

Rigidity Analysis for Protein Motion and Folding Core Identification

Shawna Thomas*, Lydia Tapia[†], Chinwe Ekenna*, Hsin-Yi (Cindy) Yeh*, Nancy M. Amato*

*Parasol Laboratory, Department of Computer Science and Engineering
Texas A&M University, College Station, TX 77843-3112
email: *stthomas, cekenna, hyeh, amato@cse.tamu.edu*

[†]Department of Computer Science
University of New Mexico, Albuquerque, NM 87131-0001
email: *tapia@cs.unm.edu*

Abstract

Protein folding plays an essential role in protein function and stability. Despite the explosion in our knowledge of structural and functional data, our understanding of protein folding is still very limited. In addition, methods such as folding core identification are gaining importance with the increased desire to engineer proteins with particular functions and efficiencies. However, defining the folding core can be challenging for both experiment and simulation.

In this work, we use rigidity analysis to effectively sample and model the protein's energy landscape and identify the folding core. Our results show that rigidity analysis improves the accuracy of our approximate landscape models and produces landscape models that capture the subtle folding differences between protein G and its mutants, NuG1 and NuG2. We then validate our folding core identification against known experimental data and compare to other simulation tools. In addition to correlating well with experiment, our method can suggest other components of structure that have not been identified as part of the core because they were not previously measured experimentally.

Introduction

Protein folding is critical to protein function as its fold largely determines its function and efficiency. Understanding the folding process is paramount to tackling problems such as misfolded disease-causing proteins (e.g., Mad Cow and Alzheimer's) and engineering new proteins to produce particular functions with particular rates/efficiencies (Woodward 1993).

Since it is difficult to experimentally observe molecular motions, computational methods for studying such issues are essential. Traditional computational approaches for generating folding trajectories such as molecular dynamics (Levitt 1983) and Monte Carlo simulation (Covell 1992) are so expensive that they can only be applied to relatively small structures (e.g., proteins with less than 130 amino acids (Zhou et al. 2008)) even when they employ massive computational resources, such as tens of thousands of PCs in the Folding@Home project (Larson et al. 2003). Statistical mechanical models have been applied to compute energy landscape statistics (Muñoz et al. 1998; Alm and Baker 1999;

Muñoz and Eaton 1999). While computationally more efficient, these methods do not produce individual pathway trajectories and are limited to studying global averages.

Identifying the folding core, the subset of a protein's structure first to form during folding and last to break during denaturation, is important in identifying where to make mutations in proteins to yield predictable structure and functional properties. However, both experimental and simulation techniques to capture folding core membership have found it challenging and have often been unsuccessful.

In this paper, we use rigidity analysis (Jacobs 1998) in the context of a robotics-based molecular modeling framework (Amato, Dill, and Song 2003; Tapia, Thomas, and Amato 2010) to

- efficiently sample protein conformations in a more physically realistic way,
- identify protein conformation pairs for connection, and
- simulate hydrogen exchange and identify folding cores.

We show that this method can improve sampling and observe subtle folding differences between protein G and its mutants, NuG1 and NuG2 (Thomas et al. 2007), an important 'benchmark' set developed by the Baker Lab (Nauli, Kuhlman, and Baker 2001). We also compare our folding core predictions to those determined experimentally and to other computational approaches (Hespenheide et al. 2002; Rader and Bahar 2004). We show good correlation to experiment and also indicate that our technique may be useful in suggesting other components of rigid structure for further study that have not yet been experimentally identified.

Related Work

Here we present related work in three areas: robotics-based molecular motion modeling, rigidity analysis, and folding core identification.

PRMs for Protein Folding

In previous work (Amato, Dill, and Song 2003), we introduced a technique for modeling protein folding that is based on the probabilistic roadmap (PRM) approach for motion planning (Kavraki et al. 1996). We applied our method to a large number of structures and validated our secondary structure formation order results against the experimentally determined order.

Our method first constructs a model of the folding landscape by (1) sampling representative conformations (as graph nodes) and (2) connecting them together with sequences of intermediate conformations (as graph edges). This combination of samples and connections forms a graph (or map) that approximates the energy landscape and encodes thousands of folding pathways.

Protein Model. We model the protein as an articulated linkage. Using a standard modeling assumption for proteins that bond angles and bond lengths are fixed (Sternberg 1996), the only degrees of freedom in our model are the backbone’s ϕ and ψ torsional angles which are represented as revolute joints in $[0, 2\pi)$.

