

Simulating Protein Motions with Rigidity Analysis*

Shawna Thomas, Xinyu Tang, Lydia Tapia, and Nancy M. Amato
{sthomas,xinyut,ltapia,amato}@cs.tamu.edu

Technical Report TR05-008
Parasol Lab
Department of Computer Science
Texas A&M University
College Station, TX 77843-3112

September 29, 2005

Abstract

Protein motions, ranging from molecular flexibility to large-scale conformational change, play an essential role in many biochemical processes. Despite the explosion in our knowledge of structural and functional data, our understanding of protein movement is very limited because it is difficult to measure experimentally and computationally expensive to simulate. In previous work, we have developed and validated a method based on motion planning for mapping protein folding pathways from unstructured conformations to the native state. In this paper, we propose a novel method based on rigidity theory to sample conformation space more effectively and we describe extensions of our framework to automate the mapping process and to map transitions between specified conformations. Our experimental results show that these additions both improve the accuracy of our maps and enable us to study a broader range of motions for larger proteins. For example, we show that rigidity-based sampling results in maps that capture subtle folding differences between protein G and its mutations, NuG1 and NuG2, and we illustrate how our technique can be used to study large-scale conformational changes in calmodulin, a 148 residue signaling protein known to undergo conformational changes when it binds to Ca^{2+} . Finally, we announce our web-based protein folding server which includes a publically available archive of protein motions: <http://parasol.tamu.edu/foldingserver/>

*This research supported in part by NSF Grants ACI-9872126, EIA-9975018, EIA-0103742, EIA-9805823, ACR-0081510, ACR-0113971, CCR-0113974, EIA-9810937, EIA-0079874, and by the DOE. Thomas supported in part by an NSF Graduate Research Fellowship.

1 Introduction

Protein motions, ranging from molecular flexibility to large-scale conformational change, play an essential role in many biochemical processes. For example, local conformational change often occurs in binding interactions between proteins and between proteins and ligands, sugars, and other small molecules. While no consensus has been reached regarding models for protein binding, the importance of protein flexibility in the process is well established by the ample evidence that the same protein can exist in multiple conformational states and can bind to structurally different molecules [24].

Our understanding of molecular movement is still very limited and has not kept pace with the explosion of knowledge regarding protein structure and function. There are several reasons for this. First, the structural data in repositories like the Protein Data Bank (PDB) [10] consists of the spatial coordinates of each atom in the protein. Unfortunately, the experimental methods used to collect this data cannot operate at the time scales necessary to record large-scale protein motions. Second, traditional simulation methods such as molecular dynamics [46, 25] and Monte Carlo methods [17, 40] are computationally too expensive to simulate long enough time periods for anything other than small peptide fragments.

Nevertheless, there has been some recent attention focused on developing methods for modeling and studying protein flexibility and motion. One notable effort is the Database of Macromolecular Movements [20] developed by the Gerstein Lab. They generate and archive protein ‘morphs’ that interpolate between two different conformations of the same protein. As described in more detail in Section 2.2, while the method used is more chemically realistic than straight-line interpolation, it was selected over other more accurate methods for computational efficiency and is known to have problems for some kinds of large deformations [20, 41, 67].

In previous work [4, 3, 63, 62], we developed a new computational technique for studying protein folding that builds an approximate map of a protein’s potential energy landscape. This map contains thousands of feasible folding pathways to the known native state enabling the study of global properties of the folding landscape. We obtained promising results for several small proteins (60–100 amino acids) and validated our pathways by comparing secondary structure formation order with known experimental results [4].

Our Contribution. In this paper, we augment our framework with three powerful new concepts that

enable us to study a broader range of motions for larger proteins:

- We propose a new method based on rigidity theory to sample conformation space.
- We generalize our PRM framework to generate transitions between specified conformations.
- We present a new framework to automate the map building process.

Our new rigidity-based sampling allows us to study larger proteins by more efficiently characterizing the protein’s energy landscape with fewer, more realistic conformations. Rigidity theory provides a tool to identify the rigid and flexible regions of a conformation, which we exploit by focusing our iterative sampling process on the (currently) flexible regions. As we show, this results in smaller, better maps. In one dramatic case study, we show that rigidity-based sampling and analysis reveals the folding differences between protein G and mutations of G, NuG1 and NuG2 [55] – this behavior was not captured using our previous sampling technique [63].

Extending our framework to focus on particular conformations enables us to investigate questions related to the transition between particular conformations, e.g., when studying folding intermediates, allostery, or misfolding. We provide evidence that the transitions mapped by our approach are more realistic than those provided by the computationally less expensive Morph Server [20], especially for transitions requiring larger conformational changes which are not handled well by [20].

The accuracy of our approach is heavily dependent on how densely we sample the conformation space. Previously, this was user specified. However, if the resulting map of the energy landscape was not accurate enough, we were unable to increase the sampling density without restarting the entire computation. Here, we use an extension of our basic technique which incrementally samples the conformation space at increasingly denser resolution until our map of the energy landscape has stabilized.

Finally, we announce our protein folding server which uses our technique to generate transitional motions of a protein to the native state, or between selected protein conformations. We invite the community to help us enrich our publicly available database of motions by submitting to our server: <http://parasol.tamu.edu/foldingserver/>

2 Related Work

2.1 Protein Motion Models

Several computational approaches have been applied to study protein motions and folding, see Table 1. These include lattice models [12], energy minimization [47, 65], molecular dynamics [46, 25] Monte Carlo methods [17, 40]. Molecular dynamics and Monte Carlo methods provide a single, high quality transition pathway, but each run is computationally intensive. Statistical mechanical models [54, 1, 7], while computationally efficient, are limited to studying global averages of the energy landscape and kinetics and are unable to produce individual pathways.

2.2 Computing Macromolecular Motions

Gerstein et. al. have developed the Database of Macromolecular Movements [23, 20] to classify protein motions. Their Morph server produces a ‘morph’ movie between two conformations of interest. They can produce ‘morphs’ in just a few minutes on a desktop PC. Their database currently includes more than 240 distinct protein motions [41].

To create a ‘morph’ between two target conformations, they first perform an alignment. This allows the comparison of proteins with different sequences. Then, an iterative ‘sieve-fit’ procedure [45, 22] produces a superposition of the target conformations. A conformational ‘morph’ is created by interpolating the $C\alpha$ atom positions between the two superimposed conformations. Each intermediate conformation is energy minimized. This interpolation method, called adiabatic mapping, was selected because it has modest computational requirements yet produces chemically reasonable ‘morphs.’ Adiabatic mapping methods have problems with some kinds of large deformations [67], and hence in the future, other more sophisticated methods may be added.

2.3 Motion Planning and Molecular Motions

The motion planning problem is to find a valid path for a movable object from a start conformation to a goal conformation. The probabilistic roadmap method (PRM) [37] has been highly successful in solving the motion planning problem for objects with many degrees of freedom (dof).

