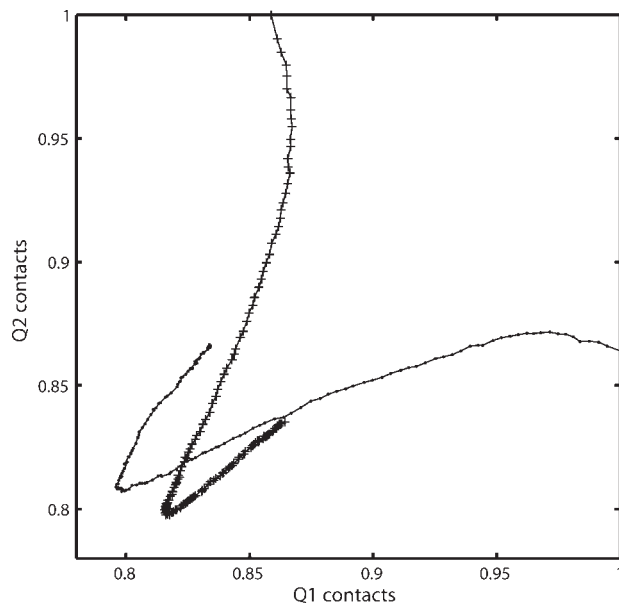


Figure 11

Q1 versus Q2 plots for icNMA pathways generated for adenylate kinase. The transition from open to closed (●) and closed to open (+) are both nonlinear and reach the barrier region in the lower left portion of the plot. Both simulations use the S_{100} parameters.

wide range of motions that must be spatially and temporally coordinated in such a way that interpolation (i.e. linear methods) cannot capture. Adenylate kinase (open: PDB=4AKE⁷³ and closed: PDB=1AKE⁷⁴) is a common structure used to test computational methods due to the wealth of experimental data available. The icNMA pathway for the adenylate kinase transition is briefly studied and comparisons are made with existing methods. The following material is not intended as a complete analysis or as validation of icNMA, but rather as an indication that the most fundamental properties of such a transition are captured by icNMA.

Franklin *et al.*⁵¹ create a transition pathway by stitching together one linear system for the initial conformation with a second linear system for the final conformation. The nonlinear pathway is well represented by a Q1 versus Q2 plot, in which the authors track the percentage of initial (Q1) and final (Q2) conformation contacts that exist in each intermediate conformation. This plot shows that Q1 contacts are broken and *then* Q2 contacts are formed in the very last stages of the transition. This nonlinear pathway enters the lower left region of the plot where both Q1 and Q2 are minimized. The barrier conformations that populate this region are metastable as they are equally distant from the stable open and closed conformations. In contrast, methods which produce linear pathways, like the interpolation-based UMMS^{75,76}

method, force the simultaneous breaking of Q1 contacts and formation of Q2 contacts.

To place the icNMA method in context with other techniques, transition pathways of adenylate kinase are mapped onto a Q1 versus Q2 plot (Fig. 11). The transition pathway from the open to the closed conformation is complete in RMSD ($\pi = 95.3\%$), but it only forms 30.8% of the missing atomic contacts required in the final conformation (the Q2 value increases from 0.81 to 0.87). The application of icNMA in the reverse direction (i.e. from the closed conformation to the open conformation) produces a transition pathway that once again achieves RMSD convergence ($\pi = 94.9\%$) while only forming 26.4% of the missing atomic contacts required in the open conformation (the Q1 value increases from 0.82 to 0.86).

Both of the transition pathways enter the barrier region and then proceed towards their respective goal conformations. This demonstrates that the icNMA method is able to produce a nonlinear pathway and access the population of metastable states in the barrier region. The s_i parameters of the icNMA global motion are computed with respect to minimizing RMSD from the goal conformation, whereas the Q1-Q2 coordinate space is directly tied to contact geometries. If desired, an alternative contact-based function for setting the s_i values could guide the icNMA pathway in a way more consistent with Q1-Q2-space representation.

CONCLUSION

The icNMA method is founded on a traveling harmonic potential which defines the evolving global motion and guides the transition pathway. Unlike interpolation and extrapolation based methods, the iterative updating of the local accessible motion space reflects the changing geometry of the structure. The global motion allows the structure to evolve towards its destination conformation along a nonlinear path while traveling only through the reduced dimensional subspace defined by the low-frequency modes of cNMA.

The analysis of β 4Gal-T1 illustrates the features of icNMA and the resulting pathway captures a variety of different structure motions which all give insight into how the biological function is achieved. The energy analysis of the pathway reveals spatially and temporally coordinated motions. The main structural elements identified in this analysis are ideal candidates for mutagenesis studies.

