A Survey of Classification Methods for Protein Domain Motions and Flexibility

David Parker

Department of Mechanical Engineering
Stanford University, Stanford Ca.

Biochemistry 218 Final Project

March 10, 2003

Introduction

Protein domain motions play an essential role in a variety of biological phenomena, including signal transduction, transport of metabolites, cellular locomotion, ligand binding and catalysis [1 – 3]. The function of a particular structure is intimately linked to the nature of its motion. This is illustrated wonderfully by the induced fit model of substrate protein interaction [4] that addresses the importance of structural flexibility of proteins for their function. The interface between protein domains often constitutes a binding site in which open and closed conformations correspond to the active and inactive forms of the enzyme.

Elucidating the mechanisms of protein domain motions is not only important for understanding basic structure and function relationships but also has practical applications in drug and protein design. For protein design it may be important to estimate possible domain rearrangements, such as the flexibility properties of an intersubunit linker used in constructing an antibody [7]. The flexibility of a binding site is important in designing new drugs [8]. A good example of this is HIV protease (HIVP), a major inhibitory drug target for current acquired immunodeficiency syndrome (AIDS) therapy. HIVP has two flexible flaps that are known to be important for closing over and binding inhibitors. It is known that drug-resistant mutants of the protease cause a conformational perturbation in this region and it is thought that this causes resistance to drug binding [9].

In this paper we will first characterize the types of domain motions seen in proteins and then review several different methods for the classification of these motions. There is a large range of approaches to this problem, differentiated by computational complexity, execution speed, computer resources required, data sources, protein complexity (number of domains and size) and information content of the results. We have selected here a set of methods that best reflects these differences.

Characterization of Domain Motions

The mechanisms of protein conformational changes have been studied experimentally using X-ray crystallography for over 20 years. Structures of the same protein in different conformations (e.g., with a bound ligand) are available from the Protein Data Bank. The motions of proteins down to the nanosecond timescales can now be obtained using time-resolved X-ray crystallography [10]. Recently it has become possible to study large-scale protein motions using NMR [11]. From these data we can identify the mechanisms involved in protein domain motions.

A key concept in the study of protein structure is the domain. Structurally a domain is a compactly folded region of a protein that has independent stability. It is usually linked to other domains by very few structural elements such as a loop or helix. Dynamically, a

domain is a relatively rigid region connected to other domains by flexible interdomain regions. Most large proteins are built from assemblies of domains that for the most part consist of regions of nearly rigid motions jointed by flexible regions

The ability of different regions in a protein to move relative to each other with only a small expenditure of energy is defined as the proteins *intrinsic flexibility* [12]. The two types of motions associated with intrinsic flexibility are governed by the internal packing of the interfaces between two regions in a protein. The first type of motion is a *hinge mechanism* that occurs when there is no continuously maintained interface constraining the motion. Hinge motions usually occur in proteins with two domains with one domain rotating about the hinge as a rigid body. The rotation is caused by a few large torsion angle changes within the hinge region. The second type of motion is a *shear mechanism* that occurs when two interfaces slide across each other in order to maintain a well-packed interface. Shear motions are typically small so a large shear motion will be composed of a number of individual shear motions.

Difference-Distance Plots

Difference-distance plots are a common method used to identify rigid regions in proteins [13]. The differences between the distances for all pairs of residues for two protein conformations are displayed. A rigid region in the protein will have small differences in distances between the residues in that region. Although easy to implement and fast this method does not provide enough information to easily determine the spatial organization of a rigid region since it contains no information about the geometric proximity of the residues. There is also no information that can be used to compute a quantitative measure of the degree of rigidity of a region.

Normal Mode Methods

Normal Mode Analysis

Normal mode analysis was developed to model and visualize the collective motions of atoms in a molecule [15]. Biologically interesting motions can be explored without the much more computationally expensive and time-consuming molecular dynamics simulations.

The basic idea behind normal mode analysis is to model a molecule as a collection of harmonic oscillators (atoms) coupled together by springs (interatomic bonds). A molecule made up of N atoms will have 3N – 6 normal modes. The low-frequency (large displacement) modes of a protein are thought to be the most biologically significant and in fact may play a crucial role in binding pocket activity [16].