PRMs work by first sampling random points in the movable object’s conformation space (C-space). C-space is the set of all possible positions and orienta-

tions of the movable object, valid or not [50]. Only those samples that meet certain feasibility requirements (e.g., collision free or low potential energy) are retained. The samples are connected to form a graph (or roadmap) by using some simple local planner (e.g., a straight line in C-space) to connect nearby points. This roadmap can then be used to find the motion between different start and goal pairs. To do so, the start and goal are connected to the roadmap, and a path between them is extracted if it exists. A major strength of PRMs is that they are simple to apply, even for high dof problems, only requiring the ability to randomly sample points in C-space and test them for feasibility.

PRMs have been applied to model molecular motions. The basic idea is to model the molecule as an articulated linkage, and to replace the typical collision detection validity check with some measure of the physical viability of the molecular conformation (e.g., potential energy). Singh, Latombe and Brutlag first applied PRMs to protein/ligand binding [61]. In subsequent work, our group used a PRM variant [2] on this problem [9]. Our group was the first to adapt the PRM framework to model protein folding pathways [4, 3, 63, 62]. Apaydin et al. [6, 5] have also applied PRMs to proteins, however their work differs from ours in several aspects. First, they model the protein at a much coarser level considering all secondary structure elements in the native state to already be formed and rigid. Second, while our focus is on studying the transition process, their focus has been to compare the PRM approach with other computational methods such as Monte Carlo simulation. In recent work, Cortes and Simeon used a PRM-based approach to model long loops in proteins [16]. Recently, we adapted the PRM framework to study RNA folding kinetics [66].

2.4 Rigidity Theory and Protein Flexibility

Several computational approaches have studied rigidity and flexibility in proteins. One approach infers the protein’s flexibility/rigidity by comparing different known conformations of the protein [57, 60, 11]. Molecular dynamics has been used to extract flexibility information from simulated motion [19, 51, 13, 38]. A third method studies rigidity and flexibility of a single conformation [27, 71, 36, 35, 52].

In this work, we use a rigidity analysis technique for counting constraints called the pebble game [30, 31, 32, 29] to better simulate protein motion. The pebble game belongs to the third class of approaches. It is fast

Approach	Folding Landscape	# Paths Produced	Path Quality	Compute Time	Native State Required
Molecular Dynamics	No	1	Good	Long	No
Monte Carlo	No	1	Good	Long	No
Statistical Model	Yes	0	N/A	Fast	Yes
PRM-Based	Yes	Many	Approx	Fast	Yes
Lattice Model	Not used on real proteins				

Table 1: A comparison of protein motion models.

and efficient; we can apply it to every conformation we sample. The pebble game has been an effective tool for studying protein flexibility and rigidity [33, 34, 59, 26, 44].

The pebble game is a constraint counting algorithm based on Laman’s theorem [43] which determines the dof in a two-dimensional graph, along with its rigid and flexible regions. In 2D, the pebble game assigns each vertex in the graph two pebbles, representing its two dof, see Figure 1. Then each edge/constraint in the graph is iteratively examined to determine if it is an independent constraint or if it is redundant. If two free pebbles can be placed on both endpoints of the edge, then the edge is marked independent and is covered by a pebble from one of its incident vertices. Once an edge is covered by a pebble, it remains covered during the rest of the game, although which vertex the pebble comes from may change. Pebbles may be rearranged by moving a free pebble from an incident vertex onto a covered edge and moving the covering pebble from the other incident vertex off of the edge. If pebbles cannot be rearranged to get two free pebbles on both of an edge’s endpoints, then the edge is marked redundant and indicates a rigid region in the graph. After every edge is examined, the remaining free pebbles indicate the graph’s dof.

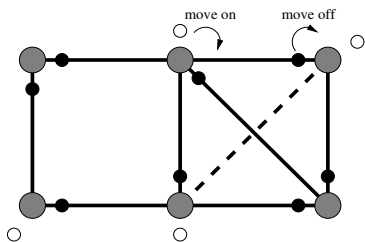


Figure 1: The result of the pebble game on a 2D graph. Each vertex has 2 associated pebbles. Free pebbles are indicated in white and covering pebbles in black. Independent constraints are marked with solid lines and redundant constraints with dashed lines. Redundant constraints indicate overstressed/rigid regions of the graph. Pebbles may be rearranged as shown.

The 2D pebble does not generalize to 3D for arbitrary graphs, but it can be applied to 3D bond-bending networks [29]. A bond-bending network is a truss structure with constraints between nearest neighbors and next-nearest neighbors. A protein, with fixed bond lengths and bond angles, forms a bond-bending network where atoms are modeled as vertices with 3 dof and bonds are modeled as edges. This is called the bar-joint model, see Figure 2 (a). In the 2D pebble game, edges/constraints may be placed in any order, however, in the 3D pebble game, order matters for correctness. The first constraint must be a nearest neighbor constraint (i.e., a bond length constraint). Then all the associated next-nearest neighbor constraints (i.e., associated bond angle constraints) must be placed before placing another nearest neighbor constraint. This model has been successfully used by several applications to study protein rigidity and flexibility [33, 34, 59, 26, 44].

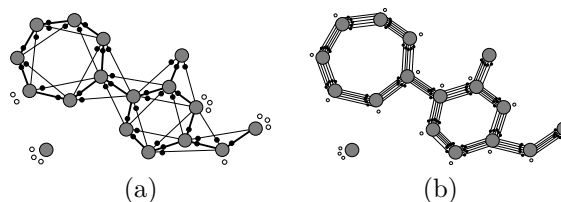


Figure 2: Rigidity models for a sample molecule: (a) bar-joint and (b) body-bar. For the bar-joint model, nearest neighbor constraints (i.e., bond lengths) are shown with thick lines and next-nearest neighbor constraints (i.e., bond angles) are shown with thin lines. Both models yield the same dof and rigid/flexible regions.

An alternative model, the body-bar model, represents atoms as rigid bodies with 6 dof and the torsional bonds between them as 5 bars/constraints [69], see Figure 2 (b). This model is conjectured to be equivalent to the bar-joint model [29].

3 Preliminaries

3.1 Protein Model

In our work, we model the protein as an articulated linkage. Using a standard modeling assumption for proteins that bond angles and bond lengths are fixed [64], the only dof in our model are the backbone’s phi and psi torsional angles which are modeled as revolute joints with values $[0, 2\pi)$.

3.2 Potential Energy Calculations

As in our previous work, we use a coarse potential function similar to [46]. We use a step function approximation of the van der Waals potential component and model side chains as spheres with zero dof. If any two spheres are too close (i.e., less than 2.4\AA during sampling and 1.0\AA during connection), a very high potential is returned. Otherwise, the potential is:

$$U_{tot} = \sum_{\text{restraints}} K_d \{[(d_i - d_0)^2 + d_c^2]^{1/2} - d_c\} + E_{hp} \quad (1)$$

where K_d is 100 kcal/mol and $d_0 = d_c = 2\text{\AA}$ as in [46]. The first term represents constraints favoring known secondary structure through main-chain hydrogen bonds and disulphide bonds, and the second term is the hydrophobic effect. The hydrophobic effect is computed as follows: if two hydrophobic residues are within 6\AA of each other, then the potential is decreased by 100 kJ/mol.

