Sampling Protein Conformations and Pathways

Ming Lei
Department of Biochemistry
Brandeis University, Waltham MA 02454

Maria I. Zavodszky, Leslie A. Kuhn
Department of Biochemistry
Michigan State University, E Lansing MI 48824

M. F. Thorpe
Department of Physics and Astronomy
Arizona State University, Tempe AZ 85287

J. Comput. Chem., submitted, October 2003

Abstract

Protein flexibility and rigidity can be analyzed using constraint theory, which views proteins as 3D networks of constraints involving covalent bonds and also including hydrophobic interactions and hydrogen bonds. This paper describes a new algorithm, ROCK (Ring Optimized Conformational Kinetics), which generates new conformations for these complex networks with many inter-locked rings while maintaining the constraints. These new conformations are tracked for the flexible regions of a protein, while leaving the rigid regions undisturbed. An application to HIV protease demonstrates how large the flap motion can be. The algorithm is also used to generate conformational pathways between two distinct known protein conformations. As an example, directed trajectories between the closed and the occluded conformations of the protein dihydrofolate reductase are determined.

keywords: protein conformations, conformational pathways, constrained dynamics, flexibility analysis, protein flexibility

1 Introduction

Proteins change their conformations to perform certain biochemical functions. It is this ability to sample different conformations with functional significance and hence to carry out chemistry on their molecular partners that distinguishes proteins, particularly enzymes, from polymers and non-crystalline networks. For example, adenylate kinase (ADK)[1–4] has large scale domain motions related to its catalytic cycle. Calmodulin, which regulates a variety of cellular processes, also goes through large scale conformational changes in its catalytic pathway.[5,6] The electron transfer function of the protein cytochrome bc_1 is directly coupled with the conformational changes of the protein.[7]

Computational modeling is useful in studying protein conformational transitions. Knowledge based algorithms,[8–10] Molecular Dynamics (MD) simulations and Monte Carlo (MC) samplings are three major categories of approaches that have been used previously to sample protein conformations.

In MD simulations, every snapshot of the motion trajectory is a viable conformation. Recent developments in MD simulations are reviewed by Wang et al.[11] In the multicanonical MD simulations,[12,13] the distribution of sampling conformations at different energies is artificially flattened,[14] so that the probability of jumping over energy barriers is enhanced. Multiple MD trajectories are sampled in parallel at different temperatures in the replica exchange MD algorithm.[15,16] Trajectories at different temperatures are periodically exchanged so that they all have the possibility to overcome high energy barriers at high temperatures. Multicanonical MC[17–20] and replica-exchange MC algorithms[21] have also been applied to study protein dynamics and conformations.

In this paper, we lock covalent bond lengths and angles and also introduce other constraints (hydrophobic interactions and hydrogen bonds) to reduce the complexity of the energy landscape. Then, the effective potential becomes infinitely high when bond lengths and angles and other constraints are violated. The major difficulty in sampling protein conformations in this way, while preserving the network of constraints, is to close all the rings.

For example, Fig (1) shows the tip of one flap of the human immunodeficiency virus (HIV)-1 protease. Four independent rings are formed from covalent and hydrogen bonds in this small region of the protein. Composite rings, that can be decomposed into smaller rings, are not counted as independent rings. Being able to sample the conformations of a network that is dense in such interlocking and independent rings is the key to our approach.

The ring closure equations proposed by Gō and Scheraga[22] state the conditions under which a ring is closed and give a procedure to solve the ring closure equations. Their subsequent work[23] exhausts the conformations of a short cyclic peptide segment. They later developed a method to close large rings when the rings have C_n, I, or S_{2n} symmetries.[24] The conformations of gramicidin S, a cyclic peptide with 18 rotatable bonds, are sampled by this procedure,[25] assuming the molecule has an exact C_2 symmetry. The conformations of the molecule *cyclo*-hexaglycyl were also generated,[26] considering its internal symmetry, and checked against conformations generated by MC algorithms.[27] Wedemeyer and Scheraga showed that a ring is closed when its dihedral angles are roots of polynomial equations and developed the form and the solutions to the polynomial functions for seven-fold and eight-fold rings.[28] Wu and Deem[29] transform the bond length and angle constraints at the two ends of a polymer to a single-variable polynomial function which can be solved numerically. Unlike other algorithms that optimize all of the unknown dihedral angles simultaneously to close a ring, the cyclic coordinate descent algorithm by Canutescu and Dunbrack[30] optimizes them one at a time. Bruccoleri and Karplus[31] allow bond angles to be relaxed when a ring cannot be closed exactly under the condition of fixed bond lengths and angles. All the algorithms listed above close *single* rings. A protein consists of zero or only a few rings when covalent and disulfide bonds are treated as components of rings. Gibson and Scheraga write the bond length and angle constraints at the disulfide bond as a pseudo-potential.[32] The ring closes if the pseudo-potential is zero. They show several examples where three or four rings containing disulfide bonds are closed simultaneously.

A protein has *multiple* inter-locking rings when non-covalent interactions and covalent

bonds are taken together to form a network. For example the tip of a flap of HIV-1 protease has four interlocking rings, as shown in Fig. (1). All of these rings have to be closed exactly and simultaneously as different new conformations are generated.

Hydrogen bonds are important in stabilizing proteins, but they have not been included as viable components of rings in previous studies. When hydrogen bonds are treated as component of rings, the number of inter-locking rings rises steeply. A protein has only three or four rings when the covalent bonds and disulfide bonds are counted as the components of rings, but hundreds of rings when hydrogen bonds are also included. These rings inter-lock with each other in a complex way so that a rotation of one dihedral angle can potentially break several rings. For this reason, we have devised a new algorithm to close a large number of rings simultaneously while generating new protein conformations.

Hydrophobic interactions are also crucial for the protein stability. However, the hydrophobic interactions are qualitatively different from hydrogen bonds and covalent bonds in that they are often not angle-specific. A hydrophobic interaction specifies the allowed range for the distance between two atoms but not the angles. Therefore hydrophobic interactions are not viable components of rings which we take to imply fixed angles, though they are as equally important as hydrogen bonds in maintaining the stability of proteins, and are included in our procedure as is described later.

We view proteins as complicated networks made up of covalent bonds and hydrogen bonds. The network is also kept in a compact shape by the hydrophobic interactions. Regions of a protein are essentially rigid where the concentration of covalent bonds, hydrogen bonds and hydrophobic interactions is high enough. These rigid regions cannot be continuously deformed. Methods without prior knowledge of which parts of the protein are rigidified by non-covalent interactions waste time on trying unsuccessfully to move rigid regions.

This paper presents a new approach to sample protein conformations by closing *all* the rings in a complicated network simultaneously. Our algorithm is based on the work of Scheraga and colleagues. It differs from that of Gibson and Scheraga[32] in several aspects.

First, the fictitious potentials of ring closure conditions used in the two algorithms are different. Second, hydrogen bonds are included as components of rings in our algorithm. Third, our algorithm handles a very much larger number of rings than previous algorithms have done.

Our algorithm is capable of generating conformational pathways between distinct protein conformations. These new conformations are guaranteed to have good bond lengths and angles and good stereo-chemistry, including favored values of the main chain dihedral angles.

Section 2 reports the details of the algorithm to sample protein conformations. Section 3 discusses the properties of the conformations generated for HIV-1 protease. The conformational pathways between the closed and the occluded conformations of *Escherichia coli* dihydrofolate reductase (ecDHFR) are discussed in Section 3. Section 4 summarizes the results and presents conclusions.

2 Methodology

2.1 Flexibility Analysis

Covalent bonds are the most stable interactions in proteins. They are rarely broken or altered. When only the covalent bonds are taken into account, a protein is an amino acid chain with many internal degrees of freedom (DOF). It is the hydrophobic interactions and hydrogen bonds that are responsible for folding and stabilizing proteins. From a topological point of view, the hydrophobic interactions and hydrogen bonds reduce the DOF in proteins, which are associated with rotatable dihedral angles. The DOF in some regions of a protein can be reduced to zero because of the high density of hydrophobic interactions and hydrogen bonds. These are the rigid regions of the protein, which do not sample multiple conformations at ambient temperatures. There remain DOF in other regions of most proteins, making it possible for most proteins to sample new conformations by changing the rotatable dihedral angles comprising the internal DOF. Typically proteins have one large rigid region forming

the core of the structure, which functions to stabilize the three dimensional structure of the protein. Smaller rigid regions often also exist. The flexible regions carry out biochemical functions through conformational changes and also contribute favorably to the free energy by maintaining some entropy in the native state.

Jacobs et al.[33,34] developed an algorithm to count the DOF in local regions of generic 3D networks. This algorithm is implemented in the software Floppy Inclusion and Rigid Substructure Topography (FIRST).[35] FIRST has been used to identify the flexible and the rigid regions of proteins,[33,36] which is called *flexibility analysis* or a *rigid region decomposition*. Flexibility analysis has also been used in the studies of protein unfolding pathways.[37,38] In all these studies, protein stability and flexibility are interpreted in terms of the topological properties of the 3D networks which are made up of constraints involving covalent bonds as well as hydrophobic interactions and hydrogen bonds.

One important question that remains to be answered is: how much conformational space can a protein sample while subject to the same constraints of bonds and interactions used in the flexibility analysis? It requires an innovative algorithm to understand how a flexible region of a protein samples conformations by following the trajectories involving the various DOF, once the distribution of DOF has been defined by the FIRST flexibility analysis or other methods. Here we present an algorithm to solve this problem, which has been implemented as the program ROCK (Rigidity Optimized Conformational Kinetics). ROCK achieves efficiency by sampling conformations for the flexible regions of proteins only, since the rigid regions of proteins will not have variable conformations. Though the atoms in the rigid regions have high frequency vibrations about the equilibrium structure, these vibrations do not change the average relative distances between atoms, nor lead to any kind of low frequency or diffusive motions. Therefore the high frequency vibrations can be ignored for the purpose of sampling significant conformational changes. The flexible regions of proteins can be observed by crystallography, NMR and FRET techniques to have large scale conformational changes at ambient temperatures. These conformational changes are of low

frequency, and may involve biochemical functionality. In the current study, we present the details of the ROCK algorithm and use it to explore the conformational pathways involved in the bioactivity of HIV-1 protease and ecDHFR. ROCK has also been used to sample new conformations around the equilibrium structure of cyclophilin A, providing an ensemble of main chain structures that correlates well with the NMR ensemble and hydrogen-exchange protection factors for this protein, and which provides a basis for exploring the range of protein main chain flexibility compatible with ligand recognition.[39]

Proteins are made up of many inter-locking rings and dangling side branches when they are viewed as 3D networks of covalent bonds and hydrogen bonds. A ring is a closed loop of bonds which connects any two atoms by two distinct paths. Two rings are said to be *interlocked* when they share common bonds. A *ring cluster* is the collection of rings that are interlocked with each another. All other atoms that are not part of the ring clusters are defined to be the *side branches*. The trivial rings in the residues of histidine, phenylalanine, proline, tryptophan and tyrosine may be in the ring clusters, or in the side branches, depending upon whether they share bonds with other rings in the proteins. Traditionally atoms in proteins are classified as main chain atoms and side chain atoms based on their location. In this work, the classification of ring clusters and side branches is based on the topology of the bonds and interactions. A side chain atom may belong to a ring cluster or could be a side branch, depending on whether it is part of a ring system of covalent and non-covalent bonds. Most of the main chain atoms are in the ring clusters, due to extensive main chain hydrogen bonding that forms the regular secondary structures in proteins, though the main chain atoms close to the N and C termini may form side branches.

2.2 Sampling Conformations

2.2.1 Single Ring Closure

It is relatively easy to generate conformations for side branches. The rotation of any dihedral angle in a side branch produces a valid new conformation, because the rotation does

not change the topology, nor the bond lengths and angles of the side branch. However, it is not trivial to generate conformations for ring clusters, because the dihedral angles are all correlated. A rotation of a dihedral angle in a ring cluster requires all the other dihedral angles to be rotated appropriately so as to close all the rings. We have reported previously an algorithm[40,41] that closes all the rings in ring clusters simultaneously. This paper reports further developments and applications of this approach to proteins. For the sake of completeness and clarity the algorithm, as it is currently being used, is reviewed here.

