Evaluating Roadmap Sampling Methods for High Dimensional Robots and Proteins †

Andrea Howells[1], Rachel Webster[2], Lydia Tapia[3]
[1]State University of New York at Plattsburgh, Plattsburgh, NY 12901
[2]Lewis & Clark College, Portland, OR 97219
[3]Computer Science Department, University of New Mexico, Albuquerque, NM 87131
ahowe003@plattsburgh.edu, tapia@cs.unm.edu, rwebster@lclark.edu

†This work supported in part by the National Institutes of Health (NIH) Grant P50GM085273 supporting the New Mexico Spatiotemporal Modeling Center and NIH Grant P20RR018754 supporting the Center for Evolutionary and Theoretical Immunology. Howells and Webster supported by the Computing Research Association CRA-W Distributed Research Experiences for Undergraduates.

1 Introduction

In studies of AIDS progression, the presence of a variant of human leukocyte antigen (HLA) (the human type of major histocompatability complex [MHC]) that binds to natural killer cells has been associated with slow AIDS progression [19]. Recent experimental research has shown that protein motion influences binding affinity, which is thought to contribute to disease resistance. By studying protein motion, we hope to be able to develop vaccinations for AIDS and other diseases.

Most previous coputational research into the prediction of binding affinities has only considered static representations of protein structure. Current probabilistic roadmap method (PRM) software developed for robotic motion planning needed to be improved and modified in order to map protein folding pathways. The methods were modified to build a graph corresponding to an approximate map of the molecule's energy landscape that encodes many (typically thousands) of folding pathways.

The pathways can capture the protein folding properties, validated through experiments, such as secondary structure formation order, subtle folding differences, and relative folding rates. Both robot bodies with articulated links and multiple rigid body structures were used to simulate proteins.

Data was collected and studied to determine appropriate values to use for roadmap generation of the proteins. Simulations of multiple proteins were also performed. By modeling and studying these protein molecules, we hope to, in the future, contribute to the development of treatments and vaccinations for viruses.

The proteins Trimeric Foldon of the T4 phagehead fibritin (1RFO), Immunoglobulin E (IgE), and 2, 4-Dinitrophenol (DNP) were modeled in Maya and Vizmo in order to allow for future study of the of the ligand-receptor interactions.

2 Related Work

2.1 Feature Sensitive Motion Planning

Motion planning environments can be drastically different. Also, no single motion planning method can outperform other methods for all problems/environments. Each method has strengths or weaknesses that make it better for given types of environments. In the paper "A Machine Learning Approach for Feature-Sensitive Motion Planning", an automated framework was proposed [21].

In this method, the characteristics of the space are used to decide what methods to apply. First, planning space is subdivided by dividing each of the n DOFs (degrees of freedom) into m equal sized parts. The authors believe that C-space subdivision can be used to identify regions suitable for a particular motion planner. Next, the planning space is evaluated to determine if a particular method could be applied. In this work, an automated process for feature-sensitive motion planning incorporates multiple methods and determines which to use based on the environment the robot must traverse. This approach was shown to outperform the other planners [21].

2.2 Protein Motion Planning

The ability to determine binding affinities is critical to understanding the specificity of immune responses and therefore vaccine research. Most previous computational research into the prediction of binding affinities has only considered static representations of protein structure. However, recent experimental research has shown that protein motion is critical in peptide-MHC-TCR (peptide-major histocompatability complex T-cell receptor) interactions. While experimental research has been successful at investigating a limited number of structures, experiments have drawbacks. They are time-consuming and expensive, and it is impossible to test the tens thousands of protein variants characterized in viruses such as HIV, influenza, and dengue.

PRM-based approaches have been applied to several molecular domains [1, 2, 5, 22, 23, 24, 26, 32]. Singh, Latombe and Brutlag first applied PRMs to protein/ligand binding [22]. Another PRM variant later explored this problem with additional success [5]. We have applied PRMs to model protein folding pathways [30, 33, 31] and RNA folding kinetics [27, 28]. PRMs have also been used by other groups to study molecular motions [3, 4, 6, 7].

2.3 Ligand-Receptor Interactions

Our immune systems are governed in part by the formation of antigen-antibody complexes. A range of critical functions, such as phagocytosis, antibody-dependent cytotoxicity, and control of antibody and cell secretion, are the result of interactions of these complexes with many cells of the immune system. The binding of immunoglobulin Fc domains to Fc receptors (FcRs) mediates these interactions. The FcR for Immunoglobulin E (designated FcϵR) is known to have a role in allergic reactions [14]. FcϵR binds IgE in a 1:1 ratio with a high avidity and a long lifetime [14, 15] and clustering of (designated FcϵR) results in rapid release of histamine and brings about other allergic reactions [20]. By studying models of these clusters, we will have a better understanding of the allergic reaction and be poised to develop methods to control allergic reactions.

3 General Methods

3.1 Probabilistic Roadmap Methods (PRM) for Protein Folding

.