4 A PRM-based Approach for Modeling Molecular Motions

We have successfully applied the PRM framework to study protein folding pathways [4, 3, 63, 62]. The strategy follows the general PRM methodology sketched in Section 2.3. First, different conformations of the protein are sampled. A sample q , with potential energy $E(q)$, is accepted with the probability:

$$P_{\text{accept}}(q) = \begin{cases} 1 & \text{if } E(q) < E_{\min} \\ \frac{E_{\max} - E(q)}{E_{\max} - E_{\min}} & \text{if } E_{\min} \leq E(q) \leq E_{\max} \\ 0 & \text{if } E(q) > E_{\max} \end{cases} \quad (2)$$

where E_{\min} is the potential energy of the open chain and E_{\max} is $2E_{\min}$. Next, node connection is done in the same way as traditional PRMs except that each connection is assigned a weight to reflect its energetic feasibility. The weight for the edge (q_1, q_2) is a function of all the intermediate conformations along the edge $\{q_1 = c_0, c_1, \dots, c_{n-1}, c_n = q_2\}$. For

each pair of consecutive conformations c_i and c_{i+1} , the probability P_i of transitioning from c_i to c_{i+1} depends on the difference in their potential energies $\Delta E_i = E(c_{i+1}) - E(c_i)$:

$$P_i = \begin{cases} e^{-\frac{\Delta E_i}{kT}} & \text{if } \Delta E_i > 0 \\ 1 & \text{if } \Delta E_i \leq 0 \end{cases} \quad (3)$$

This keeps the detailed balance between two adjacent states, and enables the weight of an edge to be computed by summing the logarithms of the probabilities for consecutive pairs of conformations in the sequence:

$$w(q_1, q_2) = \sum_{i=0}^{n-1} -\log(P_i) \quad (4)$$

Edge weights are not transition rates, but the logarithm of transition rates. This enables edge weights to follow the summation rule (instead of the multiplication rule for transition rates) and facilitates the use of graph algorithms to extract shortest paths.

The samples and connections form a roadmap from which we can typically extract thousands of transition pathways. In just a few hours on a desktop PC, we obtained promising results for many small proteins (60–100 residues) and validated our pathways by comparing the secondary structure formation order with known experimental results [4]. In one case study our technique was sensitive enough to identify subtle differences in folding behaviors for structurally similar proteins G and L [63].

4.1 Rigidity-Based Sampling

The roadmap produced by our technique is an approximation of the protein’s energy landscape. Roadmap quality is measured both by how realistic (as compared to experimental data) are the pathways it contains and by how many samples are required to achieve the desired level of accuracy. The latter is important because it determines how much computation is required and thus what size molecules can be analyzed.

Hence, sampling is the key to producing a good approximation of the protein’s potential landscape. We first note that only a relatively small portion of the conformation space ‘near’ the target conformation(s) for study is of interest in modeling motions. This implies that we should not use uniform sampling of conformation space — it would require very dense sampling to adequately cover the region near the target conformations and is therefore infeasible for all but very small molecules.

In previous work [4, 3, 63, 62], we obtained a denser distribution of samples near the target conformation

through an iterative sampling process where we apply small perturbations to existing conformations, beginning with the target conformation. The perturbations were generated according to a Gaussian distribution centered around the existing conformation. This approach works fairly well, but still requires many samples (e.g., 10,000) for relatively small proteins (e.g., 60–100 residues). To apply our method to larger proteins, we need strategies to generate ‘better’ samples. The conformations sampled should be more physically realistic and moreover, since we are interested in modeling motion, they should represent ‘stepping stones’ for conformational transitions.

In this work, we follow the same strategy as before, but use rigidity analysis to restrict how a conformation is perturbed. To perturb a given conformation, we first compute its rigid and flexible regions using the pebble game. We then perturb flexible dof with a high probability (P_{flex}) and rigid dof with a low probability (P_{rigid}). Perturbing rigid dof ensures good coverage of the space.

Rigidity Model. We employ the body-bar model to analyze a conformation’s rigidity. The body-bar model is composed of rigid bodies with 6 dof and bars (constraints) connecting them each requiring 1 dof/pebble for covering [69]. This model is conjectured to be equivalent to the bar-joint model for bond bending networks. With the body-bar model, we can represent the protein at a residue level, a closer match to our phi-psi model for sampling than the bar-joint model. (The bar-joint model yields a more detailed all-atoms view.)

For rigidity analysis, we model the protein simply as a chain of rigid bodies, each representing one torsional dof along the protein’s backbone, see Figure 3. We model each peptide bond and disulphide bond with 5 bars, each hydrogen bond with 2 bars, and each hydrophobic contact with 1 bar. On all conformations tested, this yields the same rigid and flexible regions as the equivalent bar-joint model on an all-atoms representation of the protein.

To perturb an existing conformation, we first use rigidity analysis to determine which bonds along the backbone are independently flexible, dependently flexible, and rigid, see Figure 4. Independently flexible bonds can be perturbed without affecting the rest of the bonds in the system. Dependently flexible bonds form a set of bonds such that perturbing any one of these bonds results in a corresponding perturbation in the rest of the set.

If the bond is independently flexible, we perturb it with a high probability, P_{flex} . If the bond is rigid, we perturb it with a low probability, P_{rigid} . For each set

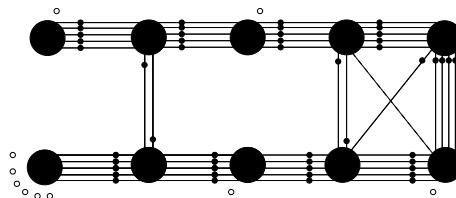


Figure 3: Our phi-psi model of a protein for rigidity analysis with 5 residues. We model each residue with two rigid bodies, each representing a torsional dof. Peptide bonds and disulphide bonds are modeled with 5 bars, hydrogen bonds with 2 bars, and hydrophobic interactions with 1 bar. Redundant constraints are identified (dashed lines) after playing the pebble game.

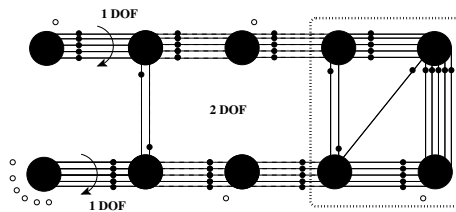


Figure 4: Rigidity analysis results for our phi-psi model of the protein in Figure 3. There are 10 total dof. The rigid cluster is enclosed with a dotted line. Constraints within these clusters are not flexible. The constraints in the dependent hinge set are marked with dashed lines. In this example, the set contains 2 dof. Independently flexible constraints are marked with an arc, each having 1 dof. (Redundant constraints are not shown for clarity.)

of dependently flexible bonds, we randomly select d bonds to perturb with the probability P_{flex} and perturb the remaining bonds with the probability P_{rigid} where d is the internal dof in the set.

