IAK 2008

Conference on Interdisciplinary Applications of Kinematics
A. Kecskeméthy (eds.)
Lima, Peru, 9–11 January 2008

PREDICTING PROTEIN CONFORMATIONAL TRANSITIONS BY TRAJECTORY PLANNING THROUGH TORSION ANGLE PROPENSITY MAPS

Christopher Madden, Peter Bohnenkamp, Kazem Kazerounian* and Horea Ilies

Department of Mechanical Engineering
University of Connecticut
Storrs, CT 06269 USA
*Corresponding author. *E-mail address: kazem@engr.uconn.edu*

Keywords: protein, kinematic pathway, conformation, transition, energy landscape

Abstract. *The function of a protein macromolecule often requires conformational transitions between two native configurations. Understanding these transitions is essential to the understanding of how proteins function, as well as to the ability to design and manipulate protein-based nano-mechanical systems. It is widely accepted that the pathway connecting two native protein conformations in nature should satisfy a minimum energy criterion. The premise of this paper is that such a pathway can be found by using dihedral angle combinations that have been shown to have a high probability of occurrence in naturally observed proteins. In order to quantify this probability, we are proposing statistical propensity torque maps for tuples of dihedral angles. These maps are constructed in the angle space, similar to the Ramachandran Charts, but are based on data obtained from more than 38,600 proteins from Protein Data Bank (PDB) so that each map contains the experimentally observed pairs of dihedral angles* $(\varphi i, \psi i)$, (φ_i, ψ_{i+1}), $(\varphi_i, \varphi_{i+1})$ *and* (ψ_i, ψ_{i+1}).

1 INTRODUCTION

Proteins are nature's nano kinematic devices. At the macroscale, kinematic devices follow a constrained motion between two or more configurations in order to perform their functions. Similarly, the function of protein macromolecules often requires conformational transitions between two native configurations that are made possible by the intrinsic mobility of the protein. Understanding these transitions is essential to the understanding how proteins function, as well as to the ability to design and manipulate protein-based nano-mechanical systems [2].

Macroscopic kinematic devices have relatively few components and therefore a rather small number of variables, defining their spatial configurations. Broadly, these macroscopic devices move in a relatively uncluttered environment even when obstacles are present. On the other hand, the transitional pathway of a protein moving between two native configurations is significantly more complex due to the very large number of "design variables" and is therefore a computationally enormous task. Moreover, the stimuli external to the protein that cause these conformational changes are not completely understood. Nevertheless, the paramount importance of the conformational transitions in biological functions requires models based on first principles that can be both practical and valuable in understanding transition pathways of the proteins.

It is generally agreed upon that functional proteins have two or more native structures that are relatively close in terms of the corresponding potential energy values [3-6]. Furthermore, it takes a relatively small amount of energy to trigger the transition from one conformation to another. It is common to assume that the pathway between the native conformations is therefore a "valley" in the potential energy landscape. To understand such pathways, a detailed description of the kinematic motion of the molecule as well as the energy landscape corresponding to the kinematic structure is needed. Numerous methods have been developed to describe the conformational transition of the proteins along preferred energy pathways. A straight forward attempt is linear interpolation of the end point native conformations [7]. Such methods can also be improved by additional energy minimization [8]. While these methods are useful in visualization of the conformational transition, they do not necessarily represent the physical motion of the protein. More complex approaches include introduction of artificial potential forces in conjunction with Molecular Dynamics Simulation to force the protein motion from one conformation to another [9, 10]. However, the enormous computational requirement of the MDS severely limits the applicability of these methods. Furthermore, the accuracy of assuming large potential forces for guiding the conformation has not been quantified.

Traditional engineering based methods in studying the conformational pathways have also proven very useful. One such method is normal modal analysis of the protein structures [11-13]. While effective, due to the linear nature of

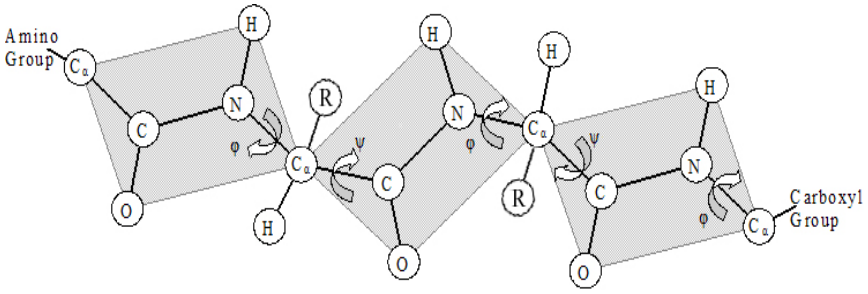


Figure 1: Amino Acid Chain with dihedral angles as generalized coordinates

modal analysis, such methods at best are local in time domain. By developing elastic models of the protein as a network of mass and springs, the efficacy of normal modal analysis is expanded globally throughout the conformational transition [14, 15]. One other class of approaches searching for the protein conformational pathways, which are deeply rooted in engineering applications, are based on robot motion planning algorithms [16, 17]. Such approaches have been remarkably successful in navigating the environment surrounding a robot, partly because that environment is assumed to be known. In the context of protein motion, the challenges faced by these approaches stem from the facts that the energy landscape is essentially unknown and it exists in a much higher dimensional space (compared to the typical robot motion planning problem).

Type of contact	Normal Limit (Å)	Extreme Limit (Å)
H to H	2	1.9
H to O	2.4	2.2
H to N	2.4	2.2
H to C	2.4	2.2
O to O	2.7	2.6
O to N	2.7	2.6
O to C	2.8	2.7
N to N	2.7	2.6
N to C	2.9	2.8
C to C	3	2.9

Table 1: Inter-atomic limiting distances [1].

In this work, we propose a new method for interactive planning of transitional pathways of a given protein. It is widely accepted that the pathway connecting two native protein conformations in nature should satisfy a minimum energy criterion. The premise of this paper is that such a pathway can be found by using dihedral angle combinations that have been shown to have a high probability of occurrence in naturally observed proteins. We propose the statistical propensity torque maps[1] for tuples of dihedral angles that capture the probability of occurrence of specific dihedral angle combinations in nature. This is so because our maps are constructed in the angle space, similar to the Ramachandran charts, but are based on data obtained from more than 38,600 proteins from Protein Data Bank (PDB) so that each map corresponds to pairs of dihedral angles ($\varphi i,\psi i$), (φ_i,ψ_{i+1}), (φ_i, φ_{i+1}) and (ψ_i,ψ_{i+1}). These maps are explained in more detail in Section 3.

2 PROTEN CHAIN MODEL

The authors have successfully modeled the protein molecule as a kinematic chain of rigid bodies connected by revolute joints [18-23]. In our novel approach, referred to as the Successive Kineto-Static Fold Compliance Method, the conformational changes of the peptide chain are driven by an inter-atomic force field without the need for Molecular Dynamic Simulation. Instead, the chain complies under the Kineto-Static effect of the force field in such a manner that each rotatable joint changes by an amount proportional to the effective torque on that joint. This process successively iterates until all of the joint torques have converged to zero. The resulting conformation is in a minimum potential energy state. This methodology has been used to develop the PROTOFOLD protein simulation software, and has been shown to be orders of magnitude more efficient and robust than traditional Molecular Dynamics Simulation [24].