There are several limitations of normal mode analysis. One is that it can require a great deal of computer memory for even medium sized proteins. Another is that it assumes the motion of a protein is purely harmonic, in which case its motion can be exactly expressed as a superposition of normal modes. However, protein motions are known to be fluid-like and aharmonic so the energy potential near equilibrium must be approximated by a harmonic function. Still, normal mode analysis has provided a great deal of insight into the nature of collective motions in proteins [17].

Rigid Domain Identification using Low-frequency Normal Modes

In this method the protein is modeled as an elastic object using a measure that treats differences in conformation as a deformation. A large deformation indicates highly flexible interdomain regions. The sufficiently rigid parts of the protein are then classified into rigid domains and low-deformation interdomain regions as flexible on the basis of a normal mode analysis [24].

The normal modes are computed by an approximate method [25] for a single conformation of a protein. Only the C^{-} backbone atoms are used in order to reduce memory requirements and execution time. Once the modes have been computed the *deformation energy* E_i can be computed for every i_{th} atom as:

$$E_i = \sum_i k(\mathbf{R}_{ij}) \, |(\mathbf{d}_i - \mathbf{d}_j) \cdot \mathbf{R}_{ij}| \, / \, |\mathbf{R}_{ij}|^2 \qquad i,j = 1 \ldots \#\text{atoms}$$

$$k(\mathbf{r}) = c \exp(- |\mathbf{r}|^2 / r_0^2)$$

where $\mathbf{R}_{ij}$ is the distance vector from atom i to atom j in the reference conformation, $\mathbf{d}_i$ the i_{th} mode (infinitesimal atomic displacement) and $k(\mathbf{r})$ a force constant that is used to ensure a short-range interaction between domains. The constant r_0 defines the range of the force field and was obtained from numerical experiments. Rigid domains were then identified as regions that have E_i values smaller than a threshold. The threshold value (100kJ/mol) was chosen by its ability to define reasonable domains in a wide range of proteins (i.e., trial and error). Rigid domain definition was performed by first subdividing the protein volume into a set of cubes. The rotations and translations were then computed for each cube from the displacements of the atoms inside. The cubes were organized into rigid domains by performing a cluster analysis on their transformation parameters (rotation and translation). The results of an analysis of citrate synthase if shown in Figure 1 and agrees well with the known motions of that protein.

Using a normal mode approximation greatly reduces the computational cost of this method so that large proteins can be analyzed. The approximation does not seem to affect the accuracy of the results. The introduction of an elastic model of a protein via the deformation energy E_i provides a nice mechanical measure of motion. However, this energy measure is not physically based and must be correlated to known protein motions.

Subdividing the protein volume into cubes for domain classification is coarse-grained and seems like an arbitrary way to group atoms. The identification of domains using cluster analysis has no physical basis and is highly dependent on measures of homogeneity and separation that can be defined in several different ways. Different domains may result from changes in these parameters.

Mode Concentration Analysis

The basis of this method is to correlate the motion of a protein from a pair of different experimentally determined conformations with the most representative normal mode of the difference between the pair [18]. A measure of this correlation is called a *mode concentration* and represents how much of a protein's motion is concentrated into any low-frequency mode.

To reduce the memory requirements for the normal mode computation residues were approximated as a single mass centered at its C^{α} atom. Only the 20 lowest frequency modes were computed. The normal modes are then computed for this system using a *deformation force field* [19] that depends on the distance between atoms in the protein pair. The mode **v** best representing the protein motion is determined by a least squares fit to the protein motion. The mode concentration is then computed as the information contained in the vector **v**:

$$I = \_ -|v_i| \ln|v_i| \qquad 1 \leq i \leq \#\text{atoms}$$

The goal of this approach was to produce a large set of statistics for many different proteins. The mode concentration and other statistics were automatically generated for approximately 4000 proteins and deposited into a database. The results showed that statistically the direction of motion lies most often along the direction of two modes. The large amount of data was also used to develop a set of training sets to perform feature analysis using decisions trees, a form of supervised machine learning. The result of this was that mode concentration could be used in feature extraction, classifying motions as fragment, domain or subunit. For larger proteins the results produced no significant correlation, probably because of the involvement of more modes in the motion. Results for an immunoglobin elbow joint motion are shown in Fig. 2.