Rigidity Map. We can also use rigidity analysis to define a new residue mapping and distance metric. A rigidity map, r , is similar to a contact map, except residue pairs are marked if they have the same rigidity relationship. We mark the residue pairs as follows:

$$r(i, j) = \begin{cases} 2 & \text{if } i \text{ and } j \text{ are in the same rigid cluster} \\ 1 & \text{if } i \text{ and } j \text{ are in the same dep. hinge set} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Figure 5(a) shows the rigidity map of the native state for protein G. Rigid clusters are colored in black, and dependent hinge sets are colored in green, for each dof pair. Rigidity maps provide a convenient way to define a rigidity distance metric, $r_{dist}(q_1, q_2)$, between two conformations q_1 and q_2 :

$$r_{dist}(q_1, q_2) = \sum_{0 \leq i < j \leq n} (r_{q_1}(i, j) \neq r_{q_2}(i, j)) \quad (6)$$

where $\frac{n}{2}$ is the length of the protein.

4.2 Automatic Roadmap Construction

The roadmap’s accuracy depends on the sampling density. Previously, this was user specified and difficult to tune. Here, we automate roadmap construction by building the roadmap incrementally [70]. We first build a roadmap with a low sampling density as described above. Then, we test the roadmap to see if it has stabilized as specified by a set of evaluation criteria. We continue to augment the roadmap with more samples and connections until it satisfies the evaluation criteria. This provides two key advantages over our previous work: (1) the roadmap is constructed automatically at the appropriate resolution and (2) we reuse all previous computation reducing runtime cost by several factors.

For protein folding problems, we build a roadmap until the secondary structure formation order along its pathways stabilizes. We consider a piece of secondary structure ‘formed’ when the distance between its rigidity map (defined in Section 4.1) and that piece’s rigidity map in the target state is within 0.8, normalized to the range [0,1]. The pathway’s secondary structure formation order is then the order at which pieces are ‘formed’ along the pathway. We examine every pathway in the roadmap from an unstructured conformation to the target state and group them by this ordering. We consider the roadmap stable when the percentage of each group in the current roadmap does not vary from the previous roadmap by more than 10%.

4.3 Mapping Specified Transitions

We extend our PRM framework to study specific large-scale conformational changes by iteratively sampling around each target conformation and connecting samples together as described earlier in Section 4. This ensures that our roadmaps contain the target conformations, as well as the transitions between them. This roadmap approximates the energy landscape encompassing the conformational change under study.

With this approach, we can study problems such as transitions between known folding intermediates, transitions between bound and unbound conformations to a ligand, misfolded proteins, and allosteric interactions. For example, calmodulin undergoes two large-scale conformational changes when binding to Ca^{2+} thereby regulating many cellular processes [68, 18, 21]. Changing conformations allows it to bind to over 100 different proteins [18, 14]. Several devastating diseases such as scrapie in sheep and goats, bovine spongiform encephalopathy (Mad Cow disease), and Creutzfeldt-Jakob disease in humans are

caused by misfolded proteins called prions [58, 15]. The prion protein has a significantly different structure in the diseased state than in the normal state. Insight into how this change occurs could help develop better drugs.

To study specific large-scale conformational changes, we interleave sampling and connection to incrementally build a roadmap as described in Section 4.2. The only difference here is we sample around each target conformation (as in Section 4.1) during each round of roadmap construction. Then we connect nearby samples together and compute edge weights as before. We continue this process until the roadmap adequately represents the protein’s energy landscape near the target conformations and between them. From this roadmap, we can extract multiple low energy transition pathways between target conformations and characterize the energy barriers between them.

To determine if the roadmap is sufficient, we use a different type of evaluation criteria than the one for protein folding given in Section 4.2. For transition problems, we instead build the roadmap until the maximum network flow between each target conformation pair is above a threshold. The maximum network flow problem originates from graph theory. Edges are assigned a capacity, and the goal is to determine how much flow can be achieved between two points in the graph. For this application, we define edge capacity as the inverse of the edge weight. Thus, the maximum network flow between two conformations approximates the transition rate between them [42].

5 Results and Discussion

In this section, we investigate the ability of our rigidity-based sampling strategy to efficiently sample the protein’s conformation space. We also look at examples of large-scale conformational change between specific target states for several small proteins and compare our results with ‘morphs’ from the Database of Macromolecular Movements [23, 20]. In all experiments, we set P_{flex} to 0.8 and P_{rigid} to 0.2. We use a straight line local planner and attempt to connect each conformation with its 50 nearest neighbors. During roadmap construction, we measure distance between two conformations as the difference between their rigidity maps as described in Section 4.1.

5.1 Improved Sampling

Rigidity analysis coupled with automatic roadmap construction greatly improves the efficiency of our PRM framework by restricting the sample space in a physically realistic way. We can build smaller roadmaps that better reflect the protein’s energy landscape. We built roadmaps for several previously studied proteins [3, 62]. For each protein, we compare our new automatic framework with rigidity-based sampling to our previous sampling technique with user specified sampling density. Table 2 shows the roadmap size and connectivity from both methods for each protein. For all results shown here, both methods give the same secondary structure formation order distribution. When available, these results also indicate the same dominant secondary structure formation order seen in experiment [48]. In all cases, the rigidity-based roadmaps produce equivalent folding pathways as the previous method with smaller, more efficient roadmaps. It also increases connectivity. With rigidity analysis, we can study much larger proteins than previously.

Case study of proteins G, L NuG1, and NuG2. Proteins G, L, and mutated forms of protein G, NuG1 and NuG2 [55], present a good test case for our technique because they are known to fold differently although they are structurally similar. All proteins are composed of an α -helix and a 4-stranded β -sheet. β strands 1 and 2 form the N-terminal hairpin (hairpin 1) and β strands 3 and 4 form the C-terminal hairpin (hairpin 2). Native state out-exchange experiments and pulse labeling/competition experiments for proteins G and L indicates that β -hairpin 1 forms first in protein L, and β -hairpin 2 forms first in protein G [48]. This is consistent with Φ -value analysis on G [53] and L [39]. In [55], protein G is mutated in both hairpins to increase the stability of hairpin 1 and decrease the stability of hairpin 2. Φ -value analysis indicates that the hairpin formation order for both NuG1 and NuG2 is switched from the wild type.

Our previous sampling strategy [63] was able to capture the folding differences between proteins G and L, but not between protein G and NuG1 or NuG2. With our new rigidity-based sampling and analysis, we are now also able to capture the correct folding behavior of NuG1 and NuG2, see Table 3. In addition, to detecting the correct folding behavior, our rigidity-based technique can also help to explain the stability shift in NuG1 and NuG2. For example, consider their native state rigidity maps shown in Figure 5. In all four proteins, the central alpha helix remains completely rigid, as expected. We also see increased rigidity from protein G to NuG1 and NuG2 in hairpin 1 as

suggested in [55]. For the other regions, the rigidity maps show more similarity between the mutated forms of protein G and protein L than between protein G, reflecting the similarity/difference in folding behaviors between the mutated protein and the two others.