3 DIHEDRAL ANGLE PROPENSITY MAPS

Protein backbone structure in the serial chain kinematic model is uniquely defined by the set of dihedral angles φ (the amino angle) and ψ (the carboxyl angle) for each amino acid

[1] In the robotics literature, a concept similar to our propensity maps is known as the so called dihedral work envelopes.

(residue). The set of these angles for all residues in the chain constitute the generalized coordinates of the backbone. In this work, we have modeled 20 of the 22 amino acids known to exist in the nature (the discovery of the last two has only been announced recently). In our model, the conformational transitions of the proteins are the result of the changes in the main chain dihedral angles and side rotamer angles.

3.1 Steric interaction and the dihedral angle-space propensity maps

The driving force behind protein folding is the interactions between the atoms that form the protein. As the joint angles change, the relative positions (distance) of atoms as well as the inter-atomic forces change as well. The inter-atomic forces, and therefore the potential energy of the protein, increase very quickly as the distance between two atoms approaches the diameter of the van der Waals spheres. The set of the corresponding dihedral angles can be computed for any such distance. On the other hand, the natural tendency of protein structures (and of any physical system) is to move towards states having lower potential energy. Therefore the angle combinations that create high potential energy states can be considered to be disallowed based on physical grounds. The corresponding configurations are referred to as "sterically disallowed" configurations. In an analogous robotic system, a set of joint angles defining a sterically disallowed configuration would produce joint angles that would correspond to self collision of the robot.

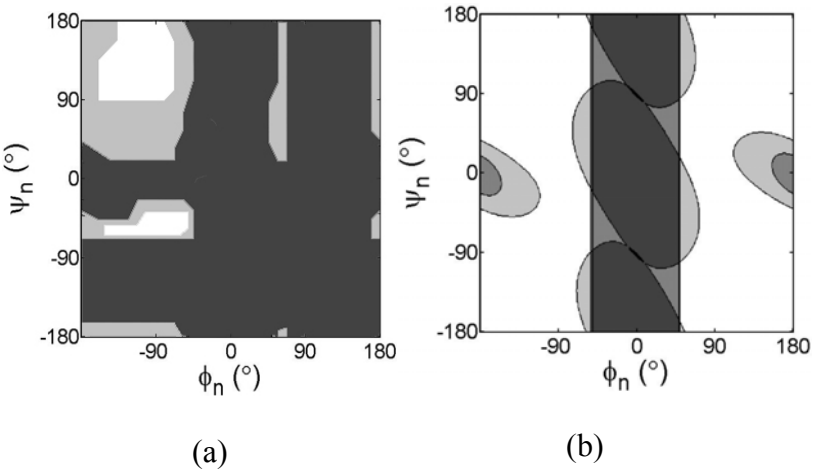


Figure 2: (a) Ramachandran Chart; (b) Glycine Self Collision Chart

In 1963, Ramachandran et al. [1] used both experimental and analytical methods to develop a set of inter-atomic minimum distances in amino acids. Their results are included in Table 1, where the left column corresponds to atoms in the two peptide planes connected by an α-C. Note that there are two limits listed in the table: a normal limit, and an extreme limit. At the atomic level there can be no interference between particles in the structure. There are no absolute self collisions, but rather a

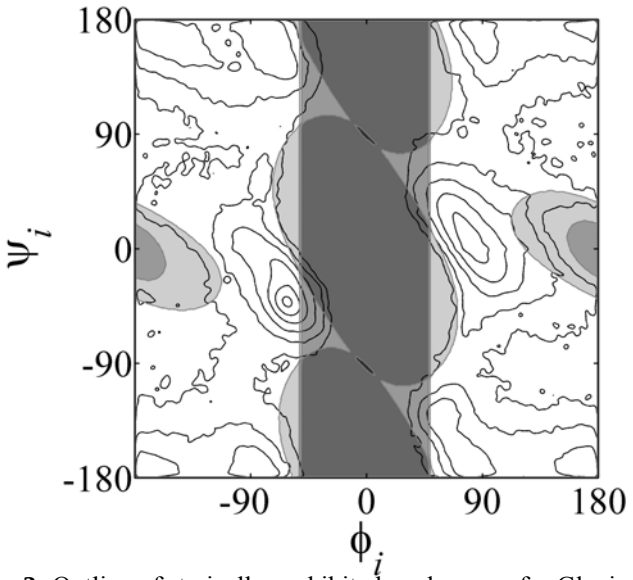


Figure 3- Outline of sterically prohibited angle space for Glycine overlaid onto dihedral angle population density from PDB.

steady increase in the repulsive force between the two interfering atoms. The conventional Ramachandran Map is a projection of the domain of the function (the torus) onto the plane and shows the fully allowed, partially allowed and disallowed regions for a dihedral angle tuple. These regions are constructed based on geometric derivations of the inter-atomic distances for various non-bonded atoms of the polypeptide, and by comparing these analytically obtained distances with those shown in Table 1.