Potential Energy Calculation. Our method can use any potential function. Here, we use a coarse potential function similar to (Levitt 1983). It has a van der Waals component where side chains are modeled as spheres, a term favoring known secondary structure through main-chain hydrogen bonds and disulphide bonds, and a term encoding the hydrophobic effect. Previously, we demonstrated that this function produces qualitatively similar results as an all-atoms function for several proteins, including structurally similar mutants, in a fraction of the time. Details can be found in (Song et al. 2003).

Sampling Conformations. Sampling is biased to increase the density near the known native state by starting with the native state and iteratively perturbing it to generate new conformations. In this iterative sampling process, small Gaussian perturbations are applied to existing conformations. Samples are kept based on their energy: low energy samples are kept with a high probability while high energy samples are kept with a low probability.

Sample Connection. We connect neighboring samples by identifying a sequence of feasible intermediate conformations. Typically, connections are attempted to the k -nearest neighbors as identified by some distance metric.

For a given pair of samples, we compute a sequence of intermediate conformations by linear interpolation in ϕ - ψ space. We assign a weight to the transition (connection) to reflect its energetic feasibility. We have found that this scheme works well in practice (Amato, Dill, and Song 2003), and we can use simple graph search algorithms to extract the most energetically feasible pathways in the map. See (Amato, Dill, and Song 2003) for details.

Map Size Determination. Our approximate landscape model’s accuracy depends on the sampling density. To determine the density automatically, we build the map incrementally until the secondary structure formation order along its pathways stabilizes. This is the same technique successfully used in our previous work (Thomas et al. 2007).

Extracting Folding Pathways. We use Map-based Monte Carlo simulation (MMC) (Tapia et al. 2007) to stochastically extract pathways from our landscape model. MMC is similar to traditional Monte Carlo simulation except that it is a walk on our approximate landscape model (i.e., the map) instead of on the complete energy landscape. Applying MMC to our landscapes follows the techniques in (Tapia et al. 2007). We ensure that the likelihood of transitioning between conformations is probabilistically biased

by their Boltzmann transition probabilities. This transition probability is based on the edge weight from constructing the map. For this work, we use 500 MMC pathways, each containing 10,000 path-steps.

Rigidity Analysis

Several computational approaches study protein rigidity and flexibility. Here, we use a rigidity analysis technique called the pebble game (Jacobs and Thorpe 1995; Jacobs 1998) to identify which portions of a particular conformation are rigid, which are independently flexible (i.e., can move without requiring movement of other residues), and which form a dependently flexible set (i.e., can only move in a coordinated motion with other residues). It has been successfully used by several applications to study protein rigidity and flexibility (Jacobs et al. 2001; Rader et al. 2002; Hespenheide et al. 2002; Lei et al. 2004).

Folding Core Identification

Various methods have been introduced to determine the folding core. Hydrogen exchange is the most widely used experimental technique. It identifies which parts of the protein structure are most exposed or protected (Wales and Engen 2006). From this, one can infer which portions fold first and which are last to form, up to the millisecond timescale. However, such methods are complicated, expensive, and cannot be applied to all structures.

Simulation methods abound but have resulted in mispredictions due to limiting assumptions made and models used. Two methods compared against in this work are Floppy Inclusions and Rigid Substructure Topography (FIRST) (Hespenheide et al. 2002) and Gaussian Network Model (GNM) (Rader and Bahar 2004). FIRST uses a full atomic description to identify rigid clusters of residues for a fixed protein conformation. They simulate denaturation by iteratively breaking the weakest hydrogen bond and recomputing the resulting rigid residue clusters, but the actual structure is kept static. They identify the folding core as the set of mutually rigid residues belonging to at least two different secondary structure elements that remain rigid longest. GNM models residues as beads connected by elastic springs representing chain connectivity and bonding. GNM simulations provide slow mode minima and fast mode peaks which are used to identify folding cores. Their simulations are limited to the immediate vicinity of the native state.

Constructing Landscape Models using Rigidity Analysis

We use the same motion framework as described in the related work. However, we incorporate rigidity analysis in the sampling technique and the distance metric to yield more physically-realistic landscape models.