Protein	Experimental Formation Order	Rigidity Results	
		Formation Order	%
G	[$\alpha, \beta 1, \beta 3, \beta 4$], $\beta 2$ ¹ [$\alpha, \beta 4$], [$\beta 1, \beta 2, \beta 3$] ²	$\alpha, \beta 3-4, \beta 1-2$	99.4
		$\beta 3-4, \alpha, \beta 1-2$	0.6
L	[$\alpha, \beta 1, \beta 2, \beta 4$], $\beta 3$ ¹ [$\alpha, \beta 1$], [$\beta 2, \beta 3, \beta 4$] ²	$\beta 1-2, \alpha, \beta 3-4$	100.0
NuG1	$\beta 1-2, \beta 3-4$ ³	$\alpha, \beta 1-2, \beta 3-4$	97.6
		$\beta 1-2, \alpha, \beta 3-4$	1.6
		$\alpha, \beta 3-4, \beta 1-2$	0.4
		$\beta 3-4, \alpha, \beta 1-2$	0.4
NuG2	$\beta 1-2, \beta 3-4$ ³	$\alpha, \beta 1-2, \beta 3-4$	96.6
		$\beta 1-2, \alpha, \beta 3-4$	1.1
		$\alpha, \beta 3-4, \beta 1-2$	0.6
		$\beta 3-4, \alpha, \beta 1-2$	0.6
		$\beta 3-4, \beta 1-2, \alpha$	1.1

Table 3: Comparison of secondary structure formation orders for proteins G, L, NuG1, and NuG2 with known experimental results. Brackets indicate there was no clear order. In all cases, our technique predicted the secondary structure formation order seen in experiment.

We can also use rigidity-based analysis to study dynamic changes along a folding pathway. Figure 6 shows how rigidity changes along the folding pathway. We see a distinction between the profiles for protein G where hairpin 2 forms first and the others where hairpin 1 forms first. For protein G, the rigidity profile (a) shows a plateau halfway along the folding pathway, where the others do not. Additionally, protein G (b) exhibits larger changes in rigidity earlier in the pathway while the others exhibit larger changes later in the pathway.

5.2 Large-Scale Conformational Change

Calmodulin is a 148-residue signaling protein that binds to Ca^{2+} to regulate several processes in the cell [68, 18, 21]. It is composed of 4 EF-hands joined by a flexible central α helix. Each calmodulin domain binds 1 calcium ion in each EF-hand pair [49], with the two domains acting independently [8]. When calmodulin binds to Ca^{2+} , it undergoes two large-scale conformational changes: (1) the central α helix linking the

¹Hydrogen out-exchange experiments [48].

²Pulsed labeling/competition experiments [48].

³ Φ -value analysis [55].

PDB Identifier	Length	Structure	Gaussian Sampling				Rigidity Sampling			
			Nodes	Edges	$N + E$	E/N	Nodes	Edges	$N + E$	E/N
1AB1	46	$2\alpha + 2\beta$	24206	386974	411180	15.99	6000	158286	164286	26.38
1CCM	46	$1\alpha + 3\beta$	43646	728964	772610	16.70	10000	456080	466080	45.61
1RDV	52	$2\alpha + 3\beta$	33691	457392	491083	13.58	4000	166702	170702	41.68
1EGF	53	3β	27356	391146	418502	14.30	4000	164902	168902	41.23
1PRB	53	5α	44551	696708	741259	15.64	4000	126562	130562	31.64
1SMU	54	$3\alpha + 3\beta$	35501	557416	592917	15.70	4000	158852	162852	39.71
1FCA	55	$2\alpha + 4\beta$	38216	489840	528056	12.82	4000	162526	166526	40.63
1VGH	55	$1\alpha + 4\beta$	38216	631936	670152	16.54	4000	157454	161454	39.36
1GB1	56	$1\alpha + 4\beta$	34236	912908	947144	26.66	4000	160552	164552	40.14
1SHG	57	5β	24696	270232	294928	10.94	18000	654884	672884	36.38
1BPI	58	$2\alpha + 2\beta$	28426	399418	427844	14.05	4000	112010	116010	28.00
4PTI	58	$2\alpha + 2\beta$	39121	389468	428589	9.96	4000	160100	164100	40.03
1HCC	59	7β	33691	453628	487319	13.46	28000	1079904	1107904	38.57
1BDD	60	3α	58486	888298	946784	15.19	6000	195950	201950	32.66
1TCP	60	$2\alpha + 2\beta$	32786	354262	387048	10.81	4000	163692	167692	40.92
2ADR	60	$2\alpha + 2\beta$	42723	701942	744665	16.43	8000	339498	347498	42.44
2PTL	62	$1\alpha + 4\beta$	23921	281334	305255	11.76	4000	159728	163728	39.93
1COA	64	$1\alpha + 5\beta$	27746	403438	431184	14.54	4000	160838	164838	40.21
2CI2	65	$2\alpha + 5\beta$	27746	389670	417416	14.04	8000	228706	236706	28.59
1NYF	67	5β	23921	262376	286297	10.97	6000	249450	255450	41.58
1MJC	69	7β	23481	226942	250423	9.66	4000	153140	157140	38.29
1HOE	74	7β	30626	184012	214638	6.01	4000	103668	107668	25.92
1UBQ	76	$1\alpha + 5\beta$	25206	236216	261422	9.37	4000	154192	158192	38.55
1O6X	81	$2\alpha + 3\beta$	40931	342138	383069	8.36	4000	133544	137544	33.39
1PBA	81	$4\alpha + 3\beta$	26476	203974	230450	7.70	8000	282960	290960	35.37
2ABD	86	5α	27956	681796	709752	24.39	18000	953900	971900	52.99

Table 2: Comparison of rigidity-based sampling to previous work for several proteins. In all cases, rigidity-based sampling significantly reduces the required roadmap size ($N + E$) to produce equivalent pathways. It also increased roadmap connectivity (E/N).

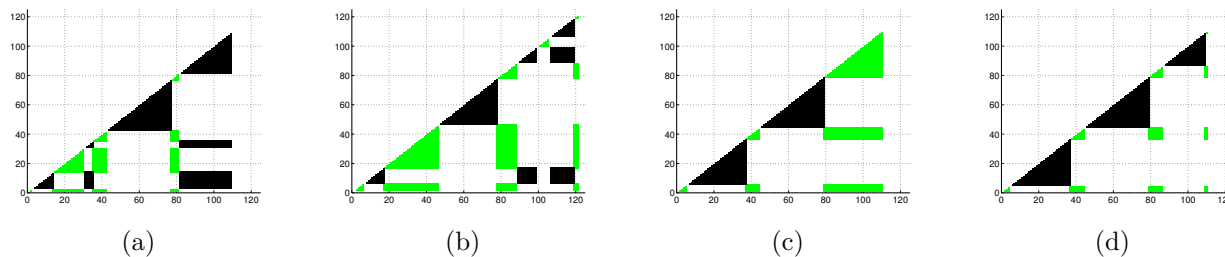


Figure 5: Rigidity maps for proteins (a) G, (b) L, (c) NuG1, and (d) NuG2.