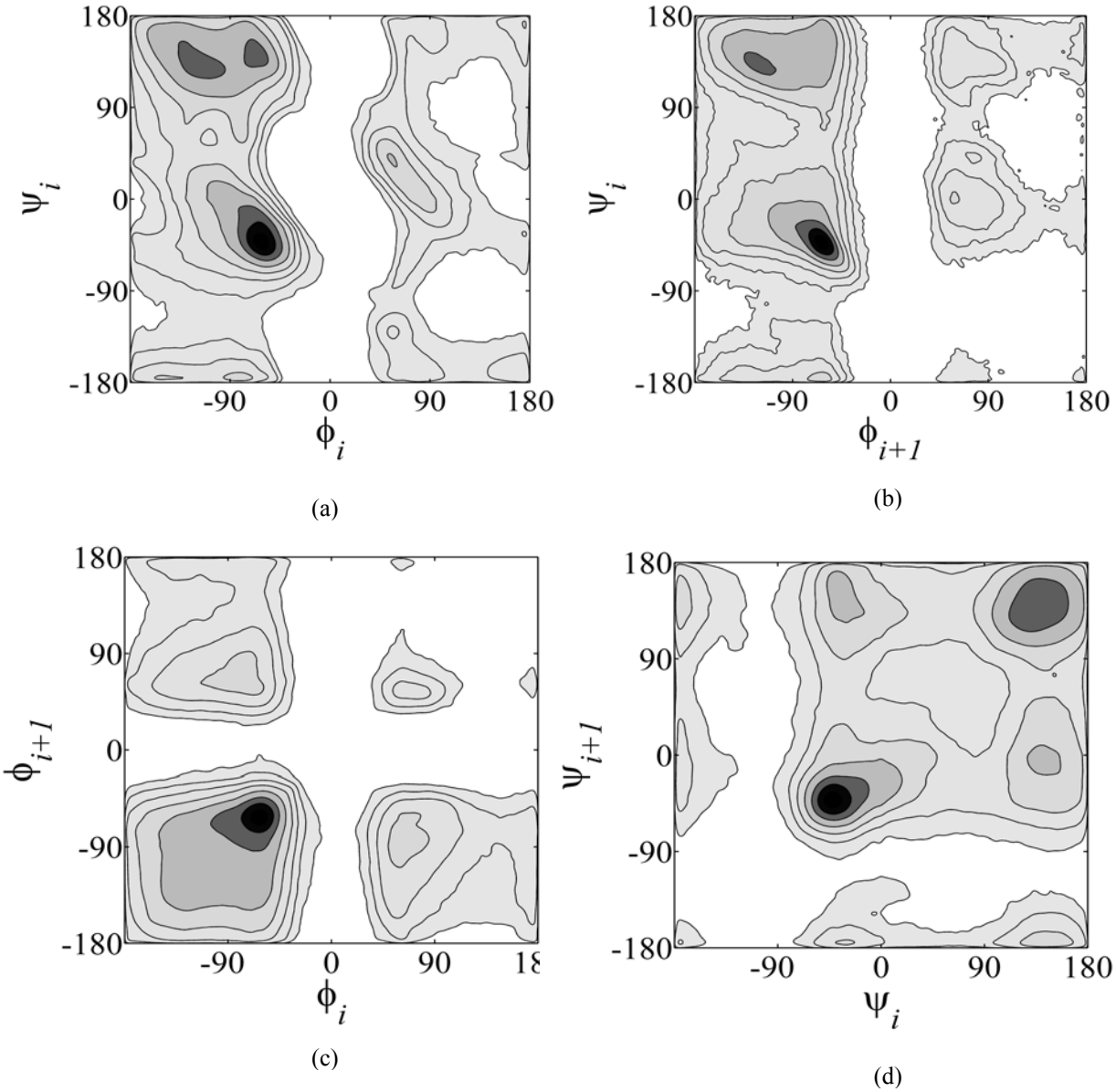


Figure 4(a-d): Four sample combination of dihedral angle work envelopes

The kinematic chain model of a polypeptide (for example as we implemented in PROTOFLD) can also be used to detect sterically disallowed angle combinations. The minimum distances shown in Table 1 can be coupled with a rigid sphere collision model set up in the kinematic chain model. The dihedral angles are varied between $0^\circ<\varphi<360^\circ$, and $0^\circ<\psi<360^\circ$, while monitoring the inter-atomic distances, which will result in the set of φ-ψ pairs that cause unstable structures. This set will be different for each amino acid. Ramachandran [1] performed

such an analysis for each amino acid, and combined the results in one chart that entails the ranges of φ-ψ values that are sterically disallowed. This chart is commonly referred to as the "Ramachandran" chart or map (Figure 2a), and has been traditionally shown over the ranges $-180^\circ<\varphi<180^\circ$, and $-180^\circ<\psi<180^\circ$ in the Cartesian plane. The darkest regions indicate dihedral angle combinations that are strictly sterically disallowed. Grey regions indicate sterically "strained" combinations, or, in other words, higher energy states that are still possible. White regions denote strictly sterically allowed angle combinations. Figure 2b is the map generated from a rigid sphere collision test implemented for only one amino acid structure, i.e., the Glycine. Note that many more angle combinations are considered sterically allowed for Glycine than for all amino acids, and that the Glycine map is a subset of the Ramachandran map.

In nature, dihedral angle combinations sometimes occur inside sterically disallowed regions. This happens when a disallowed angle combination results in an overall decrease in the potential energy of the structure. This phenomenon is relatively rare.

3.2 Generating Dihedral Angle Propensity Maps

The φ-ψ propensity maps can be generated using the rigid sphere model. Alternatively such maps *encoding the statistical distribution of angle combinations observed in nature* could be generated from the experimentally observed data found in the Protein Data Bank, which contains information on more than 40,000 proteins. While both techniques result in analogous map contours, PDB based maps contain details of the population propensity. This is illustrated in Figure 3 where we superimpose the maps obtained via the two methods mentioned above for a Glycine residue: the shaded areas are obtained from the collision model, and the contour curves (isocurves) have been computed based on PDB data.

Furthermore, observe that the difference between the charts in illustrated in Figures 2a and 2b clearly indicates that the typical Ramachandran chart (Figure 2a) is inadequate for computing propensity regions for all the dihedral angles in all amino acid sequences. Therefore, we have developed torque charts for the dihedral angle sets (φ_i,ψ_i) as well as (φ_i,ψ_{i+1}). $(\varphi_i, \varphi_{i+1})$ and (ψ_i,ψ_{i+1}). Since we modeled 20 (of the 22) amino acids known to exist in nature, 400 possible sequences exists for a pair of residues. Therefore the total number of maps in our master collection is 1220 (20+400+400+400) maps, which were developed based on 38,642 proteins from PDB. Figure 4 illustrates four such maps (one corresponding to each angle set). Darker regions indicate higher population density observed in nature and hence more favorable energy configurations.

4 DIHEDRAL ANGLE-SPACE PROPENSITY MAPS AS NAVIGATION GUIDES

In our numerical experiments we focused on the protein 1FOX, whose two known native conformations are shown in Figure 5. These two native conformations, known as 1FOX and 2FOW, have been computed based on data obtained from PDB.

Observe that PDB contains a set of dihedral angles for the backbone for each of the two native conformations. However, our dihedral angles are slightly different than those found in PDB due to both our serial kinematic chain model of connected rigid bodies as well as to the improved numerical model of the peptide planes [21], but the differences are negligible for the purpose of this work. There are 76 residues in this protein molecule, which result in 152 dihedral angles for the back bone. Consequently there are a total number of 213 torsion propensity maps.

A point in each of these maps corresponds to a specific set of values for the corresponding angles. Therefore, each of the two native conformations 1FOX and 2FOW are represented in *each* of our maps by one point in the torsion angle work space. Each curve connecting these two points (for a given map), represents a pathway for that particular set of dihedral angles. Then the pathway taking the protein from one native configuration to the other will be obtained by combining all these angle level pathways.

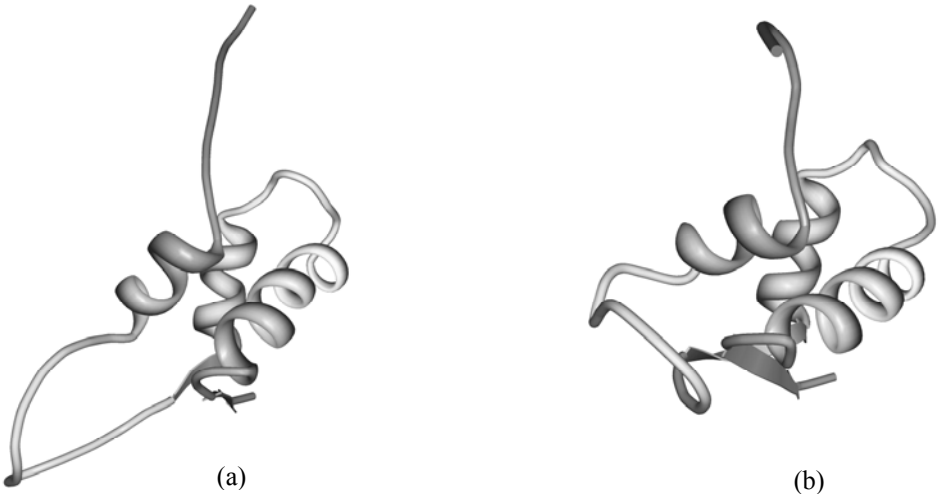


Figure 5: (a) 1FOX ribbon plot (b) 2FOW ribbon plot

4.1 Shortest distance linear pathways

Clearly, the simplest way to connect two points is by straight line, which is the basis of most of the existing visualization techniques for conformational transitions. Figure 6 shows four of these maps with the linear pathway in angle space for conformational transition. The larger circle at one end of the path in each chart indicates the initial conformation (i.e. 1FOX). The lines shown are the shortest distance between the toe end conformations. The darker regions in the chart indicate lower energy conformation and therefore naturally higher population density in PDB. Moreover, figures 6 (a) through (d) show the linear pathways with color-coded lines (coded on a gray scale) that correspond to the potential energy of the protein. Figure 7 shows the same map as Figure 6 (c) but with the color-coded pathway (RGB scale) as a function of the same potential energy of the molecule along the path. It can be clearly seen that the portions of the path on higher energy domains (lighter, or less populated regions) indicate higher total potential energy for the transitional conformation at that point on the pathway.