The Introduction discusses NMA-based structure deformation methods which utilize distance constraints and X-ray diffraction pattern matching. These methods are powerful ways to convert partial structural data from experiments into high-resolution conformations, especially in the metastable transition region where complete

structural information is difficult to produce. The icNMA method can perform this task by replacing its mode shape sign choice with any scoring function that promotes a given set of structural features. In addition to producing transition states, the icNMA method also produces corresponding transition pathways.

The power and flexibility of icNMA is due to its effective combination of independent local (i.e. cNMA mode space) and global (i.e. RMSD minimization) constraints.

REFERENCES

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235–242; <http://www.pdb.org/>.
- Ha T. Single-molecule fluorescence resonance energy transfer. *METHODS* 2001;25:78–86.
- Palmer AG, III. Probing molecular motion by NMR. *Curr Opin Struct Biol* 1997;7:732–737.
- Doniach S, Eastman P. Protein dynamics simulations from nanoseconds to microseconds. *Curr Opin Struct Biol* 1999;9:157–163.
- Englander SW, Krishna MM. Hydrogen exchange. *Nat Struct Biol* 2001;8:741–742.
- Schuyler AD, Chirikjian GS. Normal mode analysis of proteins: a comparison of rigid cluster modes with C_{α} coarse graining. *J Mol Graph Model* 2004;22:183–193.
- Schuyler AD, Chirikjian GS. Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA. *J Mol Graph Model* 2005;24:46–58.
- Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* 1976;104:59–107.
- Brooks BR, Bruccoleri RE, 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.
- Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, III, DeBolt S, Ferguson D, Seibel G, Kollman P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* 1995;91:1–41.
- Duan Y, Kollman P. Computational protein folding: from lattice to all-atom. *IBM Syst J* 2001;40:297–309.
- Noguti T, Go N. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature* 1982;296:776–778.
- Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci USA* 1983;80:3696–3700.
- Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 1983;80:6571–6575.
- Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 1985;181:423–447.
- Tirion MM, ben Avraham D. Normal mode analysis of G-actin. *J Mol Biol* 1993;230:186–195.
- Bahar I, Rader A. Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* 2005;15:586–592.
- Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc Natl Acad Sci USA* 2006;103:7664–7669.
- Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996;77:1905–1908.
- Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Design* 1997;2:173–181.
- Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys Rev Lett* 1998;80:2733–2736.
- Atilgan A, Durell S, Jernigan R, Demirel M, Keskin O, Bahar I. anisotropy of fluctuation dynamics of proteins with an elastic network model. *Bio-phys J* 2001;80:505–515.
- Tozzini V. Coarse-grained models for proteins. *Curr Opin Struct Biol* 2005;15:144–150.
- Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. *J Chem Theory Comput* 2006;2:696–704.
- Tama F, Gadea FX, Marques O, Sanejouand Y-H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function, and Genetics* 2000;41:1–7.
- Li G, Cui Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+} -ATPase. *Biophys J* 2002;83:2457–2474.
- Chun HM, Padilla CE, Chin DN, Watanabe M, Karlov VI, Alper HE, Soosar K, Blair KB, Becker OM, Caves LS, Nagle R, Haney DN, Farmer BL. MBO(N)D: a multibody method for long-time molecular dynamics simulations. *J Comput Chem* 2000;21:159–184.
- Zhou L, Siegelbaum SA. Effects of surface water on protein dynamics studied by a novel coarse-grained normal mode approach. *Biophys J* 2008;94:3461–3474.
- Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* 1994;12:84–89.
- Zheng W, Brooks BR. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys J* 2006;90:4327–4336.
- Petrone P, Pande VS. Can conformational change be described by only a few normal modes? *Biophys J* 2006;90:1583–1593.
- Kirilova S, Cortés J, Stefani A, Siméon T. An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins. *Proteins* 2008;70:131–143.
- Krebs W, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins: Structure, Function, and Genetics* 2002;48:682–695.
- Yang L, Song G, Jernigan RL. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys J* 2007;93:920–929.
- Song G, Jernigan RL. An enhanced elastic network model to represent the motions of domain-swapped proteins. *Proteins: Structure, Function, and Genetics* 2006;63:197–209.
- Su JG, Jiao X, Sun TG, Li CH, Chen WZ, Wang CX. Analysis of domain movements in glutamine-binding protein with simple models. *Biophys J* 2007;92:1326–1335.
- Saibil HR. Conformational changes studied by cryo-electron microscopy. *Nat Struct Biol* 2000;7:711–714.
- Dror O, Lasker K, Nussinov R, Wolfson H. EMatch: an efficient method for aligning atomic resolution subunits into intermediate-resolution cryo-EM maps of large macromolecular assemblies. *Acta Crystallogr D* 2007;63:42–49.
- Tama F, Wrighers W, Brooks CL, III. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol* 2002;321:297–305.
- Tama F, Miyashita O, Brooks CL, III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* 2004;337:985–999.