Rigidity Model

For rigidity analysis, we model the protein simply as a chain of rigid bodies, each representing one torsional dof. Bonds

are modeled with increasing strength from hydrophobic contacts (weakest), to hydrogen bonds, to peptide and disulfide bonds (strongest). Details can be found in (Thomas et al. 2007).

Sampling with Rigidity Analysis

Previously, sampling iteratively applied small Gaussian perturbations to the entire conformation. Instead, we use rigidity analysis to focus perturbations on flexible portions. We perturb flexible torsional angles with a high probability, P_{flex} , and rigid torsional angles with a low probability, P_{rigid} . Perturbing rigid torsional angles ensures coverage of the landscape. This sampling technique was originally introduced in (Thomas et al. 2007). We use the same values ($P_{flex} = 0.8$ and $P_{rigid} = 0.2$) as used in that work.

Rigidity Analysis Distance Metric

We also use rigidity analysis to define a new residue mapping and distance metric. A rigidity map, r , is similar to a contact map. Rigid body pairs (i, j) from the rigidity model are marked if they have the same rigidity relationship: 2 if they are in the same rigid set, 1 if they are in the same dependently flexible set, and 0 otherwise. 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 where n is the number of residues:

$$r_{dist}(q_1, q_2) = \sum_{0 \leq i < j \leq 2n} (r_{q_1}(i, j) \neq r_{q_2}(i, j)). \quad (1)$$

We use this distance metric when identifying a conformation's k nearest neighbors for connection. Details can be found in (Thomas et al. 2007).

Folding Core Identification from Rigidity Analysis

We present a new technique based on approximate landscape models and rigidity analysis to simulate relative hydrogen exchange rates. Our method can compute relative exchange rates from any input pathway. We use MMC (described in the related work) to extract multiple pathways, analyze each one individually, and then average the results. We then use these rates to identify folding cores.

Relative Hydrogen Exchange Rates

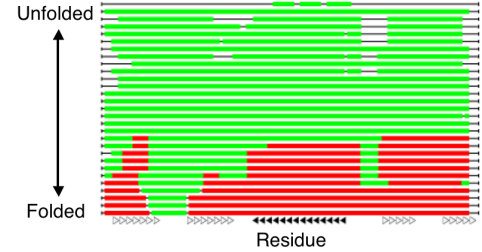
At a given conformation (or path-step), flexible residues are more likely to experience hydrogen exchange than rigid residues. Recall that rigidity analysis (Jacobs 1998) labels every residue at every path-step c along an input pathway p as rigid, independently flexible, or dependently flexible. We assign each residue at every path-step a score based on its rigidity classification. For a residue i , we define its *rigidity score*, $RS(p, i)$, for a particular pathway p , as the average of its rigidity scores along p .

To compare the rigidity scores to experimental data, we define the relative exchange rate $ex_{RS}(p, i)$ for residue i along a pathway p , as

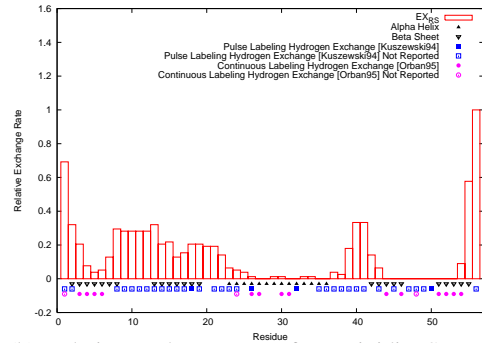
$$ex_{RS}(p, i) = 1 - \frac{RS(p, i) - RS_{min}(p)}{RS_{max}(p) - RS_{min}(p)} \quad (2)$$

where $RS_{min}(p)$ is the smallest rigidity score obtained over p for all residues and $RS_{max}(p)$ is the largest. We normalize the scores because in some cases residues do not become completely rigid or completely flexible, e.g., some residues in the native state are still flexible. Note that in most cases this does not occur. For all the proteins studied, we found that the minimum/maximum encountered score along individual pathways deviated from the minimum/maximum possible in only 2 cases resulting in an average variance of 0.16%. For an input set of pathways P , we define the average relative exchange rate, $EX_{RS}(i)$, for residue i as the average of the relative exchange rates, $ex_{RS}(p, i)$, for all pathways $p \in P$.