4.1 Selected direction linear pathways

The first step in adjusting the pathway to obtain a transition whose trajectory is more energy favorable was to select the direction of the linear paths in the angle space. To change each angle from an initial to a final value, two rotational directions can be selected (cw or ccw about the joint axis). Therefore four distinct linear paths are possible between the two end points in each chart. In our computer implementation each one of the four directions are

evaluated for the sum of the molecule's potential energy along the path. The output of this

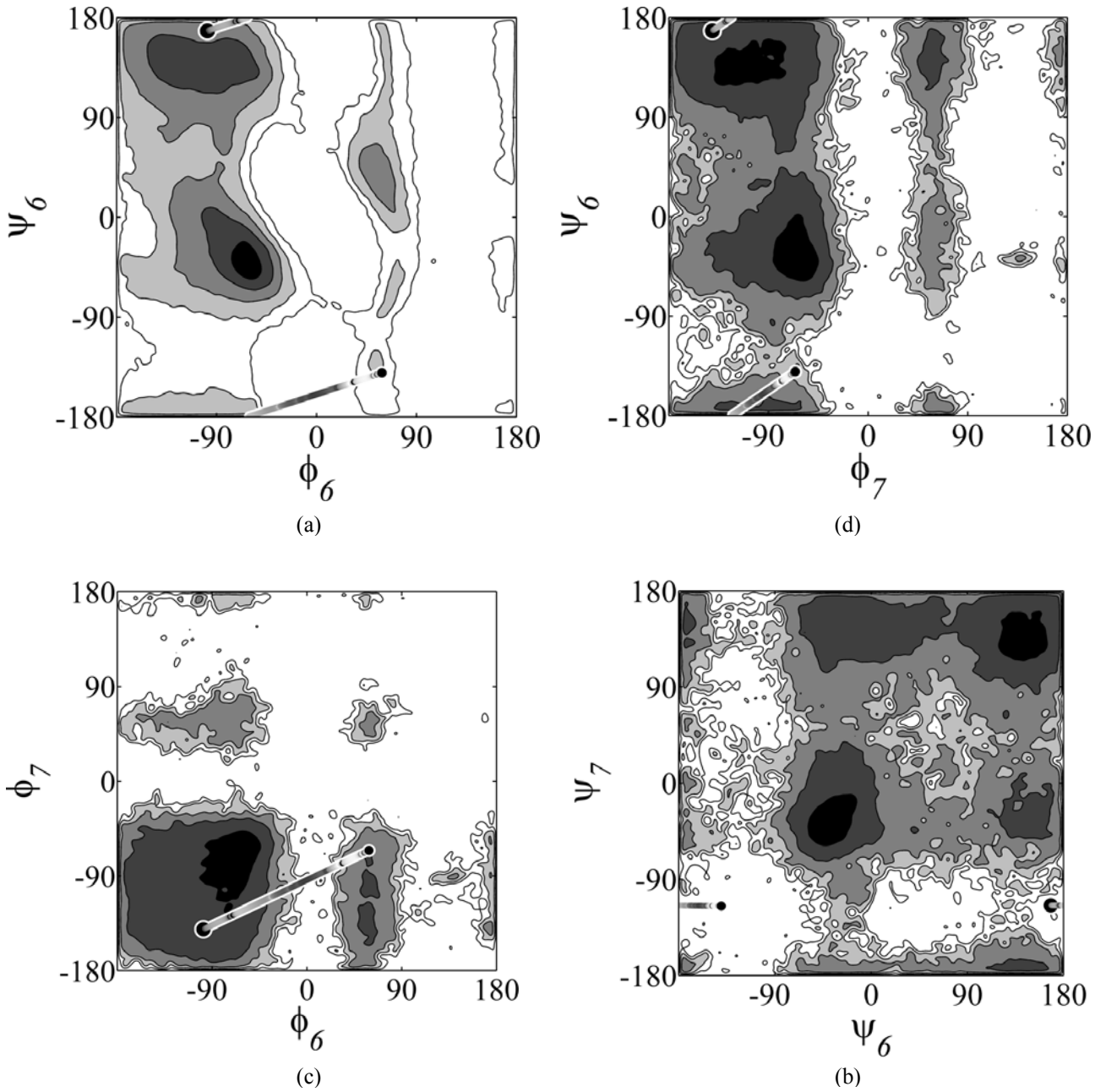


Figure 6- Conformational transition pathway of the dihedral angles along shortest distance linear paths in the third and forth residues (LYS-THR) in the protein molecule (1FOX to 2FOW)

evaluation is interactively used to select one of the four paths. Figure 8 shows sample selected paths.