PRMs work by building a graph corresponding to an *approximate map* of the molecules energy landscape that encodes many (typically thousands) folding pathways. Our PRM-based method follows the general PRM paradigm (Figure 1). First, conformations (roadmap nodes) are sampled from the molecules conformation space (C-space). Next, transitions between nearby conformations are encoded as roadmap edges. As in nature, our strategy favors low energy conformations and transitions. In particular, during the sampling phase, lower energy samples have a higher retention probability, and during the node connection phase, each connection is assigned a weight to reflect its energetic feasibility. The energetic feasibility of a transition is determined by the energies of all the intermediate conformations along the transition. Thus, least weight paths in the roadmap correspond to the most energetically feasible paths, and roadmaps encode thousands of feasible pathways.

Path Extraction. PRM roadmaps encode thousands of pathways. As in traditional PRM applications, a path between a start conformation and a goal conformation is extracted by first connecting them to the roadmap and searching for the smallest weight path between them. However, unlike traditional PRM applications which frequently aim to find *any* path, molecular applications are concerned with the *quality* of extracted paths. Typically in molecular applications, path quality is indirectly enforced through the edge weight function. Very poor transitions are either given a low probability/high weight or are removed from the roadmap all together.

Energy function: We have used both a coarse energy function similar to [18] and all atom energy models [13, 16, 35]. For the coarse model, we use a step function approximation of the van der Waals component and model all side chains as equal radii spheres with zero DOF. If two spheres are too close (e.g., their centers are < 2.4Å during sampling and < 1.0Å during connection), a very high potential is returned and the node or connection will be discarded. Otherwise the potential is:

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

where K_d is 100 kJ/mol and $d_0 = d_c = 2$Å as in [18]. 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 (E_{hp}) is computed as follows: if two hydrophobic residues are within 6Å of each other, then the potential is decreased by 20 kJ/mol. Full details can be found in [2].

Biased Sampling: As previously discussed, samples are retained based on their energy. In our protein work, a sample q, with potential energy E_q, is accepted with probability:

$$Prob(accept\ q) = \begin{cases} 1 & if E_q < E_{min} \\ \frac{E_{max} - E_q}{E_{max} - E_{min}} & if E_{min} \leq E_q \leq E_{max} \\ 0 & if E_q > E_{max} \end{cases} \tag{2}$$

where E_{min} is the potential energy of the open chain and $|E_{max}$ is $2E_{min}$. The roadmap produced by our technique is an approximation of the protein's energy landscape, and the quality is dependent on the

sampling strategy. Generally, we are most interested in regions near the target conformation(s) and so concentrate sampling there. In earlier work [1, 2, 23, 24], a denser distribution of samples near the target conformation was obtained through an iterative process where we apply small Gaussian perturbations to existing conformations, beginning with the target conformation. This approach works fairly well, but still required many samples (e.g., 10,000) for relatively small proteins (e.g., 60-100 residues). In [33], we used rigidity analysis [9, 10, 11, 12, 17] to determine which portions of the protein to perturb which increased the protein size we can handle. Rigidity analysis identifies the rigid and flexible parts of a protein, which are then perturbed according to their flexibility. This helps to provide a physically realistic way to perturb conformations. The parameters that were adjusted were the probability that a portion of the structure that is labeled as flexible is perturbed, also known as PFlex, and the probability that a portion of the structure that is labeled as rigid is perturbed, known as PRigid. Where the nodes were produced was also discerned. The layers determine how folded the protein is, with layer 9 being the native state and completely folded, and layer 0 the most unfolded state. Since each layer has a different energy level, ideally, the nodes will be distributed across many layers, and just not focused around the native state. The better the distribution of nodes across multiple layers, the more of the entire folding landscape is produced.

Connection. For each node in the roadmap, we attempt to connect it with its k nearest neighbors with a straight-line in the proteins conformation space. The weight for the edge | (*q1, q2*) is a function of the intermediate conformations along the edge $q_1 = c_0, c_1, ..., c_{n-1}, c_n = q_2$. 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 $\Delta E_i = Ec_{i+1} - E(c_i)$:

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

This keeps the detailed balance between two adjacent states, and enables the weight of an edge to be computed by summing the negative logarithms of the probabilities for consecutive pairs of conformations in the sequence. (Negative logs are used since each $0 \leq P_i \leq 1$.) In recent work, the height of the energy barrier between q_1 and q_2 was used along with atomically-detailed transitions (edges) [29].

Protein Model: In the preliminary results shown in this section, we model the protein as an articulated linkage.

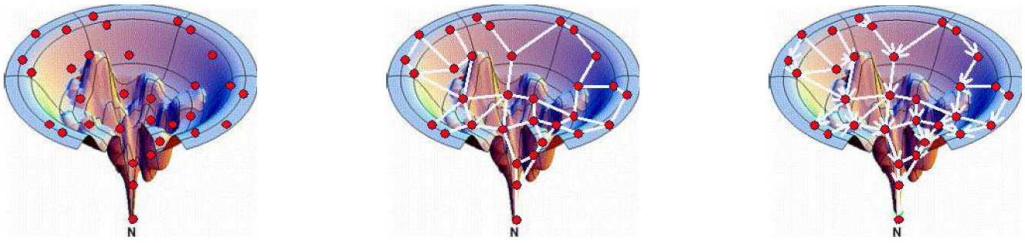


Figure 1: A PRM roadmap for molecular folding shown imposed on a visualization of the molecule's energy landscape: (a) after node generation (note sampling is denser around **N**, the known native structure), (b) after the connection phase, and (c) using it to extract folding paths to the known native structure.