Figure 1(a) shows the rigidity analysis along an example pathway for Protein G (a 56 residue protein with a central α -helix flanked by two β -hairpin turns), and Figure 1(b) shows the corresponding relative exchange rates. The second β -hairpin is experimentally known to form before the first β -hairpin (Li and Woodward 1999; McCallister, Alm, and Baker 2000). This behavior is reflected in both the rigidity scores and the relative exchange rates: β -hairpin 2 remains rigid longer than β -hairpin 1 along the pathway in (a), and its corresponding relative exchange rates (b) are lower.



(a) Rigidity Analysis Results at each Path-Step: rigid (red), dependently flexible (green), or independently flexible (not colored)



(b) Relative Exchange Rates from Rigidity Scores

Figure 1: Example unfolding pathway for Protein G. Note that this pathway was extracted from a shortest path search for demonstration purposes. Pathways used in the results are extracted stochastically using MMC and contain 10,000 path-steps with both folding and unfolding events. α -helices (filled triangles) and β -sheets (empty triangles) are indicated along the bottom for reference.

Folding Core Identification

We can infer the folding core from the relative exchange rates. Given the relative exchange rates for a set of residues, there exists some threshold t that defines the most stable residues (i.e., the folding core). This definition is frequently applied to experimental exchange rates in order to identify the folding core (Li and Woodward 1999).

There are many ways to define a threshold t . For experimental data, this is at the discretion of the authors and varies widely. In this work, we determine t automatically using k -means clustering (Jain and Dubes 1988). To determine an appropriate value for k , we use a commonly known technique as the elbow criterion (Lieu and Saito 2007). It selects a k such that increasing k does not add sufficient information. The threshold t is then defined as the minimum value splitting any two of the k clusters.

Results

Here we show the benefits of rigidity analysis in our protein motion framework, namely, improved sampling, better ability to distinguish folding behaviors between structurally similar proteins, and folding core identification.

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 built roadmaps for several previously studied proteins (Amato, Dill, and Song 2003). For each protein, we compare rigidity-based sampling with automatic map size determination to our previous sampling technique with fixed sampling density. Both give the same secondary structure formation order distribution and agree with experiment when available.

Figure 2(a,b) shows the relative performance of the two methods in terms of (a) number of samples needed and (b) connectivity. Each data point corresponds to a different protein studied. Rigidity-based sampling both reduces roadmap size and increases roadmap connectivity. In addition, the performance gains are not dependent on protein length (c,d).

Case study of proteins G, L NuG1, and NuG2

Proteins G, L, and mutants of protein G, NuG1 and NuG2 (Nauli, Kuhlman, and Baker 2001), present a good test case for our technique because they are known to fold differently despite having similar structure. All proteins are composed of a central α -helix and a 4-stranded β -sheet. Hydrogen exchange experiments indicate that β 1-2 forms first in protein L, and β 3-4 forms first in protein G (Li and Woodward 1999). In (Nauli, Kuhlman, and Baker 2001), protein G is mutated in both hairpins to increase the stability of β 1-2 and decrease the stability of β 3-4. Φ -value analysis indicates that the hairpin formation order for both is switched.

Our previous sampling strategy was able to capture the folding differences between proteins G and L, but not between protein G and NuG1 or NuG2. Our rigidity-based sampling and analysis is able to also capture the correct folding behavior of NuG1 and NuG2, see Table 1.

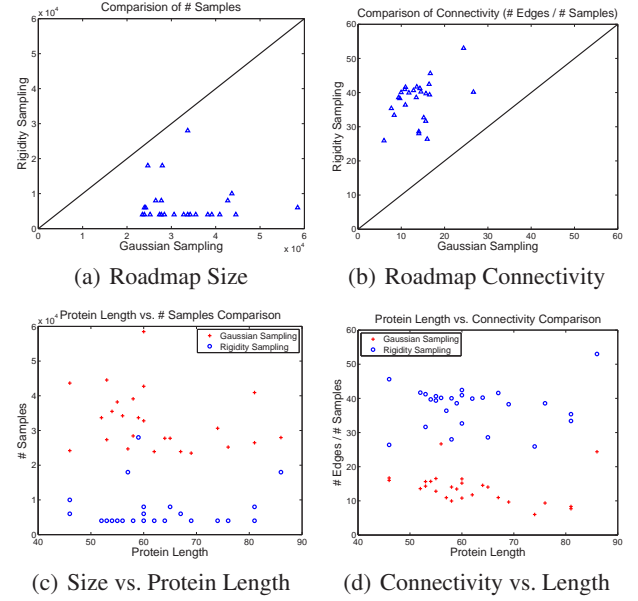


Figure 2: Rigidity-based sampling produces smaller roadmaps (a) with increased connectivity (b) vs. previous work. Performance gains are not dependent on protein length (c,d).