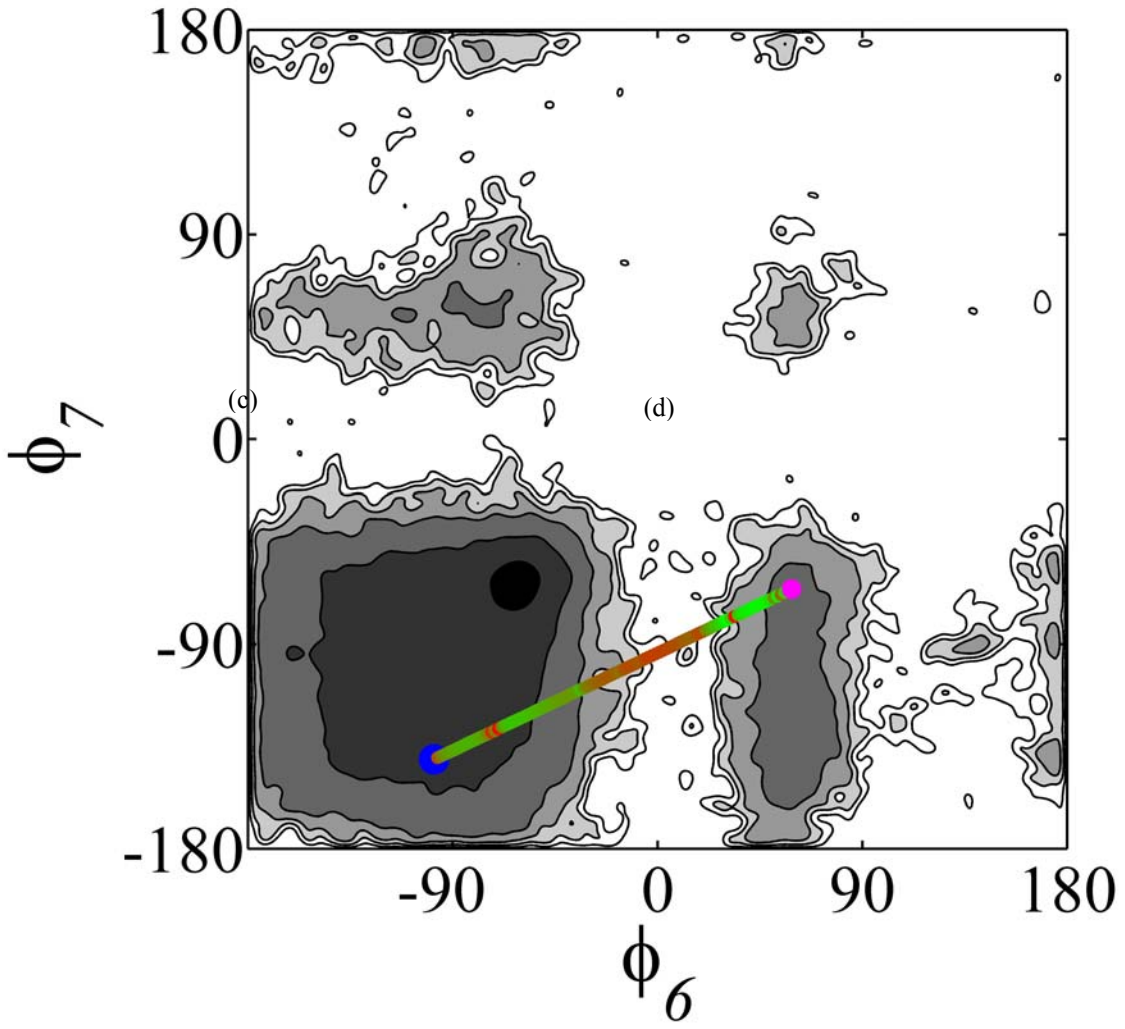


Figure 7: The chart of Figure 6 (c) with the path color coded to indicate total potential energy of the molecule throughout the conformational transition

4.2 Improved energy pathways

Through visual inspection of the charts (see sample charts in Fig. 8), the user can interactively force the paths to go through lower energy conformations as indicated by higher population intensity (darker regions). In our computer implementation, the trajectory modification is possible through introduction of additional trajectory points, or dragging existing trajectory points to new locations. Figure 9 shown sample charts where the selected linear pathways are modified visually by the user.

4.3 Minimum energy pathways

To fulfill the objective of obtaining a path of minimal energy between two low energy end conformations, the interactively designed pathways described above need to be further optimized. The kineto-static compliance method implemented in PROTOFOLD [18, 19] is applied to the curvilinear angle space pathways. PROTOFOLD can more finely tune the

pathways by rotating the joint angles that will relieve the highest remaining joint torques. **Error! Reference source not found.**10 illustrates how the linearly interpolated, planned and optimized paths in angle space relate. The contours shown are very simple representation of the complex energy landscape between the start and end conformations. If the end goal is to minimize the overall energy rise over the entire conformational motion, then the path should follow an isoenergy contour.

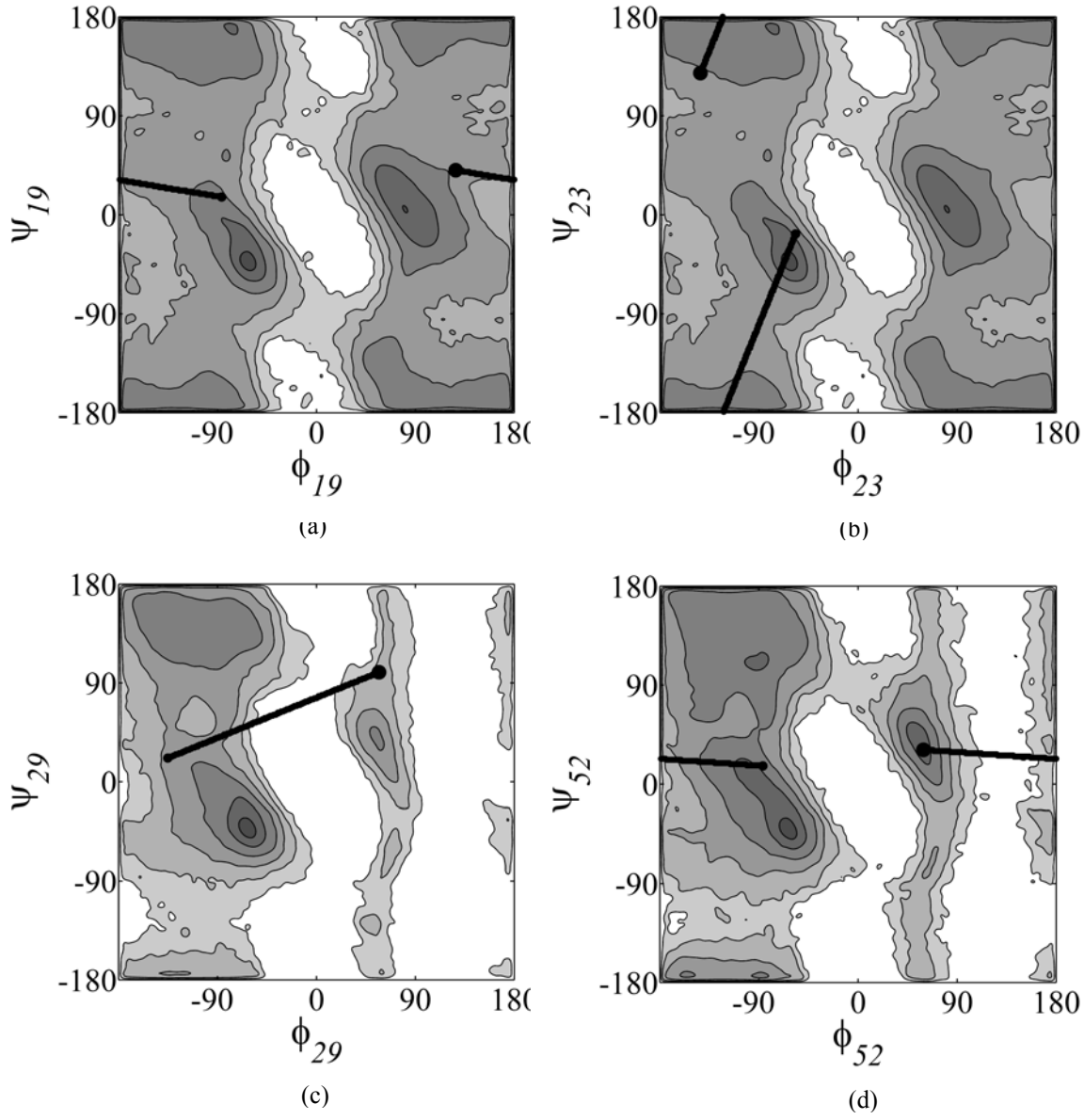


Figure 8- Sample conformational transition pathway of the dihedral angles along energy favorable linear paths in protein molecule (1FOX to 2FOW)

Given a pathway planned in angle space, energies are calculated for a set of intermediate conformations between the known end configurations using the AMBER force field applied to the kinematic model. Successive Kineto-Static Fold Compliance method in PROTOFOLD is used to move the kinematic model at each given intermediate conformation to one of lower

energy. This process is repeated for all intermediate conformations, under the restriction of minimal joint rotation to keep continuity between adjacent intermediates. If this were not in place, two adjacent conformations could diverge in angle space and the resulting pathway would be discontinuous. The energy minimized pathway very closely follows the planned input pathway in angle space, but is able to traverse a much lower energy profile.