A single N-fold ring closes if its bond lengths, bond angles and dihedral angles satisfy the ring closure equations:[22]

$$\begin{aligned} \mathbf{P}_0 + \mathbf{T}_0\mathbf{R}_1\mathbf{P}_1 + \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\mathbf{P}_2 + \cdots + \mathbf{T}_0\mathbf{R}_1\cdots\mathbf{T}_{N-2}\mathbf{R}_{N-1}\mathbf{P}_{N-1} &= \mathbf{0} \\ \mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\cdots\mathbf{T}_{N-2}\mathbf{R}_{N-1}\mathbf{T}_{N-1}\mathbf{R}_N &= \mathbf{I} \end{aligned} \qquad (1)$$

where $\mathbf{T}_i$, $\mathbf{R}_i$ and $\mathbf{P}_i$ are rotational matrices and distance vector defined as:

$$\mathbf{T}_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i & 0 \\ \sin\theta_i & \cos\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\omega_i & -\sin\omega_i \\ 0 & \sin\omega_i & \cos\omega_i \end{pmatrix} \quad \mathbf{P}_i = \begin{pmatrix} d_i \\ 0 \\ 0 \end{pmatrix} \qquad (2)$$

The $\mathbf{0}$ and $\mathbf{I}$ are the zero vector and the unit matrix respectively. The quantities d_i, θ_i and ω_i are the bond length, the supplementary angle of the bond angle, and the dihedral angle of the ith bond in the ring. The bond lengths and angles are fixed parameters while the dihedral angles are unknown variables.

There are N bond lengths constraints and N bond angle constraints in a single N-fold ring. The total DOF of N atoms is $3N$. The internal DOF of the N-fold ring is the total DOF of N atoms minus the total number of constraints plus the six rigid body motions, which gives $3N - N - N - 6 = N - 6$. This means when $N - 6$ dihedral angles in a single N-fold ring are known, the remaining 6 dihedral angles can be obtained by solving the ring

closure equations. The general procedure of sampling conformations for a single N-fold ring is to alter $N-6$ dihedral angles systematically or randomly, and to solve the ring closure equations to obtain the proper values of the 6 unknown dihedral angles at every step.

Gō and Scheraga[22] showed that there are six independent equations in Eq. (1). They also proposed an algorithm to solve the ring closure equations. This method has been extremely useful, but is limited to a single ring. Their method is generalized here to sample the conformations of complicated networks in which many rings are inter-locked with one another. Proteins, when covalent bonds and hydrogen bonds are all counted, are such examples of complicated networks. In the next section we describe an approach that can handle many inter-locked rings.

2.2.2 Multiple Ring Closure

We define a fictitious ring closure potential $\mathcal{F}$ of a single N-fold ring as:

$$\begin{aligned}\mathcal{F} = &\left[\mathbf{P}_0 + \mathbf{T}_0\mathbf{R}_1\mathbf{P}_1 + \cdots + \mathbf{T}_0\mathbf{R}_1\cdots\mathbf{T}_{N-2}\mathbf{R}_{N-1}\mathbf{P}_{N-1}\right]^2 \\ &+ \sum_{i,j=1}^{3}\left[\mathbf{T}_0\mathbf{R}_1\mathbf{T}_1\mathbf{R}_2\cdots\mathbf{T}_{N-2}\mathbf{R}_{N-1}\mathbf{T}_{N-1}\mathbf{R}_N - \mathbf{I}\right]_{ij}^2 \end{aligned} \tag{3}$$

which is the sum of the squares of the differences between the left and the right sides of Eq. (1). Since the fictitious potential is non-negative everywhere, it is at one of its minima whenever its value is zero. Therefore by minimizing the fictitious potential, our algorithm is able to numerically solve the ring closure equations. A single N fold ring is closed if and only if the fictitious potential $\mathcal{F}$ is minimized to be zero. To sample the conformations of a single N-fold ring with $N > 6$, one can try all the possible combinations of the $N-6$ dihedral angles, and minimize the fictitious potential $\mathcal{F}$ with respect to the six unknown dihedral angles at every step. Each zero value of $\mathcal{F}$ corresponds to a solution to the ring closure equations, which in turn suggests a new conformation of the ring. The limited-memory BFGS source code[42] is used to minimize the fictitious potential $\mathcal{F}$.

The advantage of this algorithm is that it can be extended to close all the rings in a ring cluster with minor modification. To close all the rings in a complicated network, we minimize the total fictitious potential of a ring cluster, $\mathfrak{F}$, which is the sum of the fictitious potentials for every ring in the ring cluster:

$$\mathfrak{F} = \sum_{\text{all rings}} \mathcal{F} \tag{4}$$

The total fictitious potential of the whole ring cluster can then be minimized with respect to all the rotatable and unknown dihedral angles, in the hope of finding a zero potential point that can then become a new nearby conformation

The computation cost of minimizing the total fictitious potential $\mathfrak{F}$ rises with a power law. Suppose a network has N bonds and $\mathcal{M}$ DOF. Once the dihedral values of $\mathcal{M}$ bonds are set, there are $\mathcal{N} = N - \mathcal{M}$ unknown dihedral angles in the network. The computation cost of minimizing $\mathfrak{F}$ with respect to $\mathcal{N}$ unknown variables will be in the order of $\mathcal{N}^3$, which is roughly N^3. A flexible region in a small to moderately-sized protein can have hundreds of bonds. The computation cost would be unbearably expensive. Therefore it is not feasible to minimize $\mathfrak{F}$ with respect to $\mathcal{N}$ dihedral angles in a single computer run. It is much better to close rings in a ring cluster one at a time. Suppose the number of rings in the network is n, the average number of variables per ring is thus $\mathcal{N}/n$. The computational cost of solving ring closure equations for every ring individually for one time is thus of the order of $(\mathcal{N}/n)^3 \times n = \mathcal{N}^3/n^2$. Since both n and $\mathcal{N}$ scale roughly linearly with N in most protein structures, the computation cost in theory scales linearly, if ring closure equations of all rings are solved one by one. In practice, however, this method does not work because all the other rings in the ring cluster break when one single ring is closed by either Gō and Scheraga's technique or by our approach. It is not possible to close all the rings in an inter-locked cluster simultaneously by closing one ring at a time.

In order to solve the ring closure equations for all the rings in a ring cluster as efficiently

as possible, we design a procedure that minimizes the total fictitious potential $\mathfrak{F}$ gradually. The algorithm first attempts to close the ring which has the smallest number of unknown dihedral angles in the cluster. This ring is called the seed. After successfully closing this ring, the algorithm then minimizes the sum of the fictitious ring closure potentials of the seed and of up to five more rings that share bonds with the seed. The seed is expanded to include both the old seed and the newly added rings. If all rings in the seed can be closed simultaneously, ROCK then adds up to five more rings to the seed. Step by step, ROCK adds rings to the expanding seed, and then minimizes the sum of the fictitious ring closure potentials of all the rings in the seed. Because the rings in the seed are already closed when new rings are added, only small adjustments on dihedral angles are necessary to close all the rings in the seed concurrently. The whole process stops when all the rings in the ring cluster have been added to the seed. The total calculation cost of this procedure is lower than minimizing the total fictitious ring closure potential of all the rings directly.

ROCK activates conformational changes by randomly rotating a set of selected bonds by a small amount. The rings are then closed as the total fictitious ring closure potential of the flexible ring cluster is minimized to be zero. The dihedral angles of the selected bonds are not altered in the optimization process. Hence the new conformation is certain to be different from the old one. The set of bonds to be randomly rotated is selected afresh in each trial to generate a new conformation.

ROCK utilizes the following procedure to select a set of bonds to rotate:

1. Randomly select a rotatable bond.

2. Count the DOF in the ring to which the selected bond belongs. Go back to Step 1 if the DOF is negative which means the ring is over-constrained. One more constraint is counted because a randomly selected bond is equivalent to one more constraint on the dihedral angles. The ring is considered to be the seed of the ring cluster.

3. Expand the seed by one more ring which shares bonds with the rings in the seed. The

DOF of the seed is calculated. Every randomly selected bond is counted as one more constraint. Go back to Step 1 if the DOF is negative.

4. Repeat Step 3 until all rings in the ring cluster are included in the seed. The bonds selected in Step 1 are then randomly rotated. Local regions in the ring cluster will not be over-constrained by the rotation of this bond.

5. Repeat Step 1 to Step 4 until a desired number of bonds are rotated.

The procedure listed above ensures that the randomly selected and rotated bonds do not over-constrain any local area of the network. It reduces the rate of unsuccessful trials. Randomly selecting and rotating fewer bonds than the DOF further improves the rate of successful trials.

The discussion so far in this paper assumes every bond in a ring cluster is rotatable. There are bonds which, however, should be considered to be locked. The peptide bonds, for example, favor either the *trans* or the *cis* conformation. There are energy barriers between these two conformations. The dihedral angles of these bonds should be kept unchanged from their values in the initial conformation. The procedure outlined above still applies, except that each fixed bond is equivalent to one less unknown dihedral angle.

2.2.3 Side Branches

Once new conformations of the flexible ring clusters are generated, side branches are anchored to the ring clusters with their correct bond lengths and angles. To begin the repositioning of the side branch atoms, they are first randomly disturbed. The coordinates of the side branch atoms are then relaxed in the Cartesian coordinates so that 1) bond lengths and angles of side branch atoms are undistorted from the original values; 2) there are no van der Waals overlaps between the side branch atoms themselves and between the side branch atoms and the ring cluster atoms; and 3) chirality at side branch atoms is maintained. ROCK uses the software DONLP2,[43] which is a non-linear optimization program that minimizes a function

subject to equality and inequality constraints, to minimize the function

$$f(x) = \sum_{\text{bonds}} (r - r^0)^2 + \sum_{\text{angles}} (\theta - \theta^0)^2 \tag{5}$$

subject to a collection of inequality constraints of van der Waals repulsions

$$g_1(x) = r_{ij}^2 - r_v^2 \geq 0 \tag{6}$$

and a set of inequality constraints to maintain the chirality:

$$g_2(x) = [\mathbf{r}_{ij} \cdot (\mathbf{r}_{ik} \times \mathbf{r}_{il})] \left[\mathbf{r}_{ij}^0 \cdot (\mathbf{r}_{ik}^0 \times \mathbf{r}_{il}^0)\right] \geq 0 \tag{7}$$

in which r and θ are the current values of bond lengths and angles, r^0 and θ^0 are the corresponding bond lengths and angles in the initial conformation. The distance r_{ij} is between two non-bonded atoms of atom i and atom j. The distance r_v is the sum of the van der Waals radii of the atom i and atom j times a coefficient which specifies the stiffness of van der Waals repulsions. The vectors $\mathbf{r}_{ij}$, $\mathbf{r}_{ik}$ and $\mathbf{r}_{il}$ are between atom i and its three bonded neighbors, with the vectors $\mathbf{r}_{ij}^0$, $\mathbf{r}_{ik}^0$ and $\mathbf{r}_{il}^0$ being between corresponding atoms in the initial conformation. The sign of the dot and cross products of the three vectors at the atom i specifies the chirality of the atom. The chirality of atom i in the generated conformation is required to be identical to that of the same atom in the initial conformation and this is achieved if and only if the sign of the dot and cross products of the three vectors is unchanged. The function $f(x)$ is minimized to be zero when the three requirements on side branches are satisfied.

After randomly disturbing the side branch atoms from their original Cartesian coordinates, ROCK checks the distances between every pair of non-bonded atoms to build a van der Waals overlap list. It constructs the inequality constraints $g_1(x)$ according to this list. Then it minimizes the function $f(x)$ subject to the inequality constraints. Once the function

is minimized to be practically zero, ROCK builds a new van der Waals overlap list. A new conformation of the side branch is found when there are no van der Waals overlaps, when the function $f(x)$ is practically zero, and when the chirality at those atoms that have at least three nearest neighbors is conserved.

2.2.4 Hydrophobic Interactions and Ramachandran Checks

Three types of interactions are taken into account when the flexibility analysis counts the local distribution of DOF in proteins: the covalent bonds, the hydrophobic interactions and the hydrogen bonds. The flexibility analysis defines a hydrophobic interaction whenever two carbon atoms are within certain distance limit. Unlike the covalent and the hydrogen bonds, the strength of the hydrophobic interactions is not angle-specific. The only parameter that determines the presence of a hydrophobic interaction is the distance between the two carbon atoms. Because all the angles in the rings are fixed when rings are exactly closed, and because hydrophobic interactions are not angle-specific, ROCK does not count the hydrophobic interactions as viable components of rings. ROCK first reads in the list of hydrophobic interactions from the flexibility analysis results. It then generates conformations without any considerations of preserving the hydrophobic interactions. After each conformation is generated, ROCK checks the distances between the pairs of carbon atoms that are on the hydrophobic interaction list. If the distances are all within the hydrophobic interaction distance limit the newly generated conformation is accepted. Otherwise it is rejected. In this way, the hydrophobic interactions used in the flexibility analysis with FIRST are preserved in ROCK. Thus the hydrophobic interactions are treated as a posteriori as an inequality in generating new conformations.