Folding Core Identification

We studied 21 different proteins of varying size and structure from 54–155 residues, see (Thomas, Tapia, and Amato 2008) for a complete listing. Because a folding core threshold is not universally agreed upon, we instead calculate an appropriate threshold using the same method described earlier. If numerical data is not provided, we use the labeling suggested by the authors.

Figure 3 provides a visual comparison of simulated exchange rates to available experimental data on the 3D structure for some of the proteins (see (Thomas, Tapia, and Amato 2008) for a complete listing). A strength of our simulation is that we can compute relative exchange rates for every residue in the structure while experiments are limited to

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

Table 1: Comparison of secondary structure formation orders for proteins G, L, NuG1, and NuG2 with known experimental results: ^a (Li and Woodward 1999), ^b (Li and Woodward 1999), and ^c (Nauli, Kuhlman, and Baker 2001). Brackets indicate no clear order. Only formation orders greater than 1% are shown.

which residues they can accurately probe.

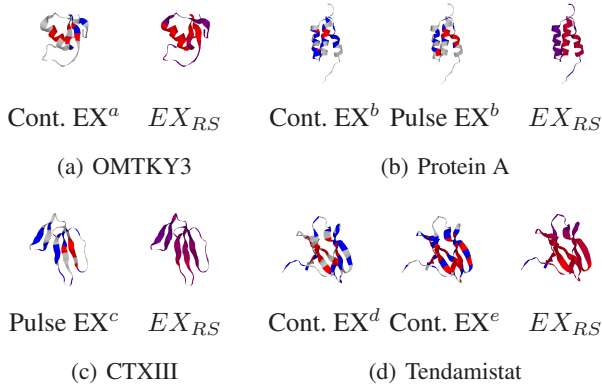


Figure 3: Comparison of simulated exchange rates to experiment. For experimental data, red residues are inside the core, blue residues are outside the core, and grey residues were not measured. For simulated exchange rate data, residues are shaded from fastest/blue to slowest/red. ^a(Arrington, Teesch, and Robertson 1999). ^b(Bai et al. 1997). ^c(Sivaraman et al. 1998). ^d(Qiwen, Kline, and Wuthrich 1987). ^e(Schonbrunner et al. 1996).

We compare our method to 4 other computational techniques for folding core identification: slow mode minimas (GNM-G), fast mode peaks (GNM-H) (Rader and Bahar 2004), and FIRST (Hespenheide et al. 2002) with two different hydrophobic tether definitions: H3 (the default and most restrictive) and H1 (the least restrictive). We compare the *sensitivity* and *specificity* of the different identification techniques. *Sensitivity* is the ratio of the number of residues accurately labeled as in the folding core by simulation to the number of residues labeled as in the folding core by experimental data. *Specificity* is the ratio of the number of residues accurately labeled as out of the folding core by simulation to the number of residues labeled as out of the folding core by experimental data. We only examine residues that were measured by experimental data which could be as little as 14% of the protein.

Table 2 summarizes the overall statistics for each method. The error is calculated as the normalized distance from perfect sensitivity and specificity. The best performing method in each category is indicated in boldface. Due in part to the large noise present in the experimental data, missing measurements, and labeling convention inconsistencies between data sets, all computational methods exhibit large variances in sensitivity and specificity, ranging from 0.025 to 0.345.