Model Analysis of Molecular Dynamics Trajectories

In this approach a modal analysis based on the singular value decomposition (SVD) of molecular dynamics (MD) trajectories of the C^{-} backbone atoms of a protein is performed. The modal analysis is then used to elucidate the collective motions of the atoms that are not readily accessible from MD trajectories [22].

The change in the time-evolved coordinates of the backbone atoms is organized into a *fluctuation trajectory matrix* **R** of dimension 3m x n (m = #atoms and n = #time steps). The matrix **R** can be decomposed into the product of three matrices using SVD as:

$$\mathbf{R} = \mathbf{U} \_ \mathbf{V}^{T}$$

where $\_ _{(3mx3m)}$ is the diagonal matrix of singular values of **R.** $\mathbf{U}_{(3mx3m)}$ and $\mathbf{V}_{(3mxn)}$ represent the time-averaged space-dependent features and time evolution of **R.** For a given singular value k a displacement vector field $\mathbf{u}_i(k)$ can be defined using these matrices for the ith C^{-} atom. The molecule volume is then divided up into volume slabs normal to each of the x, y and z-axes. The displacement vector field is averaged over all the atoms within each of the slabs to produce a 3x3 matrix **D**(k) of *collective displacements*. The deformation gradient tensor **F** and strain tensor **E** can then be computed from **D**(k). These tensors completely describe the stretching, shearing and rotational motion occurring within each of the slabs. Results for an 800 ps MD simulation of a human T-cell glycoprotein CD4 are shown in Fig. 3.

The advantage of this approach is that once you compute **D**(k) you can use the methods of solid mechanics to characterize many types of cooperative motions: rotation, breathing, wave-like, wagging, wiggling, etc. There are quite a few limitations though. The memory and compute time for this method is enormous for even small proteins. Exploring protein motion would only be possible for short time scales due to the computational limitations (small time step) of MD simulations. The coarse-graining of the displacements could also be a problem since some motions (e.g., hinge) occur in a very localized region and could be averaged over.

Graph Theoretic Method

In this method the characteristic flexibility and rigidity of a protein is determined from a single conformation. The protein is represented as a graph whose vertices are atoms and whose edges are distance constraints described by strong local forces (e.g., covalent and hydrogen bonds and salt bridges) in the protein. All other weaker forces are not included. Rigidity theory can then be used to determine the rigid clusters and the flexible joints connecting them [20].

Before the analysis can be performed the original PDB structure must have hydrogen atoms added to it since they are usually not defined. Each atom is defined as a vertex of the graph. The edges of the graph are defined as distant constraints between the atom's covalent bonds resulting from bond-stretching (central), bond-bending and torsional forces. Hydrogen bonds are modeled as three distance constraints consisting of one central-force constraint and two bond-bending constraints. Once the graph is constructed the degrees of freedom within this constrained network are counted using methods of

rigidity theory [21]. Based on the location and number of degrees of freedom within the graph, one can identify rigid clusters, flexible joints and stable regions that do not move. A flexibility index can then be computed that characterizes the degree of flexibility for each central-force bond. The results of an analysis of HIV protease if shown in Figure 4 and agrees well with the known motions of that protein.

This method is quite fast with execution time depending linearly on the number of atoms. The motion classification is fine-grained in that a flexibility index is computed for each main-chain atom making it possible to find hinge regions. The main limitation of this method is that it produces a static analysis of the protein. The results do not describe the dynamics of the motion. They can however be used as input to conformational sampling methods [23]. The method is also highly sensitive to model construction and certain parameters such as an energy threshold used to define hydrogen bond interactions.

Rigid Domain Detection from Atomic Coordinate Comparisons

In this method rigid domain movements about hinges are identified using pairs of protein conformations. The domains are extracted using least squares fitting and an adaptive selection procedure [26].