It is worth noting that though FIRST and ROCK handle hydrophobic interactions differently from the technical point of view, the concept of hydrophobic interaction is consistent in the two programs. FIRST inserts three pseudo-atoms between a pair of non-bonded carbon atoms when their distance is less than a limit. The distance between the pair of carbon

atoms changes when the dihedral angles of the bonds between the pseudo-atoms are freely rotated. In this way, FIRST allows the "bond length" of a hydrophobic interaction to change within a range, while imposes no constraints on the "bond angle". ROCK follows the same rule in handling the hydrophobic interactions. It forces the "bond length" of a hydrophobic interaction to be within a specific range by a posteriori check.

ROCK also checks the quality of main chain ϕ and ψ angles against the Ramachandran plot[44] to ensure the stereo-chemical quality of the generated conformations. Eighteen out of the twenty standard residues (alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, histidine, isoleucine, leucine, lysine, methionine, phenylanaline, serine, threonine, tryptophan, tyrosine and valine) are checked against the Ramachandran plot generated by Morris et al.[45] The other two residues, glycine and proline, are checked against two Ramachandran plots specially designed for glycine and proline. We thank Dr. Roman Laskowski of European Bioinformatics Institute for the data of the distribution of main chain ϕ and ψ angles of glycine and proline. The main chain ϕ and ψ angles of the twenty standard residues are restricted to the so-called *core* and the *allowed* regions[45] of the Ramachandran plot by default.

2.3 Conformational Pathways

Multiple and distinct conformations are experimentally detected for several proteins. Proteins such as Nitrogen regulatory protein C (NtrC),[46] cyclophilin A[47] and annexin V[48] are such examples. There is not yet an experimental technique to detect an ensemble of conformational pathways between two known distinct conformations. The NMR technique, which measures the average structure of an ensemble of conformations, can detect the structure of the most populated conformations, but not those conformations that are less populated. The intermediate conformations between two distinct conformations are not observed experimentally.

MD simulation, being the standard computational algorithm in exploring protein confor-

mational changes, is in practice not be able to sample the whole conformational pathway(s) between known and distinct protein conformations because the time range of these conformational changes is usually beyond the calculation capacity of current MD simulations. Typical large scale protein conformational changes are in the time range of microseconds and milliseconds to seconds. The state of art MD simulations currently can reach a few microseconds, with the majority of MD simulations being limited to a few nanoseconds. MD simulation is limited in its capacity to sample long time scale protein conformational changes partly because it wastes time on high frequency motions, and partly because it includes the whole protein in the calculation. The rigid regions of the proteins do not undergo significant structural transformations. The structural stability of the proteins supports our interpretation of proteins as flexible regions anchored to a rigid core. MD simulations waste time on calculating the dynamic trajectory of the rigid cores of proteins which is not of interest in studies of conformational changes.

Several algorithms are able to sample conformational pathways; each based on various simplifications. The morphing algorithm[49] builds trajectories by interpolating the two ends followed by a refinement process. The algorithm proposed by Kim and his coworkers[50,51] interpolates the distance between C_α atoms in the elastic network model framework.

ROCK eliminates the computational time expended on high frequency motions by preserving the constraints used in FIRST. These constraints include the bond length and angle constraints associated with the covalent and strong hydrogen bonds. By using the flexibility analysis results from FIRST, ROCK is able to sample the conformations of the flexible regions of the proteins only. These two advantages make ROCK a powerful tool in sampling the conformational changes that are beyond the scope of MD simulations. Being able to avoid van der Waals collisions, to maintain correct bond lengths and angles, to keep main chain dihedral angles reasonable, and to include all atoms in simulations, makes ROCK a reliable program to sample protein conformational pathways.

While designed to perform a random search from one starting conformation, ROCK

can be tailored to search directed pathways between distinct proteins conformations by incorporating a simulated annealing procedure[52,53] in the algorithm. The root mean square deviation (RMSD) d between a generated conformation and the target conformation is the pseudo-energy of a generated conformation. It is calculated as

$$d = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{r}_i^t)^2} \tag{8}$$

in which $\mathbf{r}_i$ is the coordinate of the ith atom in the generated conformation and the $\mathbf{r}_i^t$ is that of the same atom in the target conformation. The sum over the index i is over all main chain atoms of interested residues. The Metropolis criterion[54] is utilized to accept or reject conformations. Those conformations whose RMSD to the target conformation are smaller than those of their immediately proceeding accepted conformations are always accepted. The other conformations are accepted with a probability of $\exp[-\Delta d/T^*]$, in which Δd is the change of the RMSD to the target conformation since the last accepted conformation and T^* is a pseudo-temperature. The pseudo-temperature is chosen to be proportional to the current RMSD to the target conformation by

$$T^* = d\tau \tag{9}$$

where τ is a constant. In this way the pseudo-temperature is high when the RMSD to the target conformation is large so that it is possible for the protein to overcome some pseudo-energy barriers. The pseudo-temperature is low when the RMSD to the target conformation is small so that only those conformations which are closer to the target conformation than previously accepted conformations are accepted. The generated conformations are attracted toward the target conformation as the RMSD decreases.

3 Results and Discussion

3.1 Conformations of HIV-1 Protease

3.1.1 Structures and Function of HIV-1 Protease

HIV-1 protease is vital for the reproduction of the HIV virus. Its structure and dynamics have been extensively studied. Since the first 3D structure of the HIV-1 protease was published,[55] more than 200 structures of the protease have been reported[56] and are compiled in the Protein Databank (PDB)[57] and HIV protease database.[58] This protease is of enormous value and continuing medical interest as a target for anti-AIDS therapy, since blocking the active site of the protease prevents cleavage of the HIV polypeptide into capsid and other proteins, and thus blocks the formation of the active virus. As a consequence, the HIV-1 protease has been studied with great interest. York et al.[59] simulated the protease motion in crystalline environment and in solution. They claimed that the protease was held in the extended form in crystal due to the packing effects. Zoete et al.[60] studied the sequence variance of the HIV-1 protease. They also simulated the motion of the protease when it is bound with several ligands. Jacobs et al.[36] applied the flexibility approach FIRST to the liganded and ligand-free structures of the protease. Stultz and Karplus[61] docked a couple of ligands to the flexible regions in the protease. Rick et al.[62] calculated the free energy along a reaction coordinate between two conformations of the protease. The free energy for the protease to bind with a fullerene-based ligand is calculated by Zhu et al.[63] Bahar et al.[64] show that the temperature B factor of the crystal structure, which reflects the amplitudes of the atomic motions, can be accounted for rather accurately by the elastic network model (ENM).

The inhibitor-free structure of the HIV-1 protease (PDB ID 1HHP[65]) contains two identical amino acid chains, each with 99 residues. The two chains are interdigitated to form a beta sheet at the dimer interface. The molecule has an exact C_2 symmetry, as shown in Fig. (2). The catalytic site of the protease is at the bottom of a large cavity in the middle

of the protease. Two flexible flaps formed by beta hairpins can open and close over the top of the cavity. All structures of the HIV-1 protease observed in experiments are either bound to ligands in the closed conformation or are in a ligand-free, open conformation, similar to the one shown in Fig. (2). In such a conformation, the catalytic site is not covered by the two flaps at the top of the protease due to the large void that is immediately above the site. The active-site cleft is more open and can allow substrates and inhibitors to enter.

The motion of these flexible flaps is important to the biochemical function of HIV-1 protease. The distance between the tips of the flaps in the ligand-free crystal structures is only $\sim$2.7Å. NMR experiments show that the two flaps, formed by residues 45 to 56 in each monomer, have two types of motions on different time scales. One motion is relatively slow, on the order of μs-ms.[66] The other motion is a fast curling motion of the tips (residues 49 to 53), which is in the sub-ns time range.[67] The fast motion is also observed in the MD simulation by Scott and Schiffer.[68] The characteristic of the curling motion is that the ϕ and ψ angles of the GLY51 residue take on values that are disallowed for non-Gly residues. Freedberg et al.,[67] however, do not agree on the scale of the flexibility shown in the simulation. Another recent MD simulation[69] shows that the fast curling motion of the tips is absent when water molecules are first equilibrated. A full atom, explicitly solvated MD simulation[70] reveals details of motion in the slowest and the fastest modes.

We study two aspects of the sterically accessible conformations of the flexible flaps of the HIV-1 protease. First, we would like to know how large the distance between the tips of the flaps can be in all conformations. Second, we would like to characterize all conformations of the flaps and compare our results to MD simulations. We utilize the combination of flexibility analysis tool FIRST with ROCK to address these questions.

3.1.2 Flexibility Analysis on HIV-1 Protease

Polar hydrogen atoms are first added to the protein structure 1HHP by the Unix version of WhatIF.[71] Polar hydrogen atoms are critical to hydrogen bonds so they are added explicitly.

The non-polar hydrogen atoms are ignored because they do not form hydrogen bonds hence they are not components of rings, and hence have no effect in determining the rigid regions in the protein. According to a study by Jacobs et al.,[36] WhatIF gives a rather accurate placement of polar H atoms. The software package FIRST is then used to analyze the flexibility properties of HIV-1 protease. FIRST assumes a hydrophobic interaction whenever two carbon atoms are within a certain distance. For HIV-1 protease, FIRST identifies six hydrophobic interactions between the two flexible flaps. From the open conformation of the flaps, we know that these hydrophobic interactions do not persist, and they would severely restrict the conformational space sampled by ROCK. Also, the hydrophobic interactions between the two flexible flaps are not as stable as those in the interior of proteins. For this reason, we excluded them from the network. Flexibility analysis of the resulting bond network shows that a majority of HIV-1 protease is rigid with small flexible regions, as shown in Fig. (2). The two flaps on the top, residues 45–56, are flexible, as required for the biochemical function. Previous analysis[36] has showed that flexibility predicted in the beta turn regions on the side regions, residues 35–41, provides a plausible explanation for drug-resistant mutations found in this region. Mutations here could change the side-region flexibility, which would couple to the active site via the flexible flaps.[36] Not shown in Fig. (2) are many other smaller flexible regions which are exclusively formed by flexible side chains.

Most of the protein is rigid, with the exception of several loops that are flexible. The backbone of the residues 15–18 forms one small flexible loop. A hydrogen bond between GLN18 side chain NE2 atom and the GLY16 main chain O atom separates the region into two twelve-fold rings. Because of the correlations between the rings, the DOF in this region is only 3, when the DOF of dangling side branches are not counted. Residues 35–41 form another flexible loop. All the main chain dihedral angles in this loop are freely rotatable, except for those of the PRO39, which is at the tip of the loop. There are 12 DOF in this loop, excluding the DOF of the dangling side branches. The largest flexible region is the β-strands of residues 45–56. Backbone hydrogen bonds between LYS45 and VAL56, between

ILE47 and ILE54, and between GLY49 and GLY52 separate the flexible region into several inter-connected rings. The DOF of the coupled rings in this region is only 6.

3.1.3 Conformations of HIV-1 Protease

We have used ROCK to generate 600 conformations of the protein HIV-1 protease obeying all the constraints specified in the flexibility analysis. Because a large portion of the protease is rigid, the calculation power of ROCK is concentrated on the two top flexible flaps and on the other smaller flexible regions shown in Fig. (2). The total CPU time was about 7 hours on an AMD Athlon 1900+ processor. All of the generated conformations are free of van der Waals collisions. Main chain ϕ and ψ angles are restricted to the core and the allowed regions in the Ramachandran plot developed by Morris et al.[45] The superimposition of these 600 conformations is shown in Fig. (3).

The distance between the tips of the flexible flaps in all conformations indicates that the sterically accessible conformational space is large. Defined to be the shortest distance between any atom in one flap and any other atom in the other flap in a given conformation, this distance can be as large as 8.0Å, as shown in Fig. (4).