Our method performs better than FIRST and GNM-based methods in sensitivity, between GNM and FIRST in specificity, and better than FIRST and similar to GNM-G in error. While GNM-H has the best specificity, it also has the worst sensitivity. This tradeoff is a common property of learning methods. Table 2 also shows how the labeled folding core size correlates to sensitivity and specificity. Methods that “under-guess” tend to have lower sensitivities and higher specificities, and methods that “over-guess” tend to have the opposite. GNM-H predicts the smallest folding cores and also has the lowest (highest) sensitivity (specificity). GNM-

Method	Sensitivity	Specificity	Error	% in Core
Over All Data Sets				
EX_{RS}	0.545	0.671	0.482	29.8
GNM-G	0.432	0.715	0.474	28.5
GNM-H	0.302	0.846	0.525	15.4
FIRST-H3	0.501	0.543	0.551	38.6
FIRST-H1	0.518	0.573	0.521	40.7
Over the 5 Most Complete Data Sets				
EX_{RS}	0.631	0.627	0.420	43.5
GNM-G	0.189	0.815	0.591	21.1
GNM-H	0.223	0.927	0.554	15.3
FIRST-H3	0.628	0.612	0.384	30.0
FIRST-H1	0.572	0.709	0.373	39.0

Table 2: Summary of folding core identification performance. Error is the normalized distance to perfect sensitivity and specificity. The best performance in each category is in boldface.

H is overly conservative and thus greatly sacrifices sensitivity by labeling large portions of the protein as outside the core. FIRST, with the largest folding cores, has moderate sensitivity and the lowest specificity.

We believe the low sensitivities and specificities for all methods are caused largely by varying experimental conditions and missing experimental data. On average, less than half of the protein was measured experimentally with many below 25%. This imposes a bias to the labeling. We found that by considering only experiments with greater than 80% measured (the five most complete) significantly reduced the variance for all methods across all metrics (see Table 2).

Conclusion

We describe how to use rigidity analysis to enhance an existing robotics-based protein motion framework and to identify protein folding cores from simulated exchange rates. Rigidity analysis greatly reduces the size of the approximate landscape model needed and improves the connectivity of this model over previous work. These improvements are not dependent on protein length, enabling the study of larger structures in the future. The folding core predictions from our approximate landscape models and rigidity analysis are more accurate than other existing computational approaches. We believe the real use of this technique will be to aid researchers by providing an indication of fast and slowly exchanging residues to target for protein design.

Acknowledgments

This research supported in part by NSF awards CNS-0551685, CCF-0833199, CCF-0830753, IIS-0917266, IIS-0916053, EFRI-1240483, RI-1217991, by NSF/DNDO award 2008-DN-077-ARI018-02, by NIH NCI R25 CA090301-11, by DOE awards DE-FC52-08NA28616, DE-AC02-06CH11357, B575363, B575366, by THECB NHARP award 000512-0097-2009, by Samsung, Chevron, IBM, Intel, Oracle/Sun and by Award KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST). Tapia supported in part by the National Institutes of Health (NIH) Grant P20RR018754 supporting the Center for