Let $\mathbf{x}_n$ and $\mathbf{y}_n$ be the two vector sets of atomic coordinates of the reduced representations consisting of only the N C^- backbone atoms of the two protein structures. A least squares procedure is used to fit $\mathbf{x}_n$ to $\mathbf{y}_n$ yielding an approximate set $\mathbf{x'}_n$ that is related to $\mathbf{x}_n$ by a rigid body transformation of the form:

$$\mathbf{x'}_n = \mathbf{U}\mathbf{x}_n + \mathbf{v}, \quad n = 1, 2, \ldots, N.$$

U is a rotation matrix and **v** a vector connecting the centroids of $\mathbf{x}_n$ and $\mathbf{x'}_n$. The *geometric conformance* between each of the atoms in a set is given by $d_n = \| \mathbf{x'}_n - \mathbf{y}_n \|$. The geometric conformance d_n is really just a definition of rigid body motion: a set of points move as a rigid body if $d_n = 0$ for all points in the set. This definition can be used to automatically find rigid domains by an *adaptive selection* procedure that iteratively creates rigid subsets from seed atoms, adding new atoms that satisfy a specified geometric conformance. Searches for rigid subsets are repeated for remaining atoms until the protein is partitioned into well-fitting substructures. The hinges can be found using the relative rotations between domains. The results of this method are shown in Fig 5.

This method is quite fast for small proteins (around 600 residues) and can identify multiple domain motions reasonably well. There are several limitations to the method however. For large proteins the method could be very slow since the adaptive selection algorithm appears to have computational complexity of order N^2. Only rigid body motions of domains could be classified. Therefore domain motions consisting of a

cumulative result of many small relative motions (i.e., shear) or flexible regions could not accurately be determined.

Discussion

The methods described above all make certain assumptions and simplifications, mostly for reasons of computational efficiency but also from a lack of physical theories describing the long-time and large-scale behavior of molecular systems. The assumptions and simplifications made by each method contribute to its particular advantages and disadvantages. We will now address some of these issues here.

Structural Simplification

A reduced representation consisting of only the C^{-} backbone atoms is often used in order to save memory and compute time. This effectively converts a protein, which is a solid object, into a 1D chain. This simplification neglects the packing constraints within a protein [1] producing an under-constrained system.

Data Dependencies

Methods requiring only a single protein conformation were based on modal analysis except for the Graph Theoretic method but that did not produce any dynamic results. The open debate about the usefulness of modal analysis makes the results of these methods suspect although the results of the statistical study did show a correlation for small proteins.

Methods requiring a pair of protein conformations had a lot more variability in approach but usually required an initial alignment of the structures to remove the rotation and translations associated with the experimental acquisition. One shortcoming of these methods is that they are limited by the diversity of the conformational states that are available from experiment for comparison. Currently only a small fraction of the 17,000 proteins in the PDB have multiple conformations.

The Rigid Domain Detection from Atomic Coordinates method performed an analysis on MD trajectories. If you already had MD simulation data (and a big machine) then this would provide a useful analysis especially since solvation and ligand binding could be included. One could also just use the initial and final simulations and perform an analysis using a method for conformation pairs. Running MD simulations specifically for domain analysis does not seem to be feasible given that the time scales of most domain motions are beyond its limit of a few hundred picoseconds.

Mechanical Models

All methods assume some sort of mechanical model of protein deformation. Most models assume rigid domain motions but both of the normal mode methods use an elasticity measure of deformation. Protein motion is highly damped making a viscoelastic model

more appropriate. Most methods did not include a solvation model that would affect the range of motions available.

Artificial Parameters

All methods use sets of artificial (e.g., nonphysical) parameters to produce desired results. Some examples are: deformation threshold for rigid regions, tuned energy thresholds for hydrogen bonds and assorted thresholds for domain classification. Many of these parameters were determined by comparison of results to known motions. This may preclude methods from producing correct results for a wide range of motions. It also explains the sensitivity of some methods to the values of certain parameters.

Domain Identification

All the methods except the Graph Theoretic Method had difficulty characterizing transitions of relatively rigid regions. This is important for identifying both hinge regions that usually consist of a few residues and shear regions that are composed of small regions of little motion. Most methods spatially decomposed the protein domain into regular shapes (e.g., cubes) for classification. This could lead to misclassification of rigid domains and averaging out motions (e.g., hinge) occur in a very localized regions.