The RMSD of main-chain C_α atoms across superimposed conformations indicates the average fluctuation of these atoms from their coordinates in the initial crystal structure. Its mathematical form u is calculated by

$$u^{(j)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i^{(j)} - \mathbf{r}_c^{(j)})^2} \tag{10}$$

for the RMSD of the C_α atom of the jth residue. The vector $\mathbf{r}_i^{(j)}$ is the coordinate of the C_α atom of the jth residue in the ith conformation. The vector $\mathbf{r}_c^{(j)}$ is the coordinate of the C_α atom of the jth residue in the initial crystal structure. The sum is over all N conformations.

The solid line and the dash line in Fig. (5) show the calculated RMSD of the C_α atoms across the 600 conformations. The two lines represent data in the two identical chains in HIV-

1 protease. The protein undergoes large conformational changes in three regions: residues 15–18, 35–41 and 45–56. Other regions are kept rigid in the simulation, and have RMSD values of 0. Prominent motion is indicated for residues 45–56, which are the flexible flaps in HIV-1 protease. It is worth noting that the tip of the residues 35–41, residue PRO39, does not sample multiple intra-residue conformations because its main chain ϕ and ψ angles are locked by a five-fold ring. The residue PRO39 is shown as the blue tip in the residues 35–41 in Fig. (2). However, because it is dragged into inter-residue motion by the other residues in the 35–41 range, it does not a non-zero RMSD value.

The temperature factor B obtained in X-ray crystallography reflects how much an atom moves in the crystal. It is related to the RMSD u via the following equation:[72]

$$B = 8\pi^2 \langle u^2 \rangle \tag{11}$$

where $\langle u^2 \rangle$ is the mean square deviation from the average atomic position. The RMSD of the C_α atoms in the X-ray structure 1HHP is converted from the experimentally determined B factor first, then plotted as the dotted line in Fig. (5). The RMSD of the C_α atoms in the crystal structure does not have many interesting features. It oscillates around 0.7Å without any major peaks. The protease in the crystal structure does not sample many conformations because of crystal contacts and volume constraints. Our calculation shows that the protein has a large range of conformations that are not populated in the crystal.

Bahar et al.[64] studied the backbone vibrations of HIV-1 protease through the elastic network model (ENM). Fig. (1) of their paper shows a very good match between the B factor from X-ray crystallography with that calculated from the ENM. The ENM model focuses on the thermal fluctuations of a protein around its mean structure. Our algorithm, on the other hand, largely ignores such thermal fluctuations by freezing the rigid regions in space and by fixing the bond lengths and angles. Thus it is complimentary to the ENM model in that it focuses on the diffusive motion.

Scott et al.[68] published a similar figure (see their Fig. (2)) in the discussion of their MD simulation on the same protease. Because MD simulation catches both the slow and the fast motions in proteins, the RMSD calculated from the MD simulation has an oscillating background of 1.2Å. The highest RMSD value of about 4.8Å is observed at the 50th residue in one chain. This data is comparable to the RMSD of the 50th residue in our generated conformations. The RMSD of the C_α atoms in the three flexible regions identified by FIRST and ROCK are all distinguishable from the background in the MD simulation. However, MD simulation shows that the RMSD of the C_α atoms in residues 75–83 are higher than the background noise, and thus these residues have some flexibility that we don't pick up.

Kurt et al.[70] studied the 10 slowest motion modes observed in a full atom solvated MD simulation. As revealed in Fig. (3a) in their paper, the 10 slowest modes produce large scale conformational changes in four regions: around residue 15, residue 40, residue 50 and residue 80. The first three regions match well with the three flexible regions shown in Fig. (5), namely the residues 15–18, 35–41 and 45–56. The fourth region that shows conformational variations in their study, residues around 80, is not redeemed flexible in our analysis. Fig (5) of their paper, which shows the regions affected by the 10 slowest motion modes, is qualitatively in agreement with our flexibility analysis shown in Fig. (2). The regions of the highest stability in the MD simulation trajectory are also the rigid regions in our analysis.

3.2 Conformational Pathways of DHFR

3.2.1 Structures and Functions of DHFR

The enzymatic protein dihydrofolate reductase (DHFR) catalyzes the reduction of 7,8-dihydrofolate (DHF) or folate to 5,6,7,8-tetrahydrofolate (THF). THF plays a key role in the biosynthesis of purines and thymidylate, which are essential components of DNA. Therefore the activity of DHFR indirectly controls the biosynthesis of DNA. As a key enzyme, the protein DHFR is present in all living organisms. The bacterial DHFR has been one of the earliest targets for antimicrobial drug development and as such it has been extensively stud-

ied. Structures of DHFR bound with a number of ligands have been determined by X-ray crystallography and by NMR techniques. To date there are already more than 100 DHFR structures in the Protein Databank.[57]

Statistical analysis[73] shows that there are three *Escherichia coli* DHFR (ecDHFR) conformations observed in crystallographic structures: the *open*, *closed* and *occluded* conformations. Fig. (6a) shows the superimposition of the three conformations of ecDHFR. The closed, occluded and open conformations are represented by the protein structures 1RX1, 1RX6 and 1RA9 respectively,[74] which are well resolved with 2.0Å resolution. The three conformations are similar, differing slightly in the orientation of the upper domain, and differing significantly in the loop region of residues 14–24, which is conventionally called the M-20 loop. Fig. (6b) shows a close-up of the three conformations in the M-20 loop region, which substantially define the closed, occluded, and open conformation of ecDHFR.

The M-20 loop covers the binding site of the ecDHFR where ligands are bound. Sawaya et al.[73] conclude from the study of X-ray crystallography structures of ecDHFR that the closed conformation of ecDHFR is in the first half of the catalytic reaction cycle, while the occluded conformation is characteristic for the second half. Thus, there must be conformational pathway(s) between these two conformations. They also find a large number of open conformations in the X-ray crystallography structures. Based on their NMR study, Osborne et al.[75] show that the occluded and the open conformations are populated in solution, while the open conformation is not. Therefore the open conformation may be an artifact of the crystal packing.

The NMR experiment by Falzone et al.[76] indicates that the M-20 loop conformational change is the rate limiting step of the catalytic reaction. The rate of the substrates dissociating from ecDHFR is in the range of a couple of events to several hundreds of events per second.[77–79] The best MD simulations today can only reach a few microseconds, which is shorter than the millisecond to second range of the conformational changes of ecDHFR. Nevertheless, several MD simulations have been performed on ecDHFR.[80–82]

We would like to address a question related to the conformational pathways of ecDHFR. Since the closed and the occluded conformations of ecDHFR were shown to be physiologically relevant and observable states during catalysis, there must be conformational pathway(s) between these two conformations. Little is known as to whether these conformational pathways share common traits or not. The combination of FIRST and ROCK, which is described above for HIV protease, is used to answer this question. Since the open conformation may be an artifact of the crystal packing, we do not try to generate conformational pathways between the open conformation and the occluded or the closed conformations. However, we do monitor the similarities between the open conformation and the conformations on the trajectories between the closed and the occluded conformations.

3.2.2 Flexibility Analysis of DHFR

The protein structures 1RX1 and 1RX6[74] are used in this study to represent the closed and the occluded conformations of ecDHFR. The two conformations are identical in amino acid sequence, yielding identical covalent bond networks in the two conformations. Both conformations are bound with ligands. Ligands and surrounding water molecules are removed since we are interested in sampling intrinsically allowed conformations of ecDHFR itself, and since we know the ligand changes between these two states. Both protein structures are from X-ray crystallography with a resolution of 2.0Å. Polar hydrogen atoms are added to both conformations by the Unix version of the software WhatIF.[71]

Only those hydrogen bonds and hydrophobic interactions that are shared by the occluded and the closed conformations are expected to be stable in the conformational pathway between these two states. Therefore these are used as constraints in the flexibility analysis. Those hydrogen bonds or hydrophobic interactions that are present in only one of the two conformations are excluded, as they must be broken when the protein transforms between conformations. The software FIRST is applied to the protein network consisting of the covalent bonds and the shared hydrogen bonds and hydrophobic interactions between the closed

and the occluded conformations. Fig. (7) shows the flexibility properties of the protein network. The core of the protein is rigid and the M-20 loop, which is from ILE14 to LEU24, is flexible as expected. The residues from GLU118 to GLU129 are also flexible. This loop is conventionally called the F-G loop. The M-20 loop is glued with the F-G loop by hydrophobic interactions and hydrogen bonds. Together they form one large flexible region with 33 DOF in the complicated rings. The motion of the M-20 loop is coupled with that in the F-G loop through hydrophobic interactions and hydrogen bonds. The turn at ASP142–SER150 is also flexible. There are not direct interactions between this turn and the flexible region of the M-20 loop. However its motion affects the conformations of the M-20 loop through van der Waals repulsions. Residues SER63–VAL72, which are colored red at the top of the protein in the figure, are also flexible. This region can be decomposed into four rings. The total DOF in the rings is 8. Osborne et al.[75] observed from their NMR experiments that the residues 13–26 which are in the M-20 loop, residues 115–123, and residues 148–149 all have significant differences in chemical shift when DHFR transforms between the closed and occluded conformations. The FIRST flexibility analysis matches with their experimental observation.

3.2.3 Directed Pathways of DHFR

Six trajectories starting from the occluded conformation with a target of the closed conformation are generated by our algorithm. The conformations are randomly generated, but only those conformations that are closer to or slightly further away from the target conformation are accepted. These calculations generate thousands of intermediate conformations within several days of CPU time on a single AMD Athlon 1900+ processor. The parameter settings for all six pathways are the same except the initial random seeds. The random numbers used in the program ROCK are generated by a publicly available program.[83]

The main region of interest is the M-20 loop which consists of 11 residues. If we focus our attention on the main chain, with three atoms per residue, the conformational space is

defined in a $3 \times 3 \times 11 = 99$ dimensional space, which is too large for any easy geometrical analysis or visualization. To simplify, we define the two natural reference points in this higher dimensional space, which are the occluded and the closed conformations. All the conformations in the 99-dimensional space are then projected onto this simpler two dimensional plane, in which the RMSD of a conformation relative to the two reference states are its coordinates. Trajectories of conformations can be easily tracked, plotted and examined in this two dimensional plane.

Fig. (8) illustrates the correlations between the RMSD of generated conformations to the occluded and closed conformations. Since the calculation begins from the occluded conformation, the RMSD of conformational trajectories to the occluded conformation are exactly zero initially. The RMSD of trajectories to the closed conformation can reach slightly more than 4.0Å after the first few steps. Calculations are terminated when the RMSD to the closed conformation is less than 1.0Å. Because the bond lengths and angles in the initial conformation (the occluded conformation in this case) are not exactly the same as those in the ending conformation (the closed conformation), the calculation can never reach an RMSD of exactly zero, because the bond lengths and angles are not changed.

As a test, we built a conformation in which the bond lengths and angles were *identical* to those of the occluded conformation, and the dihedral angles were the same as those of the closed conformation. The RMSD of the best fit of this manually built conformation to the closed conformation is about 0.6Å. Therefore driving the RMSD down to the vicinity of 0.6Å seems to be the limit of the algorithm when the bond lengths and angles are not disturbed. The lowest RMSD observed in trajectories driving between the 1RX6 to 1RX6 calculations is between 0.8Å and 1.0Å.

As shown in Fig. (8), the six trajectories do not differ very much from each other in the two dimensional plane. Similarly, we analyzed the RMSD of the trajectories to the occluded and to the open conformations, and the RMSD to the closed and to the open conformations. The six trajectories are all similar in these analysis. However it is not conclusive that the six

trajectories are indeed similar to each other in the high dimensional conformational space, because totally different trajectories could look similar when they are projected from high dimensional space to low dimensional space.

The six trajectories do not show any triggering events such as a flip of a dihedral angle. The dihedral angles gradually and collaboratively change, resulting in quasi-continuous conformational transitions.

The RMSD values of the conformations in the six trajectories relative to the open conformation are all larger than 2.0Å. Our results indicate that it is not necessary for the ecDHFR to pass close to the open conformation on its path between the occluded and the closed conformations. This conclusion is consistent with the finding that the open conformation is not populated in solution,[75] and therefore is at least not a persistent state between the closed and occluded conformations.

3.3 RMSD in Coordinate Space vs. in Dihedral Angle Space

Trajectories between the initial and the target conformations are built by randomly sampling conformations and selecting those whose RMSD values to the target conformation are less than or slightly higher than the previous state's RMSD. These RMSD values are calculated in coordinate space. When the RMSD between two conformations is exactly zero in coordinate space, the dihedral angle RMSD (daRMSD) between them must also be zero. Therefore, it is intuitive to assume that when the RMSD is small in coordinate space the daRMSD will also be small in dihedral angle space. However, analysis on the trajectories proves this intuition is far from being correct.