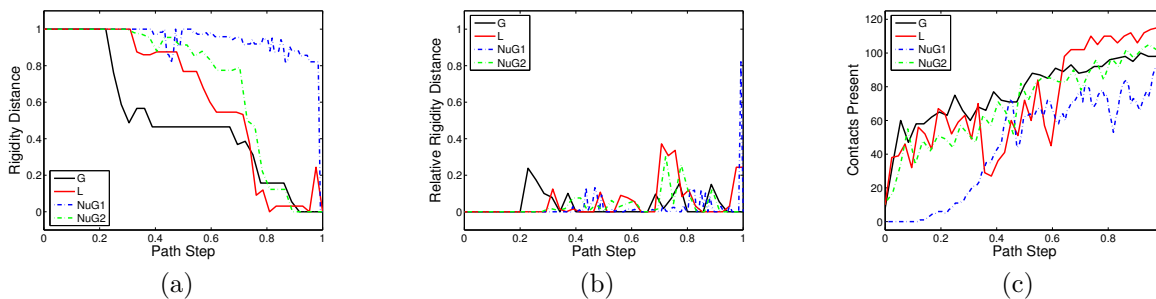


Figure 6: Folding pathway profiles for proteins G, L, NuG1, and NuG2: (a) rigidity distance to the target state, (b) relative rigidity distance to the target state, and (c) contacts present. There is a distinction between the profiles for protein G where hairpin 2 forms first and the others where hairpin 1 forms first. For protein G, the rigidity profile (a) shows a plateau halfway along the folding pathway, where the others do not. Additionally, protein G (b) exhibits larger changes in rigidity earlier in the pathway while the others exhibit larger changes later in the pathway.

C-terminal and N-terminal domains unravels to bring the protein from a dumbbell conformation to a more globular conformation [28] (Figure 7a–b) and (2) the α helices in each domain reorganize [56] (Figure7c–d).

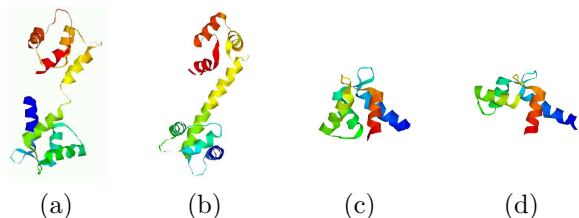


Figure 7: Conformational changes of Calmodulin: (a) calcium-free state (1CFD) to (b) bound state (1CLL) and conformation changes of the N-terminal domain: (c) calcium-free to (d) bound.

We built a roadmap biased towards both target states as outlined in Section 4.3. Figure 8 compares pathway profiles of the most energetically feasible transition between the two states in our roadmap and ‘morphs’ of various resolution obtained from the Database of Macromolecular Movements [23, 20]. Recall, as mentioned in Section 2.2, that the ‘morphs’ are generated using an interpolation method called

adiabatic mapping [67] that was chosen as a computationally efficient method which produces chemically realistic ‘morphs.’ We examined pathway profiles for energy, contacts present, actual dof as computed by rigidity analysis, RMSD distance to the target states, and rigidity distance to the target states. Note that since the Morph server alters the original target conformations, their profile endpoints do not always align with our pathways. One striking observation is the regularity of the concavities for the ‘morphs’ corresponding to the various resolution levels across all the profiles except for the RMSD profiles in which the RMSD to the target states seems to change monotonically with the path step. These regularities that are observed in the ‘morphs’ would not be expected in actual transition pathways, e.g., one would not expect a monotonic increase in RMSD from 1CFD to 1CLL as occurs in the ‘morphs’. In contrast, the profiles of our roadmap pathways are more plausible — they exhibit trends, but also have reasonable fluctuations. Indeed, this type of behavior has also been observed by other researchers. For example, in [72], Monte Carlo simulations indicate a wide range of transition pathways and event durations.

Figure 8(a,c) shows the contacts present and ac-

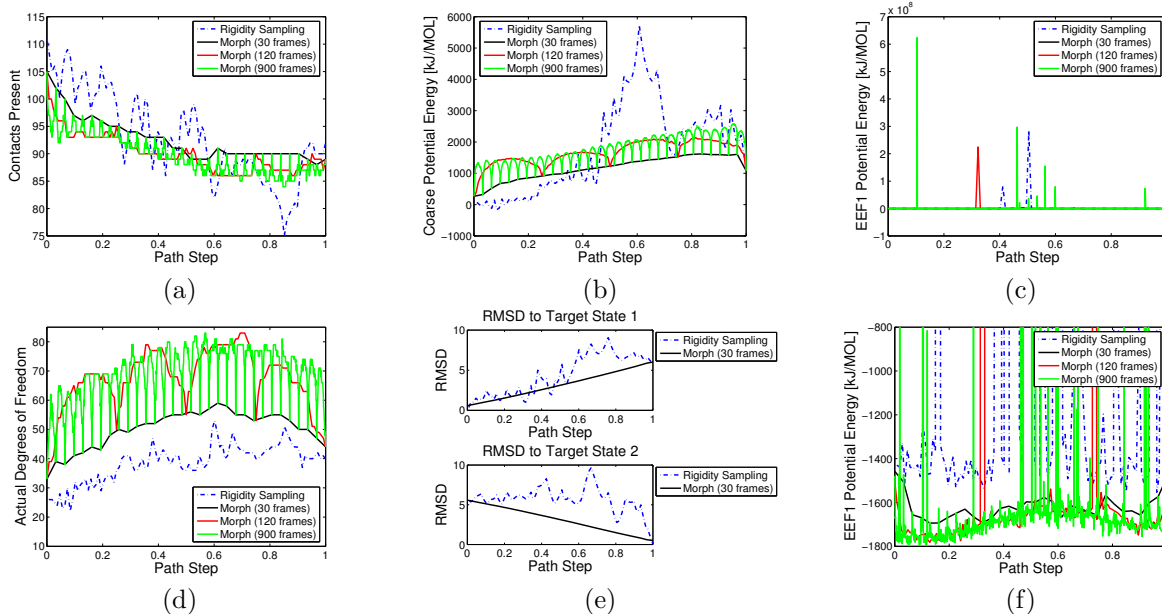


Figure 8: Pathway profiles for the N-terminal domain of calmodulin: (a) contacts present, (b) coarse potential energy, (c,f) all-atoms potential energy, (d) actual dof as computed by rigidity analysis, and (e) RMSD to both target states. For RMSD, we only show the 30 frame ‘morph’ because all resolutions are nearly identical.

tual dof as computed by rigidity analysis along the pathway. Note that the protein does not completely unfold during the transition, but maintains a large number of contacts and loses few dof. In general, the actual dof is inversely proportional to the number of contacts present in the conformation. It is interesting to note, however, that we see a slight break in this relationship on the second half of the pathway where the peaks in dof to not match up with the peaks in number of contacts. Regions of the protein become stressed when the number of contacts increases without a corresponding decrease in dof.