One of the main problems with the domain classification methods described here is that regions are treated as collections of discrete objects, namely atoms. Since each atom has a volume associated with it a protein should be represented as a volume. A Voronoi packing could be defined for the atoms in a protein [1] that would allow for the definition of a continuous deformation field. Domain classification could then be done more rigorously using this continuous representation and continuum mechanics techniques.

Computational Complexity

The Rigid Domain Detection from Atomic Coordinates method was by far the most memory and compute intensive of all methods because of amount of data it needs to process. It is clearly not a PC application. The normal mode methods, even using approximate methods, required more memory than the methods using conformational pairs although their execution times were not too bad. The Graph Theoretic method was the least compute intensive method and was able to perform analysis on million atom models.

Conclusions

In this paper we have reviewed several different methods for the classification of protein domain motion and flexibility. The methods described above represent a large range of approaches to this problem, differentiated by computational complexity, execution speed, computer resources required, data sources, protein complexity (number of domains and size) and information content of the results.

Normal mode methods continue to play an important role in the analysis of protein motion even though they suffer from several limitations and their practical relevance seems questionable. For certain types of motion at least these methods seem to produce decent results. As shown in the statistical analysis of a large number of protein movements there does seem to be a correlation in direction of movement and modes.

With time-resolved X-ray and NMR there will be a lot more data available describing protein dynamics. Some of the methods for multi-conformational analysis could be applied to these data sets for a more detailed study of protein dynamics.

Techniques from solid mechanics have been used in all these methods to characterize protein motions. In the future computational mechanics techniques (i.e., multibody dynamics, finite element) may be able to contribute to these methods by providing better material models (e.g., viscoelastic models) of protein deformation and velocity and force analysis of protein dynamics.

Figures

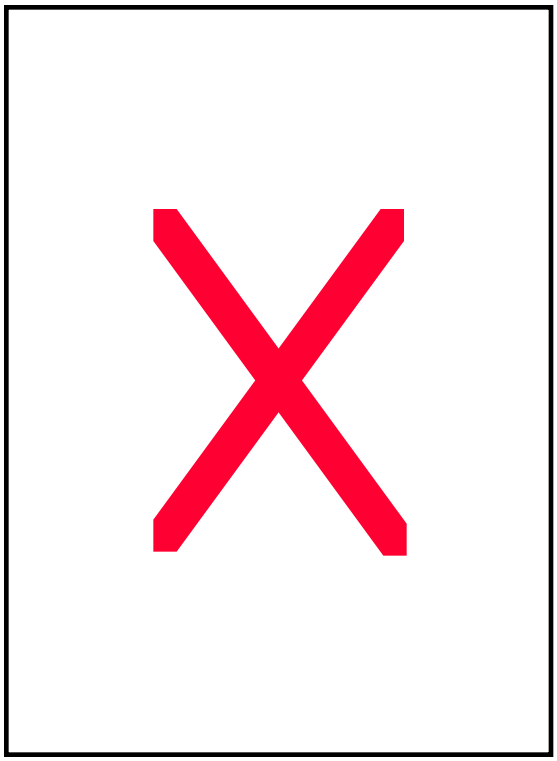

Fig. 1 The deformation energy (left column) and domain decomposition (right column) for citrate synthase, from conformation comparison (top) and normal modes (bottom). Blue regions in the deformation energy pictures are the most rigid ones, red indicates strong deformation. Green corresponds to the rigidity threshold used in the domain analysis. In the domain decompositions, the central orange region and the tips in yellow and green represent stable dynamical domains, the red, cyan, and pink parts are low-deformation interdomain regions.