The daRMSD θ of a generated conformation to the target conformation is defined as

$$\theta = \sqrt{\frac{1}{2N}\sum_{i=1}^{N}[(\phi_i - \phi_i^t)^2 + (\psi_i - \psi_i^t)^2]} \tag{12}$$

in which ϕ_i and ψ_i are main chain dihedral angles of a generated conformation and ϕ_i^t and

ψ_i^t are the corresponding dihedral angles in the target conformation. The sum is over all residues of interest, which are the 11 residues in the M-20 loop of ecDHFR in this case.

Fig. (9) shows the superimposition of a generated and the closed conformation. This conformation is chosen randomly from an ensemble of conformations that are generated along the conformational pathways and are low in RMSD to the closed conformation. The RMSD between these two conformations is merely 1.057Å in coordinate space. Though not perfectly matched, the two conformations wind around each other somewhat akin to the two chains of a double helix. But these two conformations are quite different in the main chain ϕ and ψ angle space. The daRMSD between the two conformations is 56.9°. The difference between some corresponding dihedral angles is more than 100°.

The lack of correlations between the RMSD in coordinate space and the daRMSD in dihedral angle space has been seen before. In their study of the variations of main chain ϕ and ψ angles in different conformations of proteins, Korn and Rose[84] found that a protein conformation is not deformed when the main chain dihedral angles are rotated compensatorily. The effect of a big change in the main chain dihedral angle in one residue may be offset by the rotations of main chain dihedral angles of adjacent residues. The counter-intuitive fact that the RMSD in coordinate space is not correlated with the daRMSD in dihedral angle space disqualifies the usage of the main chain dihedral angles in the analysis of the similarities between conformations. Whether two conformations are alike should be investigated in coordinate space.

4 Summary and Conclusion

Proteins evolve in such a way that parts of the proteins are flexible while other parts remain rigid.[37,85] The flexibility and rigidity properties of proteins can be determined by using FIRST flexibility analysis.[36] The algorithm ROCK, described in this paper, takes advantage of FIRST analysis as input to explore conformational space involving the flexible regions of

proteins only, while leaving the rigid regions unaltered. This reduces the calculation cost considerably and can efficiently explore large scale conformational changes.

Proteins are complicated 3D networks with many inter-locking rings formed by covalent and non-covalent interactions. The bonds shared by adjoining rings determine how the rings are positioned and oriented relative to each other. ROCK is an unique algorithm for conformational exploration in that it preserves all the intra-ring and inter-ring bond length and bond angle constraints exactly. The algorithm uses the ring closure equations of Gō and Scheraga[22] in combination with a fictitious potential. The fictitious potential corresponding to ring closure is minimized for the entire system of rings, simultaneously. ROCK first randomly selects and perturbs several rotatable bonds in a ring by small rotations. Then it minimizes the fictitious ring closure potential, which is the sum of the squares of the ring closure equations. ROCK can be applied to any complicated network containing many inter-locked rings. All the rings in the network are closed simultaneously when the sum of the fictitious ring closure potential is minimized to be practically zero. It is efficient in sampling conformations for proteins, which are composed of numerous inter-locking rings. The side branches of protein groups not involved in ring systems are anchored back onto the rings once a new conformation of the ring clusters has been found, with the requirements that bond lengths, bond angles and chirality are maintained. Main chain dihedral angles and van der Waals overlaps are checked at each step to ensure good stereo-chemistry within the generated conformations.

ROCK samples conformations consistent with the bond network used in the initial conformation by maintaining all bond length and coordination angle constraints, including non-covalent interactions. This proves to be a powerful approach for exploring cooperative motions within the structure. The flexibility analysis also inputs hydrophobic interactions when the distance between two hydrophobic atoms that are not bonded is within a certain limit. ROCK then does not accept conformations in which these hydrophobic interactions are violated.

One way to improve the algorithm is to allow the redistribution of hydrophobic interactions and hydrogen bonds. The breaking and forming of the non-covalent interactions are involved in some protein conformational changes. The current version of ROCK does not have a scheme to break existing non-covalent interactions. As a consequence, only those truly stable interactions should be included as real constraints in the FIRST/ROCK approach. Otherwise, the conformational space sampled by ROCK may be smaller than desired. One projected solution is to incorporate a bond breaking and forming mechanism in FIRST/ROCK. Existing hydrogen bonds and hydrophobic interactions could be removed from the list of constraints, while new hydrogen bonds and hydrophobic interactions are allowed to form when appropriate. At the present time, a bypass strategy is used. Only those non-covalent interactions that are present in several experimentally determined conformations of a same protein are taken as stable constraints. In our sampling of the conformational pathways for DHFR we include the common constraints observed in the occluded and the closed conformations. This step enables us to sample the necessary conformational changes between the occluded and closed DHFR conformations without the rearrangement of constraints during the process. Hydrogen bond lengths and angles, which are fixed in the current version of ROCK, are being considered to be relaxed to allow more motions. However it is not totally clear on how to count DOF exactly with varying bond lengths and angles. A simple way to allow bond length and angle variations in a hydrogen bond is to break the bond in FIRST/ROCK. If a hydrogen bond to be broken is in a marginally rigid region, the rigid regions shrink while the flexible regions expand. In some cases, the flexible regions dominate so that the whole protein is flexible. Rader et al.[37] have studied how protein flexibility properties are affected by the breaking of hydrogen bonds. It will be interesting to study how the protein conformational space is affected.

ROCK works in two modes. It can either sample protein conformations in an unbiased way, starting from an initial conformation, or can be used to sample directed conformational pathway(s) between two distinct protein conformations. In the latter mode, it uses

the RMSD between the current and randomly generated conformation and the target conformation as a pseudo-potential. The generated conformation is either accepted or rejected based on a Metropolis criterion.[54] This algorithm introduces little bias into the system while generating conformational pathways which can then be further optimized and studied.

The use of ROCK is demonstrated on two proteins: HIV-1 protease and ecDHFR. FIRST and ROCK did well at identifying the major flexible regions in the proteins and sampling their motions. The algorithm sampled a wide range of conformations of HIV-1 protease. The distance between the flaps can be as large as 8.0Å in some conformations, starting from a conformation in which they are 2.7Å apart.

ROCK was also used to generate six conformational trajectories between the occluded and the closed conformations of ecDHFR. As shown in recent NMR results, we find the open conformation is not a persistent state between the closed and occluded states, despite appearing so in some crystal structures, in which crystal contacts may contribute to the altered loop conformation.

In a recent study, Zavodszky et al.[39] compared the conformations of the protein cyclophilin A generated by ROCK with those detected in NMR experiments. They found overall good agreements. The conformations from ROCK has a slightly larger backbone variation.

Proteins exhibit a variety of flexible region distribution patterns. The two proteins we studied, HIV-1 protease and DHFR, share the feature that several smaller flexible regions are anchored on a large rigid core. Some proteins may have a couple of large rigid regions instead of one, for example proteins having hinge motions. In such cases, ROCK fixes the largest rigid region in space. The flexible hinge that connects the smaller and the larger rigid regions, together with the smaller rigid region, is treated as one expanded flexible region in ROCK. The conformation variance of the hinge region then alters the relative orientation between the larger and the smaller rigid regions. ROCK does not sample internal conformation variations in both the smaller and the larger rigid regions. In this way the hinge motion can

be simulated as well. Some other proteins have correlated motions between remote parts in proteins. At least one flexible region should span across the whole protein to link these distant sites, therefore the effect of conformational changes in one end is transmitted to the other end of the flexible network. It will be interesting to apply the FIRST/ROCK approach to study such phenomena.

MD simulations explore the protein conformational space by accurately following the rugged protein energy landscape. ROCK presents a different paradigm, in which the goal is to capture the significant stabilizing and mobility-influencing interactions in the protein, then sample the conformations accessible to this bond network. While simpler and potentially less accurate than detailed forcefield approaches, it is also considerably faster and can attain large scale motions that are at the far extreme of currently achievable MD motions. The goal is for ROCK to explore the regions of the protein conformational space that are most biologically relevant.

In an earlier paper,[41] the energy landscapes of a small molecule is explored by ROCK combined with an empirical potential. Though the molecule has only 36 atoms, its energy landscape is already complicated. ROCK first generates those conformations observing the bond length and angle constraints and without van der Waals overlaps. The conformations are then optimized by an empirical potential. This procedure produces more than 230 local energy minima. The protein energy landscape[86–93] is much more complicated than the simplicity and relative symmetry of this small molecule would suggest, and a simplified approach to characterizing the landscape through ROCK can be valuable in identifying the most significant features.

The goal of protein docking is to filter tens of thousands of potential ligands (drugs) for their ability to match the binding sites of proteins. In principle, both the flexibility of ligands and the protein targets should be taken into account in docking studies. In many studies, however, only the flexibility of ligands is considered. It is only recently that attention has been oriented towards modeling protein main chain flexibility in docking,[94] largely due to

the computational challenges entailed. Most of the work in this area still focuses on using a series of crystallographic conformations as alternative targets for docking. However, many conformations are accessible to proteins that have not been trapped cryptographically, and therefore ROCK can serve a unique role in sampling conformations from the equilibrium protein structure. In a parallel study, ROCK has proved an accurate tool for building the conformational ensemble of a ligand-free protein, cyclophilin A, yielding very good correspondence with NMR studies of this protein's conformational range.[39,95] This work was combined with SLIDE[96,97] for fully flexible docking of estrogen receptor and cyclophilin A ligands. Thus, ROCK also has strong potential for modeling the full range of conformations for protein-ligand recognition.

This software, as well as other programs that explore the flexibility of networks, is available to the academic community via flexweb.asu.edu. This includes the program ROCK described in this paper, and the program FIRST that is used as a front end to ROCK.

Acknowledgments

The authors acknowledge useful discussions with Mykyta Chubynsky, Robert Cukier, Roy Day, Brandon Hespenheide, Maria Kurnikova, A.J Rader, Claire Vieille and Walter Whiteley. This work was supported by NIH grant GM067249.

References

[1] Berry, M. B.; Meador, B.; Bilderback, T.; Liang, P.; Glaser, M.; Phillips, G. N. Proteins 1994, 19, 183.

[2] Müller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. Structure 1996, 4, 147.

[3] Schlauderer, G. J.; Schulz, G. E. Protein Sci 1996, 5, 434.

[4] Gerstein, M.; Schulz, G.; Chothia, C. J Mol Biol 1993, 229, 494.

[5] Crivici, A.; Ikura, M. Ann Rev Biophys Biomol Struct 1995, 24, 85.

[6] Houdusse, A.; Silver, M.; Cohen, C. Structure 1996, 4, 1475.

[7] Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K. K.; Huang, L.-W.; Crofts, A. R.; Berry, E. A.; Kim, S.-H. Nature 1998, 392, 677.

[8] Sudarsanam, S.; Dubose, R. F.; March, C. J.; Srinivasan, S. Protein Sci 1995, 4, 1412.

[9] Deane, C. M.; Blundell, T. L. Proteins 2000, 40, 135.

[10] Feuston, B. P.; Miller, M. D.; Culberson, J. C.; Nachbar, R. B.; Kearsley, S. K. J Chem Inf Comput Sci 2001, 41, 754.

[11] Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Ann Rev Biophys Biomol Struct 2001, 30, 211.

[12] Nakajima, N.; Nakamura, H.; Kidera, A. J Phys Chem B 1997, 101, 817.

[13] Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. Chem Phys Lett 1996, 259, 321.

[14] Kamal, K. B.; Sethna, J. P. Phys Rev E 1998, 57, 2553.

[15] Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 314, 141.

[16] Sugita, Y.; Okamoto, Y. Chem Phys Lett 2000, 329, 261.

[17] Berg, B. A.; Neuhaus, T. Phys Lett B 1991, 267, 249.

[18] Lee, J. Phys Rev Lett 1993, 71, 211.

[19] Higo, J.; Nakajima, N.; Shirai, H.; Kidera, A.; Nakamura, H. J Comput Chem 1997, 18, 2086.

[20] Hansmann, U. H. E.; Okamoto, Y. J Comput Chem 1993, 14, 1333.

[21] Swendsen, R. H.; Wang, J.-S. Phys Rev Lett 1986, 57, 2607.

[22] Gō, N.; Scheraga, H. A. Macromolecules 1970, 3, 178.