Evolutionary and Theoretical Immunology. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Alm, E., and Baker, D. 1999. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA* 96(20):11305–11310.
- Amato, N. M.; Dill, K. A.; and Song, G. 2003. Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures. *J. Comput. Biol.* 10(3-4):239–255. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2002.
- Arrington, C. B.; Teesch, L. M.; and Robertson, A. D. 1999. Defining protein ensembles with native-state NH exchange: Kinetics of interconversion and cooperative units from combined NMR and MS analysis. *J. Mol. Biol.* 285:1265–1275.
- Bai, Y.; Karimi, A.; Dyson, H. J.; and Wright, P. E. 1997. Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci.* 6(7):1449–1457.
- Covell, D. 1992. Folding protein α -carbon chains into compact forms by Monte Carlo methods. *Proteins: Struct. Funct. Genet.* 14(4):409–420.
- Hespenheide, B. M.; Rader, A.; Thorpe, M.; and Kuhn, L. A. 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Gra. Model.* 21:195–207.
- Jacobs, D., and Thorpe, M. 1995. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.* 75(22):4051–4054.
- Jacobs, D. J.; Rader, A.; Kuhn, L. A.; and Thorpe, M. 2001. Protein flexibility predictions using graph theory. *Proteins Struct. Funct. Genet.* 44:150–165.
- Jacobs, D. 1998. Generic rigidity in three-dimensional bond-bending networks. *J. Phys. A: Math. Gen.* 31:6653–6668.
- Jain, A. K., and Dubes, R. C. 1988. *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall.
- Kavraki, L. E.; Švestka, P.; Latombe, J. C.; and Overmars, M. H. 1996. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans. Robot. Automat.* 12(4):566–580.
- Larson, S.; Snow, C.; Shirts, M.; and Pande, V. 2003. Folding@home and genome@home: Using distributed computing to tackle previously intractable problems in computational biology. *Computational Genomics*. To appear.
- Lei, M.; Zavodszky, M. I.; Kuhn, L. A.; and Thorpe, M. F. 2004. Sampling protein conformations and pathways. *J. Comput. Chem.* 25:1133–1148.
- Levitt, M. 1983. Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.* 170:723–764.
- Li, R., and Woodward, C. 1999. The hydrogen exchange core and protein folding. *Protein Sci.* 8(8):1571–1591.
- Lieu, L., and Saito, N. 2007. Automated discrimination of shapes in high dimensions. In *Proc. Society of Photo-Optical Instrumentation Engineers (SPIE)*, volume 6701, 67011V–1–67011V–12.
- McCallister, E. L.; Alm, E.; and Baker, D. 2000. Critical role of β -hairpin formation in protein G folding. *Nat. Struct. Biol.* 7(8):669–673.
- Muñoz, V., and Eaton, W. A. 1999. A simple model for calculating the kinetics of protein folding from three dimensional structures. *Proc. Natl. Acad. Sci. USA* 96(20):11311–11316.
- Muñoz, V.; Henry, E. R.; Hoferichter, J.; and Eaton, W. A. 1998. A statistical mechanical model for β -hairpin kinetics. *Proc. Natl. Acad. Sci. USA* 95:5872–5879.
- Nauli, S.; Kuhlman, B.; and Baker, D. 2001. Computer-based re-design of a protein folding pathway. *Nature Struct. Biol.* 8(7):602–605.
- Qiwen, W.; Kline, A. D.; and Wuthrich, K. 1987. Amide proton exchange in the α -amylase polypeptide inhibitor tendamistat studied by two-dimensional ^1H nuclear magnetic resonance. *Biochemistry* 26:6488–6493.
- Rader, A., and Bahar, I. 2004. Folding core predictions from network models of proteins. *Polymer* 45:659–668.
- Rader, A.; Hespenheide, B. M.; Kuhn, L. A.; and Thorpe, M. 2002. Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci. USA* 99(6):3540–3545.
- Schonbrunner, N.; Wey, J.; Engles, J.; Georg, H.; and Kiefhaber, T. 1996. Native-like β -structure in a trifluoroethanol-induced partially folded state of the all- β -sheet protein tendamistat. *J. Mol. Biol.* 260:432–445.
- Sivaraman, T.; Kumar, T. K. S.; Chang, D. K.; Lin, W. Y.; and Yu, C. 1998. Events in the kinetic folding pathway of a small, all β -sheet protein. *J. Biol. Chem.* 273(17):10181–10189.
- Song, G.; Thomas, S.; Dill, K.; Scholtz, J.; and Amato, N. 2003. 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)*, 240–251.
- Sternberg, M. J. 1996. *Protein Structure Prediction*. OIRL Press at Oxford University Press.
- Tapia, L.; Tang, X.; Thomas, S.; and Amato, N. M. 2007. Kinetics analysis methods for approximate folding landscapes. *Bioinformatics* 23(13):539–548. Special issue of Int. Conf. on Intelligent Systems for Molecular Biology (ISMB) & European Conf. on Computational Biology (ECCB) 2007.
- Tapia, L.; Thomas, S.; and Amato, N. M. 2010. A motion planning approach to studying molecular motions. *Communications in Information and Systems* 10(1):53–68. special issue in honor of Michael Waterman.
- Thomas, S.; Tang, X.; Tapia, L.; and Amato, N. M. 2007. Simulating protein motions with rigidity analysis. *J. Comput. Biol.* 14(6):839–855. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2006.
- Thomas, S.; Tapia, L.; and Amato, N. M. 2008. Protein folding core identification from rigidity analysis and motion planning. Technical Report TR08-001, Parasol Lab, Dept. of Computer Science, Texas A&M University.
- Wales, T. E., and Engen, J. R. 2006. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spec. Rev.* 25(1):158–170.
- Woodward, C. 1993. Is the slow exchange core the protein folding core? *Trends Biochem. Sci.* 18:359–360.
- Zhou, R.; Eleftheriou, M.; Hon, C.-C.; Germain, R. S.; Royyuru, A. K.; and Berne, B. J. 2008. Massively parallel molecular dynamics simulations of lysozyme unfolding. *IBM J. Res. & Dev.* 52(1/2):19–30.