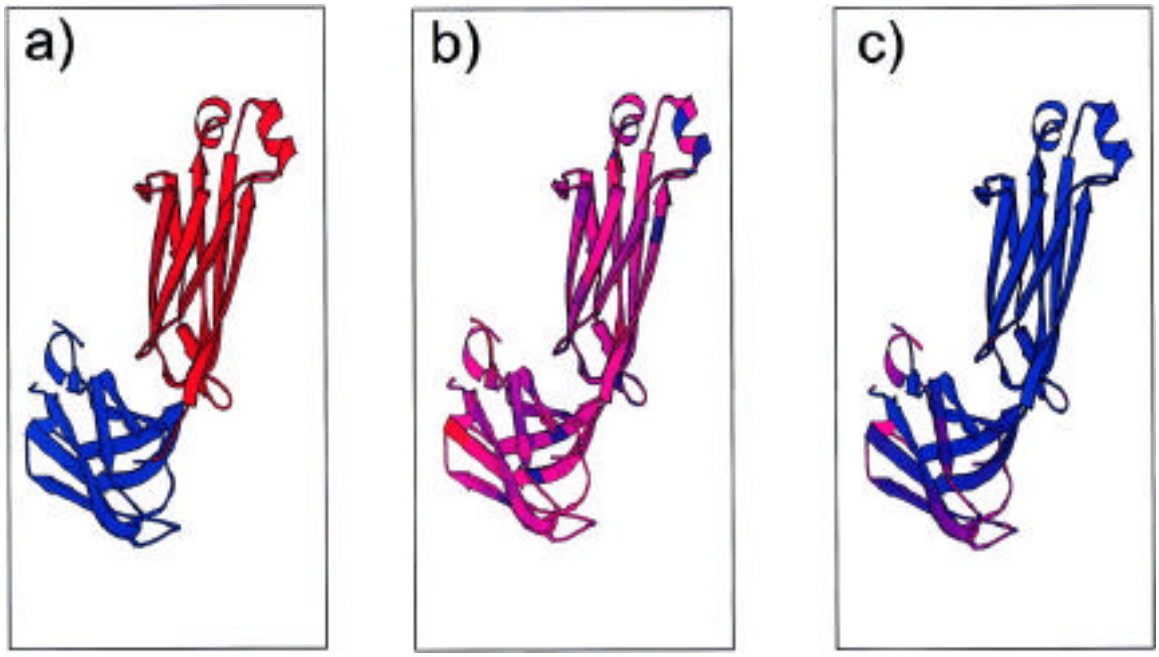

Fig. 2. a: The parts of the protein that actually move, as calculated from comparison of the starting and ending PDB structures for the motion. Areas that move are colored in red, while areas that remain stationary are colored in blue. The user may compare these three panels to deduce structural information. Hinge locations involved in the motion may be deduced, as these are highly flexible regions (as identified by a and b) located near the moving domains (show in red in c). **b:** Performs a normal mode flexibility analysis on the structure. Regions that are more flexible are colored in red, while less flexible regions are colored in blue. **c:** Similar information, using experimental temperature factors supplied in the PDB file.

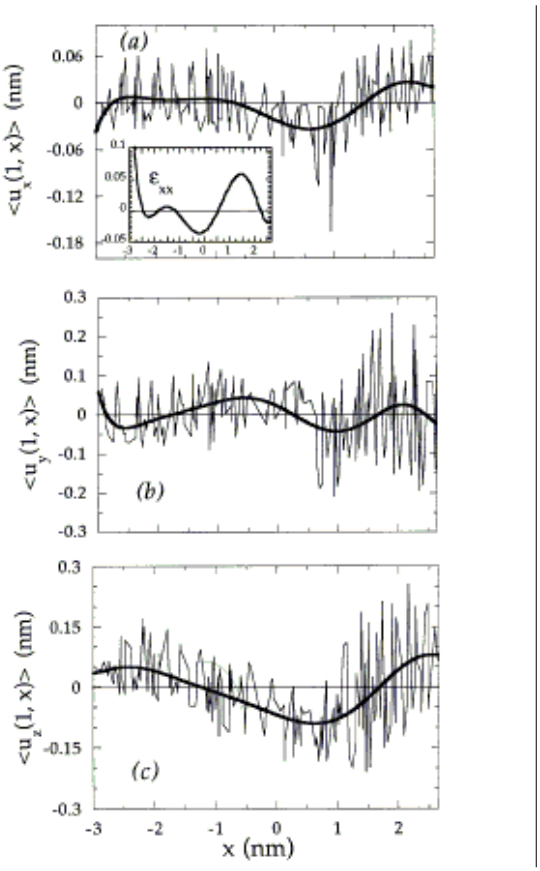

Fig. 3. Collective motions induced by the first (k = 1) dominant mode of motion identified by the SVD of a vacuum trajectory of 800 ps. Parts (a), (b) and (c) display the x-, y- and z-components of the mean displacement of the slab. The insert in part (a) displays the normal strain E_{xx} along the x-axis. The collective motions observed in parts (a) – (c) may be characterized as wave-like and wiggling types of motions.