[23] Gō, N.; Scheraga, H. A. Macromolecules 1970, 3, 188.

[24] Gō, N.; Scheraga, H. A. Macromolecules 1973, 6, 273.

[25] Dygert, M.; Gō, N.; Scheraga, H. A. Macromolecules 1975, 8, 750.

[26] Gō, N.; Scheraga, H. A. Macromolecules 1973, 6, 525.

[27] Gō, N.; Scheraga, H. A. Macromolecules 1978, 11, 552.

[28] Wedemeyer, W. J.; Scheraga, H. A. J Comput Chem 1999, 20, 819.

[29] Wu, M. G.; Deem, M. W. Mol Phys 1999, 97, 559.

[30] Canutescu, A. A.; Dunbrack Jr., R. L. Protein Sci 2003, 12, 963.

[31] Bruccoleri, R. E.; Karplus, M. Macromolecules 1985, 18, 2767.

[32] Gibson, K. D.; Scheraga, H. A. J Comput Chem 1997, 18, 403.

[33] Jacobs, D. J.; Kuhn, L. A.; Thorpe, M. F. Rigidity Theory and Applications; Thorpe, M. F.; Duxbury, P. M. Eds.; Kluwer: New York, 1999.

[34] Jacobs, D. J. J Phys A 1998, 31, 6653.

[35] http://firstweb.asu.edu, website of the software FIRST.

[36] Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Proteins 2002, 44, 150.

[37] Rader, A. J.; Hespenheide, B. M.; Kuhn, L. A.; Thorpe, M. F. Proc Natl Acad Sci USA 2002, 99, 3540.

[38] Hespenheide, B. M.; Rader, A. J.; Thorpe, M. F.; Kuhn, L. A. J Mol Graph Model 2002, 21, 195.

[39] Zavodszky, M. I.; Lei, M.; Thorpe, M. F.; Day, A. R.; Kuhn, L. A. submitted to Proteins.

[40] Thorpe, M. F.; Lei, M.; Rader, A. J.; Jacobs, D. J.; Kuhn, L. A. J Mol Graph Model 2001, 19, 60.

[41] Thorpe, M. F.; Lei, M. to be published in Phil. Mag., 2004.

[42] Liu, D. C.; Nocedal, J. Math Program 1989, 45, 503.

[43] Spellucci, P. Math Program 1998, 82, 413.

[44] Ramachandran, G. N.; Sasiskharan, V. Adv Protein Chem 1968, 23, 283.

[45] Morris, A. L.; MacArthur, M. W.; Hutchioson, E. G.; Thornton, J. M. Proteins 1992, 12, 345.

[46] Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. Science 2001, 291, 2429.

[47] Weber, C.; Wider, G.; von Freyberg, B.; Traber, R.; Braun, W.; Widmer, H.; Wuethrich, K. Biochemistry 1991, 30, 6563.

[48] Concha, N. O.; Head, J. F.; Kaetzel, M. A.; Dedman, J. R.; Seaton, B. A. Science 1993, 261, 1321.

[49] Echols, N.; Milburn, D.; Gerstein, M. Nucleic Acids Res 2003, 31, 478.

[50] Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. J Mol Graph Model 2002, 21, 151.

[51] Kim, M. K.; Jernigan, R. L.; Chirikjian, G. S. Biophys J 2002, 83, 1620.

[52] Gould, H.; Tobochnik, J. An Introduction to Computer Simulation Methods Addison-Wesley 1998.

[53] Kirkpatrick, S.; Gelatt Jr., C. D.; Vecchi, M. P. Science 1983, 220, 671.

[54] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. J Chem Phys 1953, 21, 1087.

[55] Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Science 1989, 246, 1149.

[56] Kumar, M.; Hosur, M. V. Eur J Biochem 2003, 270, 1231.

[57] http://www.pdb.org, website of the Protein Databank.

[58] http://srdata.nist.gov/hivdb, HIV protease database.

[59] York, D. M. Y.; Darden, T. A.; Pedersen, L. G.; Andersont, M. W. Biochemistry 1993, 32, 1443.

[60] Zoete, V.; O., M.; Karplus, M. J Mol Biol 2002, 315, 21.

[61] Stultz, C. M.; Karplus, M. Proteins 1999, 37, 512.

[62] Rick, S. W.; Erickson, J. W.; Burt, S. K. Proteins 1998, 32, 7.

[63] Zhu, Z.; Schuster, D. I. Biochemistry 2003, 42, 1326.

[64] Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Phys Rev Lett 1998, 80, 2733.

[65] Spinelli, S.; Liu, Q. Z.; Alsari, P. M.; Hirel, P. H.; Poljak, R. J. Biochimie 1990, 73, 1391.

[66] Ishima, R.; Freedberg, D. I.; Wang, Y.-X.; Louis, J. M.; Torchia, D. A. Structure 1999, 7, 1047.

[67] Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y.-X.; Kustanovich, I.; Louis, J. M.; Torchia, D. A. Protein Sci 2002, 11, 221.

[68] Scott, W. R. P.; Schiffer, C. A. Structure 2000, 8, 1259.

[69] Carlson, H. A. personal communications.

[70] Kurt, N.; Scott, W. R.; Schiffer, C. A.; Haliloglu, T. Proteins 2003, 51, 409.

[71] Vriend, G. J Mol Graph Model 1990, 8, 52.

[72] Drenth, J. Principles of Protein X-ray Crystallography Springer-Verlag 1994.

[73] Sawaya, M. R.; Kraut, J. Biochemistry 1997, 36, 586.

[74] Reyes, V. M.; Sawaya, M. R.; Brown, K. A.; Kraut, J. Biochemistry 1995, 34, 2710.

[75] Osborne, M. J.; Venkitakrishnan, R.; Dyson, H. J.; Wright, P. E. Protein Sci 2003, 12, 2230.

[76] Falzone, C. J.; Wright, P. E.; Benkovic, S. J. Biochemistry 1994, 33, 439.

[77] Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. Biochemistry 1987, 26, 4085.

[78] Hammes, G. G. Biochemistry 2002, 41, 8221.

[79] Rajagopalan, P. T. R.; Zhang, Z.; McCourt, L.; Dwyer, M.; Benkovic, S. J.; Hammes, G. G. Proc Natl Acad Sci USA 2002, 99, 13481.

[80] Lau, E. Y.; Gerig, J. T. Biophys J 1997, 73, 1579.

[81] Radkiewicz, J. L.; Brooks III, C. L. J Am Chem Soc 2000, 122, 225.

[82] Rod, T. H.; Radkiewicz, J. L.; Brooks III, C. L. Proc Natl Acad Sci USA 2003, 100, 6980.

[83] L'Ecuyer, P.; Cote, S. ACM Trans Math Software 1991, 17, 98.

[84] Korn, A. P.; Rose, D. R. Protein Eng 1994, 7, 961.

[85] Luque, I.; Freire, E. Proteins 2000, 41, 63.

[86] Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. Proc Natl Acad Sci USA 2001, 98, 2370.

[87] Plaxco, K. W.; Simons, K. T.; Ruczinski, I.; Baker, D. Biochemistry 2000, 39, 11177.

[88] Kortemme, T.; Kelly, M. J. S.; Kay, L. E.; Forman-Kay, J.; Serrano, L. J Mol Biol 2000, 297, 1217.

[89] Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. Proc Natl Acad Sci USA 2000, 97, 2527.

[90] Eaton, W. A.; Muño, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. Ann Rev Biophys Biomol Struct 2000, 29, 327.

[91] Karplus, M. J Phys Chem B 2000, 104, 11.

[92] Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Ann Rev Phys Chem 1997, 48, 545.

[93] Brockwell, D. J.; Smith, D. A.; Radford, S. E. Curr Opin Struct Biol 2000, 10, 16.

[94] Carlson, H. A.; McCammon, J. A. Mol Pharmacol 2000, 57, 213.

[95] Zavodszky, M. I. Modeling Flexibility in Protein-Ligand Recognition Thesis, Michigan State University, 2003.

[96] Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. Proteins 1998, 33, 74.

[97] Schnecke, V.; Kuhn, L. A. Perspect Drug Discov Des 2000, 20, 171.

Figure Captions

Figure 1: The tip of one flap of HIV-1 protease. Four rings are formed by covalent and hydrogen bonds in this small region of the protein. Carbon, nitrogen, oxygen and hydrogen atoms are colored as green, blue, red and white spheres respectively. The solid red bonds are non-rotatable peptide bonds. The dashed red bonds are the hydrogen bonds. Non-polar hydrogen atoms are not shown for the sake of clarity.

Figure 2: The flexibility properties of HIV-1 protease (PDB 1HHP). Only the main chain is shown. The protease has one major rigid core as shown in gray (rigid and without redundant constraints) and blue (rigid and with redundant constraints). The two flaps on the top of the protein are flexible as indicated by the color yellow. Four additional regions in the protease are flexible. The protein is colored by the degree of flexibility—the scale is shown at the bottom of the figure.

Figure 3: Superposition of HIV-1 protease (PDB 1HHP) conformations generated by ROCK. Only the main chain of the protein is shown. The residues are colored by degree of flexibility. Fig. (a) shows a side view, while Fig. (b) shows the view from the top, with the flaps at the center.

Figure 4: Distance fluctuations between the flexible flaps in HIV-1 protease (PDB 1HHP), as a function of the numbered sequence of conformers. The distance ranges between 2.7Å and 8.0Å.

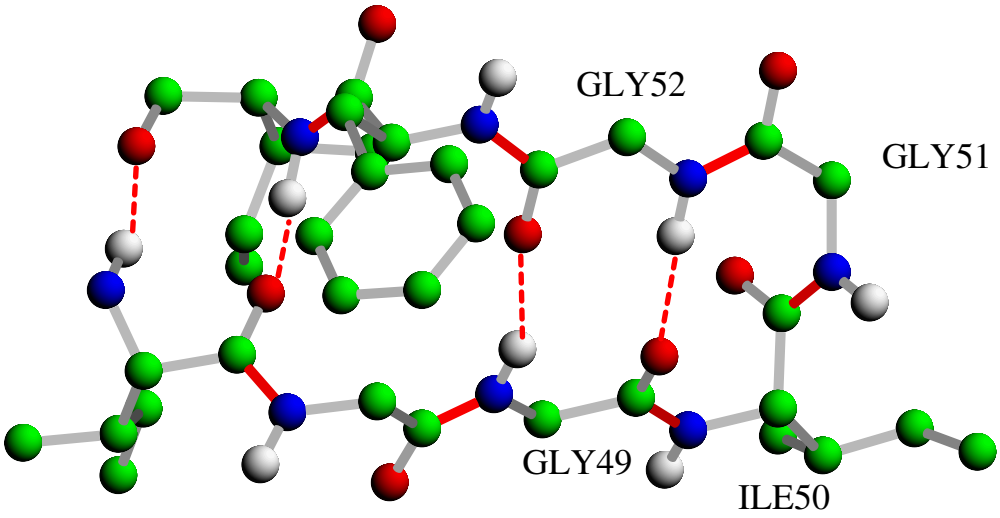
Figure 5: The average RMSD values for the main chain C_α atoms in the generated sequence of conformations relative to the initial crystal conformation. The solid and dashed curves are calculated for the two monomers in HIV-1 protease.

Figure 6: The superimposition of the open (blue), closed (yellow) and occluded conformations (red) of ecDHFR. These three conformations are represented by the PDB structures 1RX1, 1RX6 and 1RA9, respectively. Fig. (a) shows the whole protein while Fig. (b) shows a close up of the M-20 loop region from a different angle.