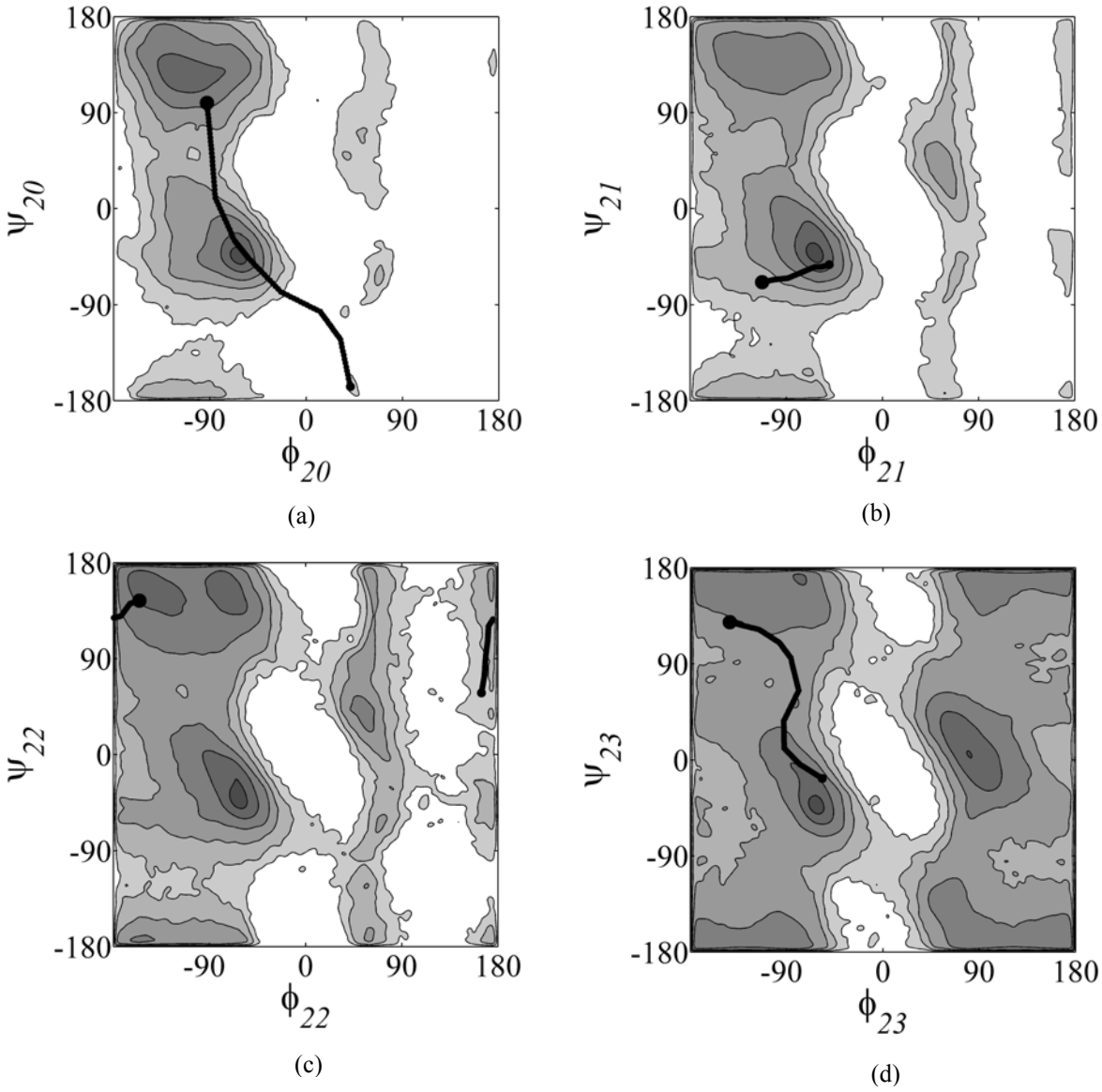


Figure 9- Sample conformational transition pathway of the dihedral angles modified interactively by the user to result in more energy favorable trajectories (1FOX to 2FOW)

This method is especially effective at relieving local areas of intense atomic interaction that cause high energy, but are not detected through angle space planning alone. Simply utilizing this optimization for sections of the angle space pathway that contains large jumps in energy is both computationally effective, since it does not have to run for all intermediate conforma-

tions, and effective at removing the largest energy barriers. Figure 11 shows how large energy barriers can be avoided with minimal changes in angle space of intermediate conformations.

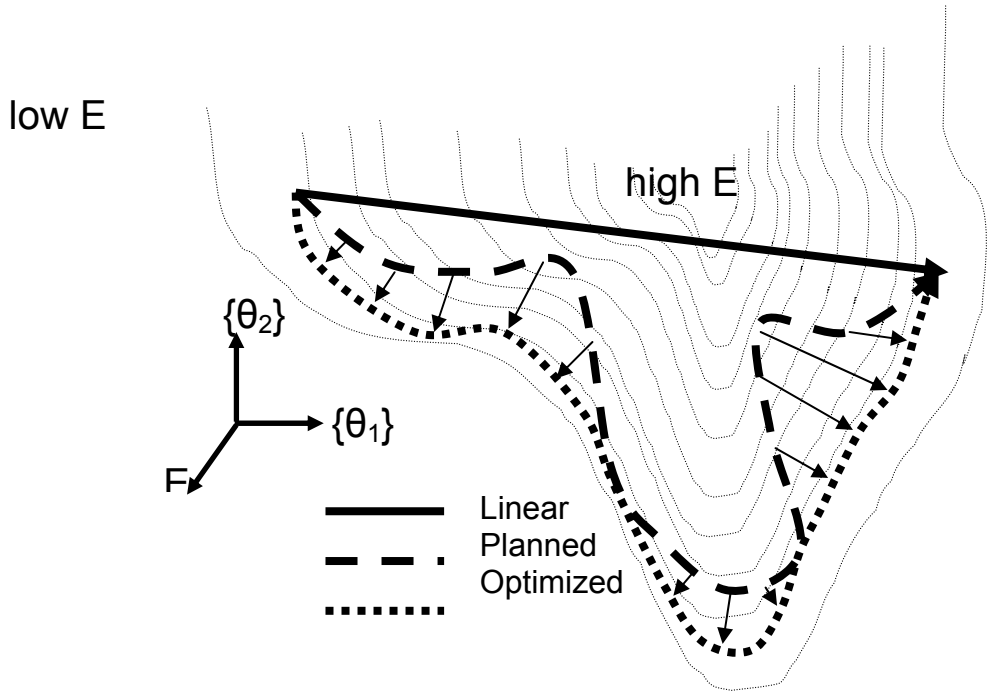
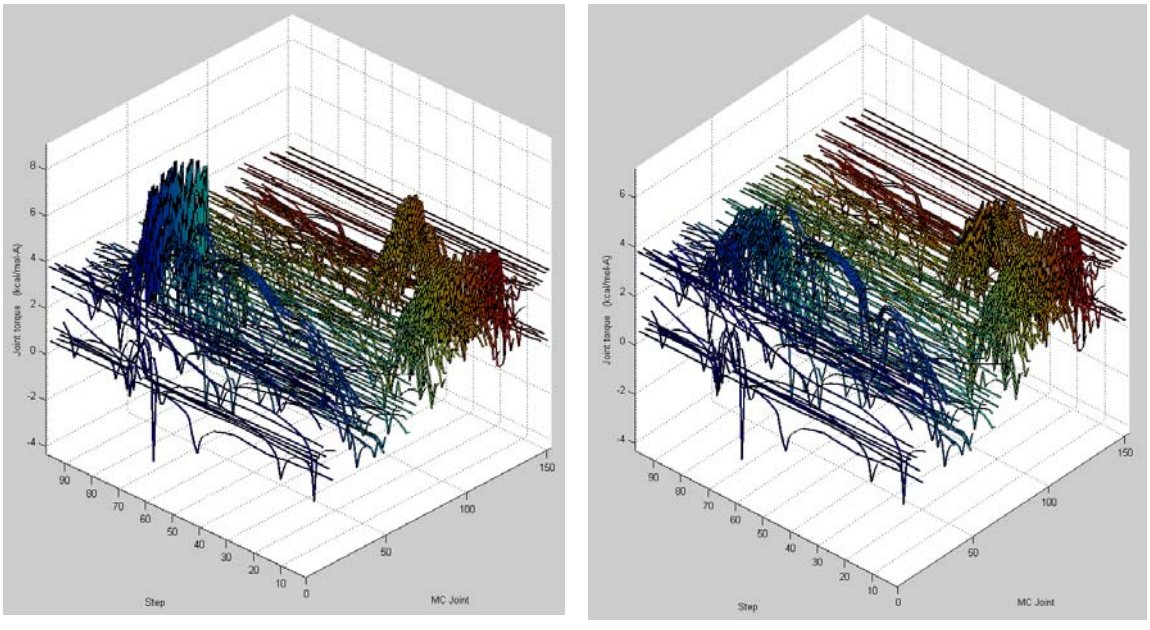