We investigated several other protein transitions in a similar way, see Table 4. We measure % dof gained as the difference between the maximum dof along the pathway and the minimum dof of the starting/ending conformations, as a percentage of the total dof possible ($2 \times \text{length}$). Most transitions do not involve a complete unfolding of the protein. In fact, several have % dof gain less than 10%. We also captured different types of transitions including smooth transitions without any significant energy barriers (i.e., 1PRV, 1BMR, and 1FOX) and those with multiple energy barriers (i.e., 2VGH and 1CMF).

We also compared ‘morphs’ of various resolutions to our transition pathways when possible. (The Morph server was not able to produce some higher resolution ‘morphs’ for transitions 1BMR–1FH3 and 1PRV–

Transition IDs		Length	% Dof Gained	# Barriers
2VGH	1VGH	55	21.8	2
1PRV	1PRU	56	5.4	0
1BMR	1FH3	67	32.8	0
1CFD	1CLL	72	18.1	1
1CMF	1CMG	73	24.7	2
1FOX	2FOW	76	3.9	0
1PFH	1HDN	85	43.5	1

Table 4: Pathway results for several transitions studied. Most transitions do not involve a significant unfolding.

1PRU.) Across all transitions, we observed the same concavity pattern phenomenon for the ‘morph’ transitions as seen in calmodulin (Figure 8) for energy, contacts, degrees of freedom, and rigidity distance to the targets. Here also, the RMSD to the target states essentially changed monotonically with the path step. Again, our pathways did not exhibit these unrealistic regularities. Additional path profiles for all the transitions studied here can be found on our folding server: <http://parasol.tamu.edu/foldingserver/>

6 Conclusion

In this paper, we describe how we can augment our PRM-based approach to study a broader range of motions for larger proteins. In particular, we proposed a method based on rigidity theory to sample conformation space more efficiently and to generate transitions between specified conformations. We also demonstrated that our approach yields more physically realistic transitions than those produced by the computationally less expensive Morph server.

We invite the community to help us enrich our publicly available database of motions by submitting to our web-based folding server: <http://parasol.tamu.edu/foldingserver/>

7 Acknowledgments

We would like to thank Ileana Streamu, Leslie Kuhn, and Walter Whitely for introducing us to rigidity theory and for many insightful discussions. We also thank Luke Hunter, Kasia Leyk, and Dawen Xie for their help with experiments.

References

- [1] E. Alm and D. Baker. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA*, 96(20):11305–11310, 1999.
- [2] N. M. Amato, O. B. Bayazit, L. K. Dale, C. V. Jones, and D. Vallejo. OBPRM: An obstacle-based PRM for 3D workspaces. In *Robotics: The Algorithmic Perspective*, pages 155–168, Natick, MA, 1998. A.K. Peters. Proceedings of the Third Workshop on the Algorithmic Foundations of Robotics (WAFR), Houston, TX, 1998.
- [3] N. M. Amato, K. A. Dill, and G. Song. Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures. *J. Comput. Biol.*, 10(3-4):239–256, 2003. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2002.
- [4] N. M. Amato and G. Song. Using motion planning to study protein folding pathways. *J. Comput. Biol.*, 9(2):149–168, 2002. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2001.
- [5] M. Apaydin, D. Brutlag, C. Guestrin, D. Hsu, and J.-C. Latombe. Stochastic roadmap simulation: An efficient representation and algorithm for analyzing molecular motion. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 12–21, 2002.
- [6] M. Apaydin, A. Singh, D. Brutlag, and J.-C. Latombe. Capturing molecular energy landscapes with probabilistic conformational roadmaps. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 932–939, 2001.
- [7] D. Baker. A surprising simplicity to protein folding. *Nature*, 405:39–42, 2000.
- [8] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax. Backbone dynamics of calmodulin studied by n relaxation using inverse detected two-dimensional nmr spectroscopy: the central helix is flexible. *Biochemistry*, 31:5269–5278, 1992.
- [9] O. B. Bayazit, G. Song, and N. M. Amato. Ligand binding with OBPRM and haptic user input: Enhancing automatic motion planning with virtual touch. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 954–959, 2001. This work was also presented as a poster at *RECOMB 2001*.
- [10] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.
- [11] N. Boutonnet, M. Rooman, and S. Wodak. Automatic analysis of protein conformational changes by multiple linkage clustering. *J. Mol. Biol.*, 253:633–647, 1995.
- [12] J. Bryngelson, J. Onuchic, N. Socci, and P. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Protein Struct. Funct. Genet.*, 21:167–195, 1995.
- [13] D. Case. Molecular dynamics and normal mode analysis of biomolecular rigidity. In M. Thorpe and P. Duxbury, editors, *Rigidity theory and applications*, pages 329–344. Kluwer Academic/Plenum Publishers, 1999.
- [14] M. R. Celio, T. Pauls, and B. Schwaller. *Guidebook to the Calcium-Binding Proteins*. Oxford University Press, Oxford, 1996.
- [15] J. Collinge. Prion diseases of humans and animals: Their causes and molecular basis. *Annu. Rev. Neurosci.*, 24:519–550, 2001.

- [16] J. Cortes, T. Simeon, M. Remaud-Simeon, and V. Tran. Geometric algorithms for the conformational analysis of long protein loops. *J. Computat. Chem.*, 25, 2004.
- [17] D. Covell. Folding protein α -carbon chains into compact forms by Monte Carlo methods. *Proteins: Struct. Funct. Genet.*, 14(4):409–420, 1992.
- [18] A. Crivici and M. Ikura. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.*, 24:85–116, 1995.
- [19] Y. Duan and M. Karplus. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740–744, 1998.
- [20] N. Echols, D. Milburn, and M. Gerstein. Molmovdb: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, 31:478–482, 2003.
- [21] J. Evenas, A. Malmendal, and S. Forsen. Calcium. *Curr. Op. Chem. Biol.*, 2(2):293–302, 1998.
- [22] M. Gerstein and C. H. Chothia. Analysis of protein loop closure: Two types of hinges produce one motion in lactate dehydrogenase. *J. Mol. Biol.*, 220:133–149, 1991.
- [23] M. Gerstein and W. Krebs. A database of macromolecular motions. *Nucleic Acids Res.*, 26:4280–4290, 1998.
- [24] C.-S. Goh, D. Milburn, and M. Gerstein. Conformational changes associated with protein-protein interactions. *Curr. Op. Str. Biol.*, 14:1–6, 2004.
- [25] J. Haile. *Molecular Dynamics Simulation: elementary methods*. Wiley, New York, 1992.
- [26] B. M. Hesperheide, A. Rader, M. Thorpe, and L. A. Kuhn. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Gra. Model.*, 21:195–207, 2002.
- [27] L. Holm and C. Sander. Parser for protein folding units. *Proteins*, 19:256–268, 1994.
- [28] M. Ikura, G. M. Clore, A. M. Gronenborn, G. Zhu, C. B. Klee, and A. Bax. Solution structure of a calmodulin-target peptide complex by multidimensional nmr. *Science*, 256:632–638, 1992.
- [29] D. Jacobs. Generic rigidity in three-dimensional bond-bending networks. *J. Phys. A: Math. Gen.*, 31:6653–6668, 1998.
- [30] D. Jacobs and M. Thorpe. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.*, 75(22):4051–4054, 1995.
- [31] D. Jacobs and M. Thorpe. Generic rigidity percolation in two dimensions. *Phys. Rev. E*, 53(4):3682–3693, 1996.
- [32] D. J. Jacobs and B. Hendrickson. An algorithm for two dimensional rigidity percolation: The pebble game. *J. Comp. Phys*, 137:346–368, 1997.
- [33] D. J. Jacobs, L. A. Kuhn, and M. F. Thorpe. Flexible and rigid regions in proteins. In M. Thorpe and P. Duxbury, editors, *Rigidity Theory and Applications*, pages 357–384. Kluwer Academic/Plenum Publishers, 1999.
- [34] D. J. Jacobs, A. Rader, L. A. Kuhn, and M. Thorpe. Protein flexibility predictions using graph theory. *Proteins Struct. Funct. Genet.*, 44:150–165, 2001.
- [35] J. Janin and S. Wodak. Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol.*, 42:21–78, 1983.
- [36] P. Karplus and G. Schulz. Prediction of chain flexibility in proteins. *Naturwissenschaften*, 72:212–213, 1985.
- [37] L. E. Kaviraki, P. Svestka, J. C. Latombe, and M. H. Overmars. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans. Robot. Automat.*, 12(4):566–580, August 1996.
- [38] O. Keskin, R. Jernigan, and I. Bahar. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.*, 78:2093–2106, 2000.
- [39] D. E. Kim, C. Fisher, and D. Baker. A breakdown of symmetry in the folding transition state of protein l. *J. Mol. Biol.*, 298:971–984, 2000.
- [40] A. Kolinski and J. Skolnick. Monte Carlo simulations of protein folding. *Proteins Struct. Funct. Genet.*, 18(3):338–352, 1994.
- [41] W. G. Krebs and M. Gerstein. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.*, 28:1665–1675, 2000.