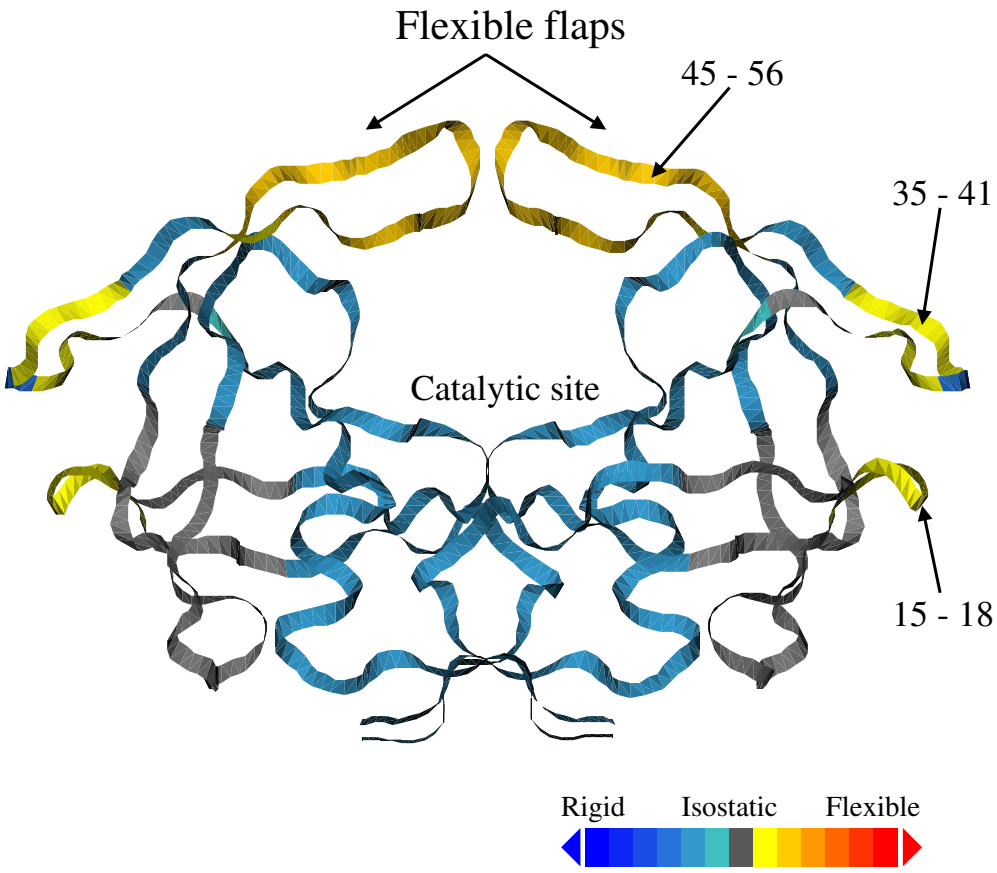
Figure 7: Flexibility properties of ecDHFR. The protein shown in this figure is 1RX6 which is in the occluded conformation. Only those hydrogen bonds and hydrophobic interactions that are present in both the occluded (1RX6) and the closed (1RX1) conformations are included in the flexibility analysis. Residues are colored by their degrees of flexibility. Only the mainchain is shown for clarity.

Figure 8: The correlation of RMSD of the six trajectories to the occluded and to the closed conformations, which run from the top left down to the right bottom. The RMSD of all six trajectories are exactly 0.0Å relative to the occluded conformation and roughly 4.0Å relative to the closed conformation in the beginning. The calculations are terminated for each trajectory when the RMSD relative to the closed conformations reaches 1.0Å.

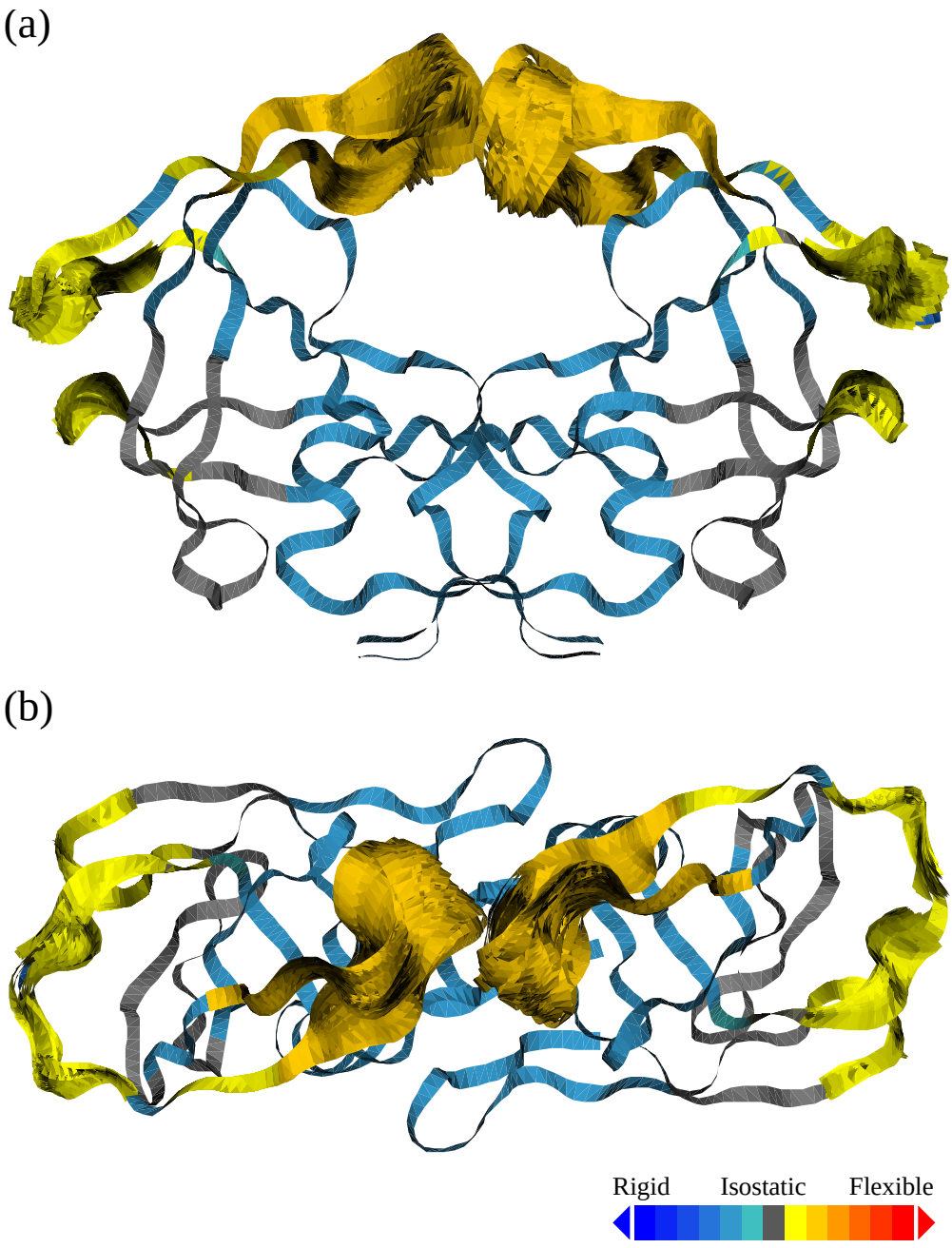
Figure 9: Shows a superimposition of the M-20 loop of a generated and of the closed conformations. One conformation is colored white while the other is colored black. The two conformations are almost identical.