Figure 10: Three dimensional contour representation of energy landscape with various angle space paths.



(a) (b)

Figure 11- Plot of torques on every main chain joint during 1fox—2fow transition following (a) planned angle space pathway and (b) energy optimized angle space pathway. Note the local, large magnitude reduction in energy accomplished through the optimization.

4.4 Simulation results

Fig. 12 shows the energy profile of the various pathways that we have simulated in this work. The potential energy is directly related to the equivalent joint torques calculated in PROTOFOLD. In Figure 12, the Euclidean norm (summation of the square of all joint torques) is evaluated at discrete points along the pathway. As seen in Figure 12, although the best direction linear paths show some improvements over the shortest linear paths in some segments of the pathway, they are not very energy favorable in general.

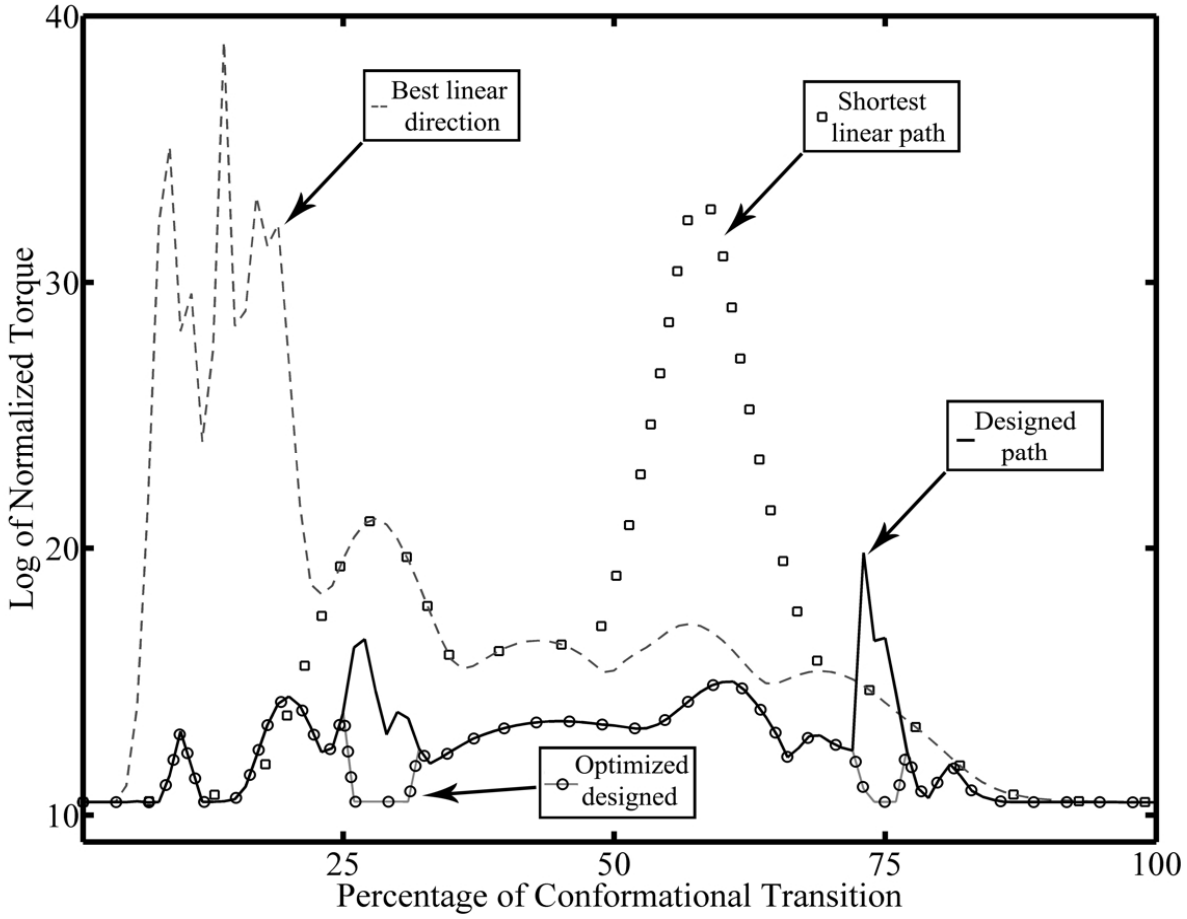


Figure 12- Simulated Pathways

From these plots it is evident that the designed path which is interactively designed by the user based on chart navigation leads to significantly more favorable energy profiles than the linear pathways in the torsion angle space. Furthermore, using these paths as the initial guess in the energy minimization performed by the PROTOFOLD further improves the energy profile of the pathways. Samples of the final energy optimum pathways are included in Fig. 13.

5- CONCLUSIONS

In this paper we have developed a systematic methodology for developing pathways for the conformational transition of the functional protein molecules. The pathways are developed based on torsion angle propensity maps obtained from the data available on more than 38,000 protein chains in the Protein Data Bank (PDB). Our numerical simulations indicate that these charts provide an very effective strategy for charting minimum energy profile pathways.