41. Tama F, Miyashita O, Brooks CL, III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* 2004;147:315–326.
42. Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapère J-J. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys J* 2005;88:818–827.
43. Kim MK, Chirikjian GS, Jernigan RL. Elastic models of conformational transitions in macromolecules. *J Mol Graph Modell* 2002;21:151–160.
44. Kim MK, Jernigan RL, Chirikjian GS. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys J* 2005;89:43–55.
45. Kim MK, Li W, Shapiro BA, Chirikjian GS. A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA. *J Biomol Struct Dyn* 2003;21:395–405.
46. Zheng W, Brooks BR. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys J* 2005;88:3109–3117.
47. Jeong JI, Lattman EE, Chirikjian GS. A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure. *Acta Crystallogr D* 2006;62:398–409.
48. Miyashita O, Onuchic J, Wolynes P. Nonlinear elasticity, protein-quakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 2003;100:12570–12575.
49. Best RB, Chen Y-G, Hummer G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* 2005;13:1755–1763.
50. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol* 2005;352:807–822.
51. Franklin J, Koehl P, Doniach S, Delarue M. MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. *Nucleic Acids Res* 2007;35:W477–W482.
52. Chu J-W, Voth GA. Coarse-grained free energy functions for studying protein conformational changes: a double-well network model. *Biophys J* 2007;93:3860–3871.
53. Zheng W, Brooks BR, Hummer G. Protein conformational transitions explored by mixed elastic network models. *Proteins* 2007;69:43–57.
54. Zheng W, Brooks BR, Thirumalai D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys J* 2007;93:2289–2299.
55. Fischer S, Karplus M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem Phys Lett* 1992;194:252–261.
56. Yang L, Song G, Carriquiry A, Jernigan RL. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 2008;16:321–330.
57. Whitford PC, Miyashita O, Levy Y, Onuchic JN. Conformational transitions of adenylate kinase: switching by cracking. *J Mol Biol* 2007;366:1661–1671.
58. Arora K, Brooks CL. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci USA* 2007;104:18496–18501.
59. Thorpe M, Lei M, Rader A, Jacobs DJ, Kuhn LA. Protein flexibility and dynamics using constraint theory. *J Mol Graph Model* 2001;19:60–69.
60. Rader A, Hespenheide BM, Kuhn LA, Thorpe M. Protein unfolding: Rigidity lost. *Proc Natl Acad Sci USA* 2002;99:3540–3545.
61. Gohlke H, Thorpe M. A natural coarse graining for simulating large biomolecular motion. *Biophys J* 2006;91:2115–2120.
62. Kannan N, Vishveshwara S. Identification of side-chain clusters in protein structures by a graph spectral method. *J Mol Biol* 1999;292:441–464.
63. Brinda K, Vishveshwara S. A network representation of protein structures: implications for protein stability. *Biophys J* 2005;89:4159–4170.
64. Hinsen K. *Mathematical and Computational Biology Series, vol. 9: Normal mode analysis: theory and applications to biological and chemical systems.* Chapman & Hall/CRC: New York; 2006. chapter 1, pp 1–16.
65. Van Wynsberghe AW, Cui Q. Interpreting correlated motions using normal mode analysis. *Structure* 2006;14:1647–1653.
66. Tama F, Sanjouand Y. Conformational change of proteins arising from normal mode calculations. *Protein Eng* 2001;14:1–6.
67. Okazaki K-i, Koga N, Takada S, Onuchic JN, Wolynes PG. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci USA* 2006;103:11844–11849.
68. Gastinel LN, Cambillau C, Bourne Y. Crystal structures of the bovine β 4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J* 1999;18:3546–3557.
69. Ramakrishnan B, Qasba PK. Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β 1,4-galactosyltransferase-I. *J Mol Biol* 2001;310:205–218.
70. DeLano W. *The PyMOL Molecular Graphics System*, 2002.
71. Ramakrishnan B, Balaji P, Qasba PK. Crystal Structure of β 1,4-Galactosyltransferase Complex with UDP-Gal Reveals an Oligosaccharide Acceptor Binding Site. *J Mol Biol* 2002;318:491–502.
72. Qasba PK, Ramakrishnan B, Boeggeman E. Substrate-induced conformational changes in glycosyltransferases. *Trends in Biochemical Sciences* 2005;30:53–62.
73. Müller C, Schlauderer G, Reinstein J, Schulz G. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* 1996;4:147–156.
74. Müller C, Schulz G. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. A model for a catalytic transition state. *J Mol Biol* 1992;224:159–177.
75. Kim MK, Jernigan RL, Chirikjian GS. Efficient Generation of Feasible Pathways for Protein Conformational Transitions. *Biophys J* 2002;83:1620–1630.
76. Jang Y, Jeong JI, Kim MK. UMMS: constrained harmonic and anharmonic analyses of macromolecules based on elastic network models. *Nucleic Acids Res* 2006;34:W57–W62.