- [42] S. V. Krivov and M. Karplus. Free energy connectivity graphs: Application to peptide models. *J. Chem. Phys.*, 114(23):10894–10903, 2002.
- [43] G. Laman. On graphs and rigidity of plane skeletal structures. *J. Eng. Math.*, 4:331–340, 1970.
- [44] M. Lei, M. I. Zavodsky, L. A. Kuhn, and M. F. Thorpe. Sampling protein conformations and pathways. *J. Comput. Chem.*, 25:1133–1148, 2004.
- [45] A. M. Lesk. *Protein Architecture: A Practical Approach*. IRL Press, Oxford, 1991.
- [46] M. Levitt. Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.*, 170:723–764, 1983.
- [47] M. Levitt and A. Warshel. Computer simulation of protein folding. *Nature*, 253:694–698, 1975.
- [48] R. Li and C. Woodward. The hydrogen exchange core and protein folding. *Protein Sci.*, 8(8):1571–1591, 1999.
- [49] S. Linse, A. Helmersson, and S. Forsen. Calcium binding to calmodulin and its globular domains. *J. Biol. Chem.*, 266:8050–8054, 1991.
- [50] T. Lozano-Pérez. Spatial planning: A configuration space approach. *IEEE Trans. Comput.*, C-32:108–120, 1983.
- [51] J. Ma and M. Karplus. The allosteric mechanism of the chaperonin groel: a dynamic analysis. *Proc. Natl. Acad. Sci. USA*, 95:8502–8507, 1998.
- [52] V. Maiorov and R. Abagyan. A new method for modeling large-scale rearrangements of protein domains. *Proteins*, 27:410–424, 1997.
- [53] E. L. McCallister, E. Alm, and D. Baker. Critical role of β -hairpin formation in protein g folding. *Nat. Struct. Biol.*, 7(8):669–673, 2000.
- [54] V. Muñoz, E. R. Henry, J. Hoferichter, and W. A. Eaton. A statistical mechanical model for β -hairpin kinetics. *Proc. Natl. Acad. Sci. USA*, 95:5872–5879, 1998.
- [55] S. Nauli, B. Kuhlman, and D. Baker. Computer-based redesign of a protein folding pathway. *Nature Struct. Biol.*, 8(7):602–605, 2001.
- [56] M. R. Nelson and W. J. Chazin. An interaction-based analysis of calcium-induced conformational changes in ca²⁺ sensor proteins. *Protein Sci.*, 7:270–282, 1998.
- [57] W. Nichols, G. Rose, L. T. Eyck, and B. Zimm. Rigid domains in proteins: an algorithmic approach to their identification. *Proteins*, 23:38–48, 1995.
- [58] S. Prusiner. Prions. *Proc. Natl. Acad. Sci. USA*, 95(23):13363–13383, 1998.
- [59] A. Rader, B. M. Hespeneide, L. A. Kuhn, and M. Thorpe. Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci. USA*, 99(6):3540–3545, 2002.
- [60] A. Siddiqui and G. Barton. Continuous and discontinuous domains: an algorithm for the automati generation of reliable protein domain definition. *Protein Sci.*, 4:872–884, 1995.
- [61] A. Singh, J. Latombe, and D. Brutlag. A motion planning approach to flexible ligand binding. In *7th Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, pages 252–261, 1999.
- [62] G. Song. *A Motion Planning Approach to Protein Folding*. Ph.D. dissertation, Dept. of Computer Science, Texas A&M University, December 2004.
- [63] G. Song, S. Thomas, K. Dill, J. Scholtz, and N. Amato. A path planning-based study of protein folding with a case study of hairpin formation in protein G and L. In *Proc. Pacific Symposium of Biocomputing (PSB)*, pages 240–251, 2003.
- [64] M. J. Sternberg. *Protein Structure Prediction*. OIRL Press at Oxford University Press, 1996.
- [65] S. Sun, P. D. Thomas, and K. A. Dill. A simple protein folding algorithm using a binary code and secondary structure constraints. *Protein Eng.*, 8(8):769–778, 1995.
- [66] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, and N. M. Amato. Using motion planning to study rna folding kinetics. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 252–261, 2004.
- [67] C. S. Tung, S. C. Harvey, and J. A. McCammon. Large-amplitude bending motions in phenylalanine transfer RNA. *Biopoly.*, 23:2173, 1984.
- [68] H. Vogel. Calmodulin: a versatile calcium mediator protein. *Biochem. Cell Biol.*, 72(9-10):357–376, 1994.
- [69] W. Whiteley. Some matroids from discrete applied geometry. *Contemp. Math.*, 197:171–311, 1996.

- [70] D. Xie, S. Thomas, J.-M. Lien, and N. M. Amato. Incremental map generation. Technical Report TR05-006, Parasol Lab, Dept. of Computer Science, Texas A&M University, Sep 2005.
- [71] M. Zehfus and G. Rose. Compact units in proteins. *Biochemistry*, 25:5759–5765, 1986.
- [72] D. M. Zuckerman. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J. Phys. Chem*, 108:5127–5137, 2004.