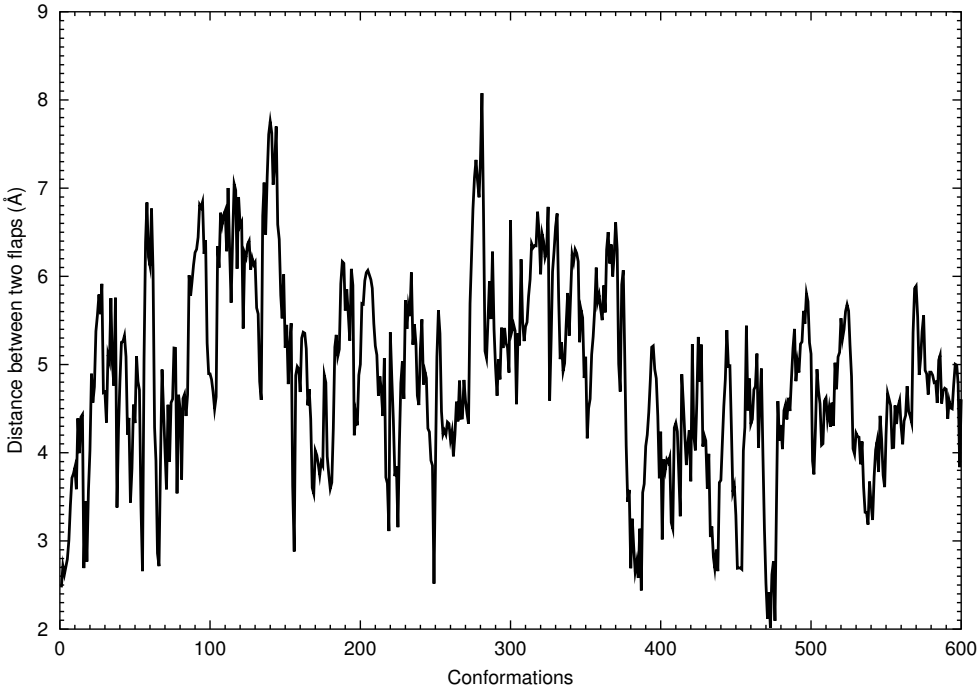
Lei, Fig. 1


Flexible flaps
45 - 56
35 - 41
Catalytic site
15 - 18
Rigid
Isostatic
Flexible

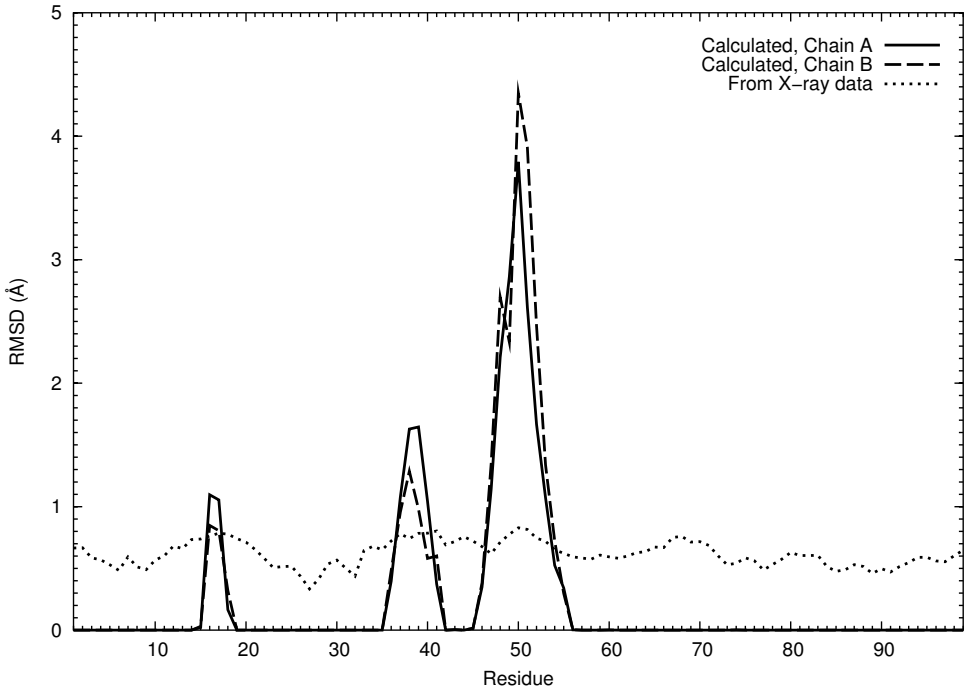
Lei, Fig. 2



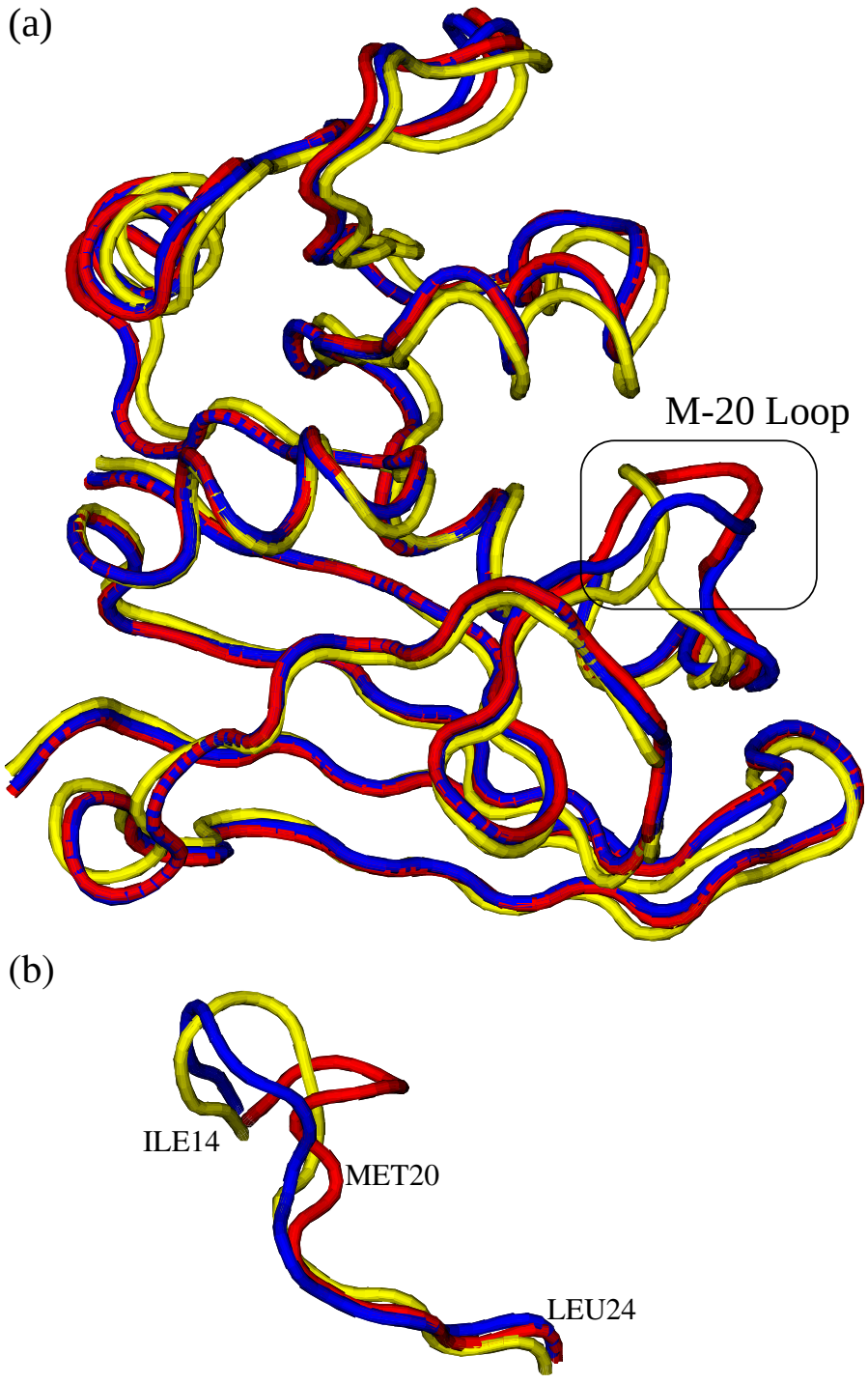
Lei, Fig. 3


Distance between two flaps (Å)
Conformations
9
8
7
6
5
4
3
2
0
100
200
300
400
500
600

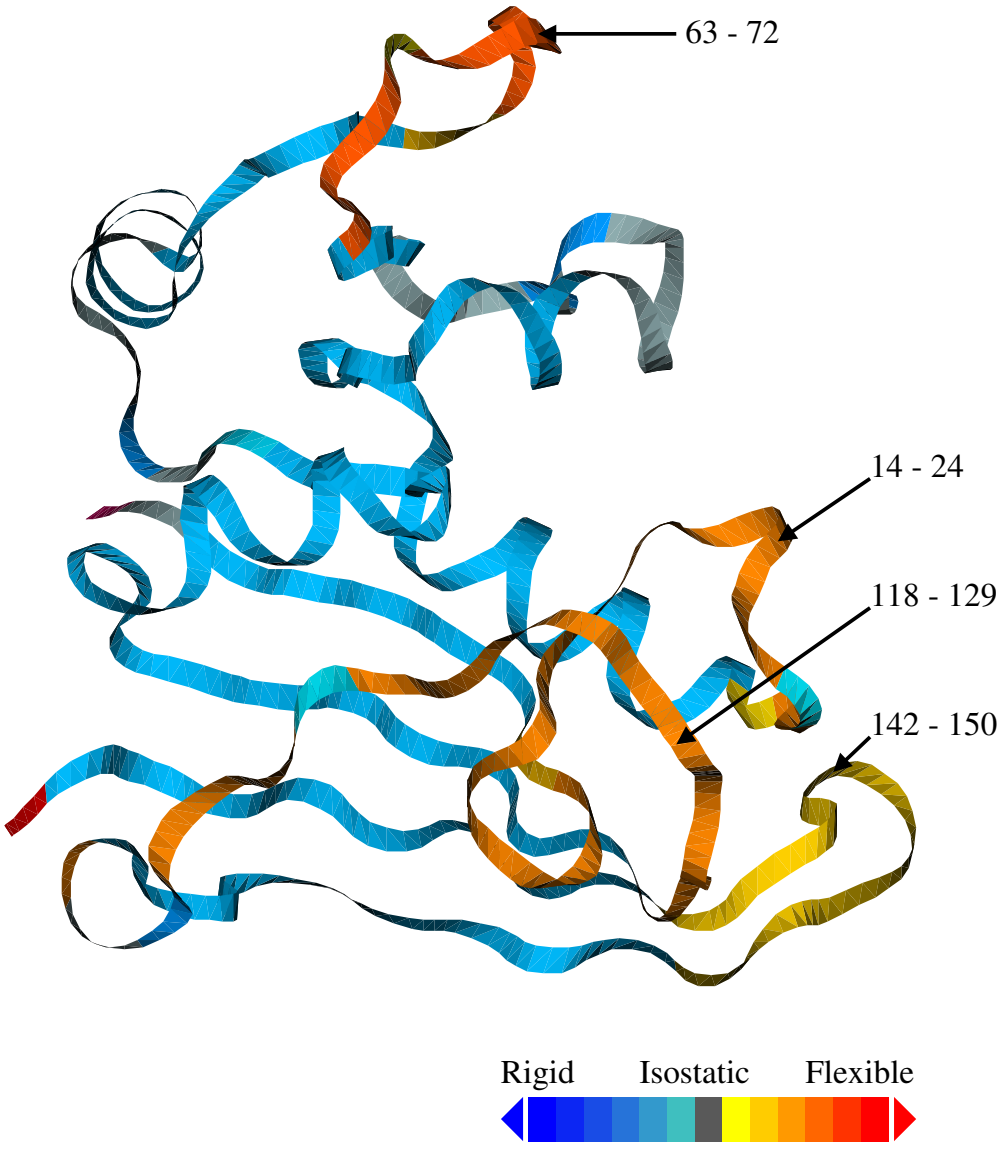
Lei, Fig. 4


Calculated, Chain A
Calculated, Chain B
From X−ray data
RMSD (Å)
Residue
0
1
2
3
4
5
10
20
30
40
50
60
70
80
90

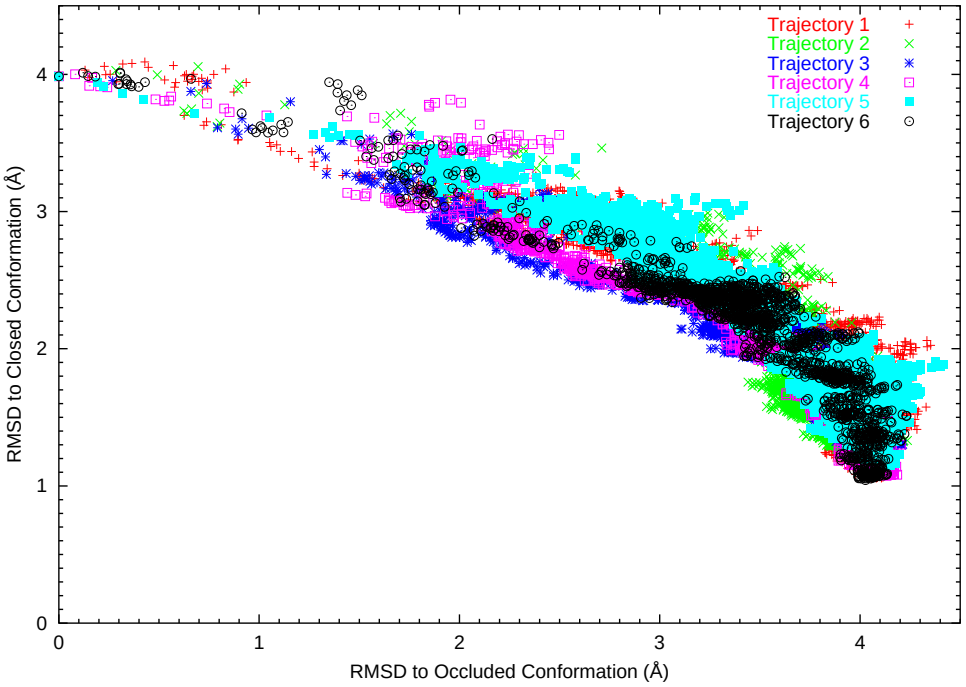
Lei, Fig. 5



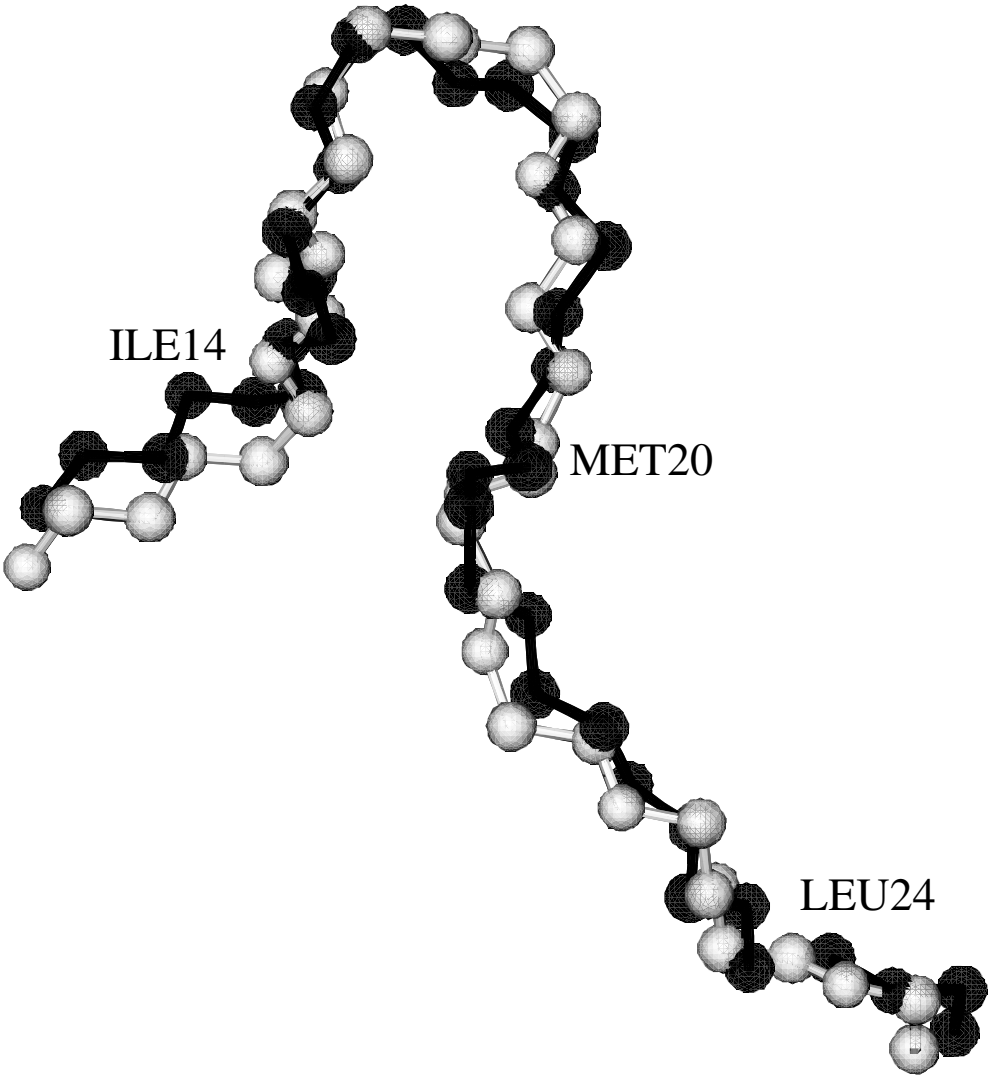
Lei, Fig. 6



Lei, Fig. 7


Trajectory 1
Trajectory 2
Trajectory 3
Trajectory 4
Trajectory 5
Trajectory 6
RMSD to Closed Conformation (Å)
RMSD to Occluded Conformation (Å)
0
1
2
3
4

Lei, Fig. 8



Lei, Fig. 9