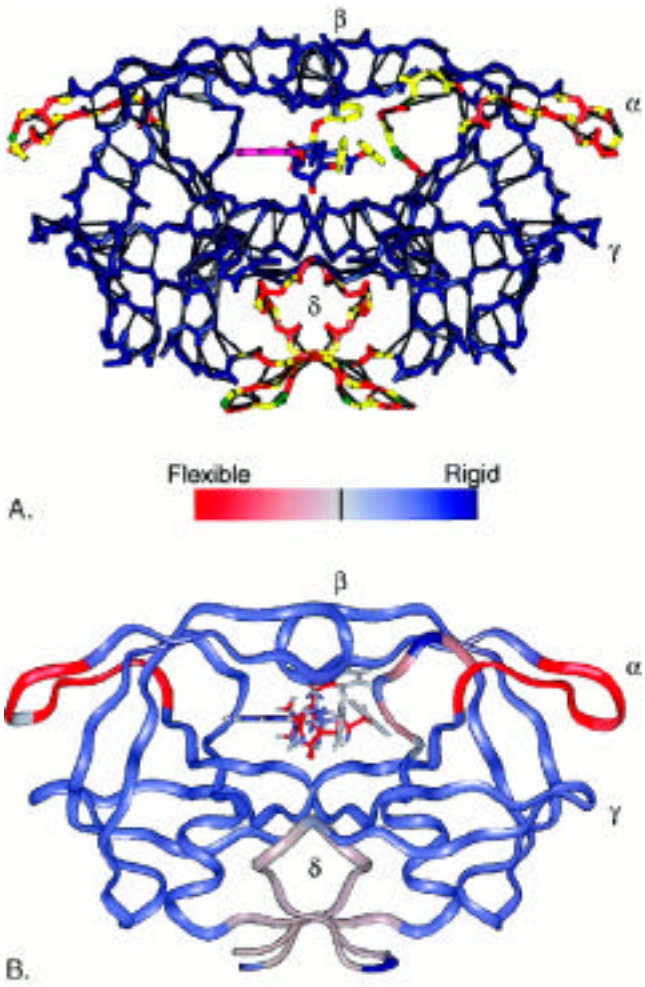

Fig. 4. A: Rigid cluster decomposition of the closed conformation of human immunodeficiency virus protease (HIVP) (PDB code 1htg). **B:** Flexibility index plot of the same, closed conformation of HIVP (PDB code 1htg). Four regions of interest, α, β, γ, and δ, are identified for one of the monomers.

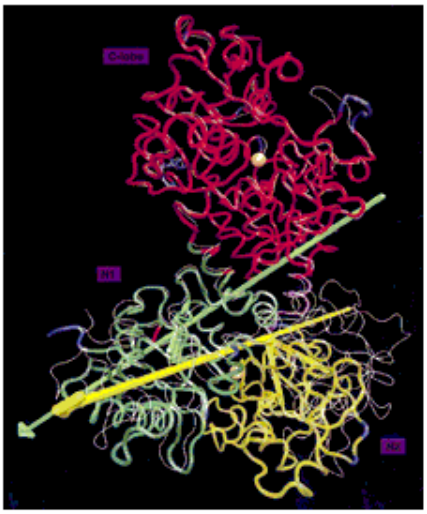

Fig. 5. Domain movements of lactoferrin. Shown are the backbone traces of iron-bound 33 (color) and iron-free lactoferrin 34 (white). Iron ions are shown in orange. Effective rotation axes and perpendicular centroid-connecting lines are rendered as tubes in the color of the corresponding domain. The arrows indicate a left-hand rotation, which shifts the center of mass of the domain in the iron-free structure onto the center of mass in the iron-bound structure. Three domains .15 residues have been found at 1.2 Å tolerance: domain 1 (red, 325 residues) is the reference domain that has been superimposed with the iron-free structure; domain 2 (green, 171 residues) rotates by 8° (relative error D 513%); domain 3 (yellow, 155 residues) rotates by 54° (relative error D 54%). Disordered regions (26 residues) are shown in blue.