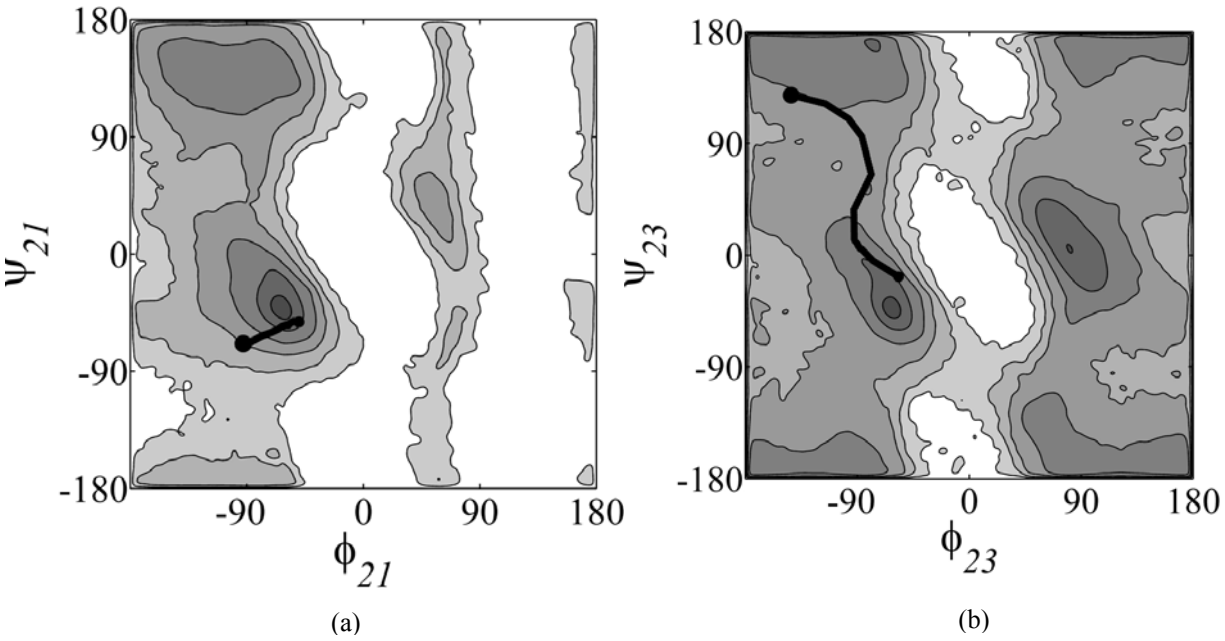


Figure 13- Sample conformational transition pathway of the dihedral angles along the energy optimized pathways obtained by PROTOFOLD (1FOX to 2FOW)

REFERENCES

[1] Ramachandran, G.N., C. Ramakrishnan, and V.J. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* Journal of Molec. Biollogy, 1963. **7**: p. 95−99

[2] Chirikjian, G.S., K. Kazerounian, and C. Mavroidis, *Analysis and design of protein based nanodevices: Challenges and opportunities in mechanical design.* Journal of Mechanical Design, 2005. **127**(4): p. 695-698.

[3] Miyashita, O., P.G. Wolynes, and J.N. Onuchic, *Simple energy landscape model for the kinetics of functional transitions in proteins.* Journal of Physical Chemistry B, 2005. **109**(5): p. 1959-1969.

[4] Gibbs, N., A.R. Clarke, and R.B. Sessions, *Ab initio protein structure prediction using physicochemical potentials and a simplified off-lattice model.* Proteins-Structure Function and Genetics, 2001. **43**(2): p. 186-202.

[5] Itoh, K. and M. Sasai, *Coupling of functioning and folding: photoactive yellow protein as an example system.* Chemical Physics, 2004. **307**(2-3): p. 121-127.

[6] Scheraga, H.A., et al., *The protein folding problem: Global optimization of force fields.* Frontiers in Bioscience, 2004. **9**: p. 3296-3323.

[7] Gerstein, M. and W. Krebs, *A database of macromolecular motions.* Nucleic Acids Research, 1998. **26**(18): p. 4280-4290.

[8] Krebs, W.G., et al., *Tools and databases to analyze protein flexibility; Approaches to mapping implied features onto sequences.* Macromolecular Crystallography, Pt D, 2003. **374**: p. 544-+.

[9] Schlitter, J., M. Engels, and P. Kruger, *Targeted Molecular-Dynamics - a New Approach for Searching Pathways of Conformational Transitions.* Journal of Molecular Graphics, 1994. **12**(2): p. 84-89.

[10] Guilbert, C., D. Perahia, and L. Mouawad, *A Method to Explore Transition Paths in Macromolecules - Applications to Hemoglobin and Phosphoglycerate Kinase.* Computer Physics Communications, 1995. **91**(1-3): p. 263-273.

[11] Guilbert, C., et al., *Low frequency motions in phosphoglycerate kinase. A normal mode analysis.* Chemical Physics, 1996. **204**(2-3): p. 327-336.

[12] Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations.* Protein Engineering, 2001. **14**(1): p. 1-6.

[13] Tama, F., W. Wriggers, and C.L. Brooks, *Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory.* Journal of Molecular Biology, 2002. **321**(2): p. 297-305.

[14] Kim, M.K., R.L. Jernigan, and G.S. Chirikjian, *An elastic network model of HK97 capsid maturation.* Journal of Structural Biology, 2003. **143**(2): p. 107-117.

[15] Schuyler, A.D. and G.S. Chirikjian, *Efficient determination of low-frequency normal modes of large protein structures by cluster-NMA.* Journal of Molecular Graphics & Modelling, 2005. **24**(1): p. 46-58.

[16] Amato, N.M. and G. Song, *Using motion planning to study protein folding pathways.* Journal of Computational Biology, 2002. **9**(2): p. 149-168.

[17] Tapia, L., et al., *Kinetics analysis methods for approximate folding landscapes.* Bioinformatics, 2007. **23**(13): p. I539-I548.

[18] Kazerounian, K., et al., *Nano-kinematics for analysis of protein molecules.* Journal of Mechanical Design, 2005. **127**(4): p. 699-711.

[19] Kazerounian, K., K. Latif, and C. Alvarado, *Protofold: A successive kinetostatic compliance method for protein conformation prediction.* Journal of Mechanical Design, 2005. **127**(4): p. 712-717.

[20] Bohnenkamp, P., K. Kazerounian, and H. Ilies. *Strategies to Avoid Energy Barriers in ab initio Protein Folding*. in *12th IFToMM (International Federation of the Theory of Mechasims and Machines) World Congress*. 2007. Besançon, France.

[21] Subramanian, R. and K. Kazerounian. *Improved Molecular Model of a Peptide Unit for Proteins* in *ASME 2006 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*. 2006. Philadelphia, Pennsylvania, USA.

[22] Subramanian, R. and K. Kazerounian, *Kinematic mobility analysis of peptide based nano-linkages.* Mechanism and Machine Theory, 2007. **42**(8): p. 903-918.

[23] Subramanian, R. and K. Kazerounian, *Residue level inverse kinematics of peptide chains in the presence of observation inaccuracies and bond length changes.* Journal of Mechanical Design, 2007. **129**(3): p. 312-319.

[24] Su, H., et al. *A Comparison of Kinetostatic and Multibody Dynamics Models for Simulating Protein Structures*. in *Mechanisms and Robotics Conference, ASME IDETC*. 2007. September 2007, Las Vegas.