References

[1] Gerstein M, Krebs W. A database of macromolecular motions. Nucleic Acids Res 1998;26:4280–4290.

[2] Debrunner PG, Frauenfelder H. Dynamics of proteins. Annual Rev of Phys Chem 1982;33:283.

[3] Lipscomb WN. Acceleration of reactions by enzymes. Accounts Chem Res 1982;15:232.

[4] Koshland, DE, Proc. Natl. Acad. Sci. *U. S.*, **4**4, 98 (1958).

[5] Genick UK, Borgstahl GE, Ng K, Ren Z, Pradervand C, Burke PM, Srajer V, Teng TY, Schildkamp W, McRee DE, Moffat K, Getzoff ED. Structure of a protein photocycle intermediate by millisecond time- resolved crystallography. Science 1997;275:1471–1475.

[6] Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science 2001;291:2429–2433.

[7] Sano T, Vajda S, Smith CL, and Cantor CR. Engineering subunit association of multisubunit proteins: A dimeric streptavidin. Proc. Natl. Acad. Sci., 94: 6153-6158, 1997.

[8] Carlson HA, McCammon, AJ. Accommodating protein flexibility in computational drug design. Mol. Pharmacol. 2000, 57, 213.

[9] Scott W, Schiffer C. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. Struct Fold Design 2000;9:1259–1265.

[10] Genick UK, Borgstahl GE, Ng K, Ren Z, Pradervand C, Burke PM, Srajer V, Teng TY, Schildkamp W, McRee DE, Moffat K, Getzoff ED. Structure of a protein photocycle intermediate by millisecond time- resolved crystallography. Science 1997;275:1471–1475.

[11] Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science 2001;291:2429–2433.

[12] Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements. Biochemistry 1994;33:6739–6749.

[13] Nishikawa K, Ooi T, Isogai Y, Saito N. Tertiary structure of proteins. I. Representation and computation of the conformations. J Phys Soc Japan 1972;32:1331–1337.

[14] Huang ES, Rock EP, Subbiah S. Automatic and accurate method for analysis of proteins that undergo hinge-mediated domain and loop movements Curr Biol 1993;3:740–748.

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

[16] Miller DW, Agard DA. Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. J Mol Biol 1999;286:267–278.

[17] Hayward S, Go N. Collective variable description of native protein dynamics. Annu. Rev. Phys. Chem. 46:223-250. 1995

[18] Krebs WG, 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. 2002 Sep 1;48(4):682-95.

[19] Hinsen K. The molecular modeling toolkit: a new approach to molecular simulations. J Comp Chem 2000;79–85.

[20] Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein Flexibilty Predictions using Graph Theory Proteins **4**4, 150 - 165, (2001).

[21] Jacobs D, Kuhn LA, Thorpe MF. Flexible and rigid regions in proteins. In: Thorpe MF, Duxbury P, editors. Rigidity theory and applications. New York: Kluwer Academic/Plenum; 1999. 7:961–967.

[22] Doruker, P.; Bahar, I.; Baysal, C.; Erman, B. Collective deformations in proteins determined by a mode analysis of molecular dynamics trajectories. Polymer 43: 431-439, 2002.

[23] M.F. Thorpe, Ming Lei, A.J. Rader, Donald J. Jacobs, and Leslie A. Kuhn Protein Flexibility and Dynamics using Constraint Theory. J. Molecular Graphics and Modelling 19, 60-69, 2001

[24] Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins Proteins 34, 369 (1999)

[25] Hinsen K. Analysis of domain motions by approximate normal mode

calculations. Proteins 1998;33:417–429.

[26] Wriggers W, Schulten K.
Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates.
Proteins, Sep 1997; 29(1): 1-14.