Using a standard modeling assumption for proteins that bond angles and bone lengths are fixed[25], the only degrees of freedom (DOF) in our model are the backbone's phi and psi torsional angles which are modeled as joints with values [0,2π) [29].

4 Application: Parameter Study

Testing was performed on Quad Core Optiplex 980 machines. One to three tests were performed simultaneously. Some of the tests were repeated individually to verify results.

We evaluated three large proteins; TCR (T-cell receptors), MHC (Major histocompatability complex), and 1QLP (1 Alpha Antitrypsin). TCR and MHC were chosen as protein motion is critical in peptide-MHC-TCR interactions (as previously stated). 1QLP was chosen as it is a large protein and has interesting properties as a member of the serpin family of serine proteinase inhibitors, which play important roles in the inflammatory, coagulation, fibrinolytic, and complement cascade[8].

When the parameters PFlex (the probability that a portion of the structure that is labeled as flexible is perturbed), PRigid (the probability that a portion of the structure labeled as rigid is perturbed), and standard angle (the amount by which the angles indicated as PFlex or PRigid are changed) were varied, we found smaller PFlex and PRigid and larger standard angle settings produced more nodes and edges in more layers of the protein energy landscape for all three of these proteins. The layers represent the unfolding states of the protein. The largest number represented is the native state of the protein. Smaller PFlex and PRigid values and larger standard angle settings also resulted in a longer time to generate the landscape, so a tradeoff may be needed in order to produce acceptable landscapes in acceptable time frames.

For the first tests for TCR, the number of samples was set at 100. PFlex was kept between 0.1-0.06, and PRigid was kept between 0.05-0.0005, as shown in Figure 2. The lower the PFlex and PRigid values, the lower the number of rejections that occurred at each standard angle, also seen in Figure 2. The distribution of the nodes across the layers is also shown in Figure 3, where layer 9 is the native state. When PFlex = 0.06 and PRigid = 0.0005, the number of rejections are at their lowest, and the nodes are distributed across 8 layers. When PFlex = 0.1 and PRigid = 0.05 or 0.0005, the nodes are also distributed across 8 layers, but the number of rejections is much higher [34].

For the second tests for TCR, the number of nodes generated was set at 100. PFlex was kept between 0.1-0.06 and PRigid was kept between 0.05-0.0005. The distribution of nodes across the layers is shown in Figure 5. When PFlex = 0.06 and PRigid = 0.0005, the number of rejections are the lowest, however the distribution of nodes is not as diverse, which means the nodes are not at their highest quality. When PFlex = 0.06 and PRigid = 0.05, the number of rejections is fairly low, and the nodes are highly distributed [34].

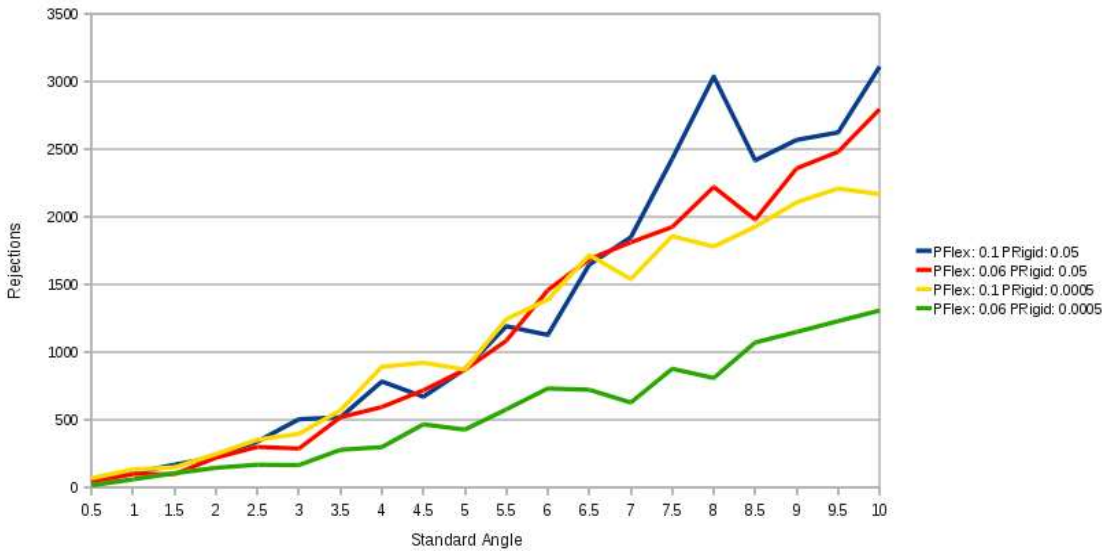
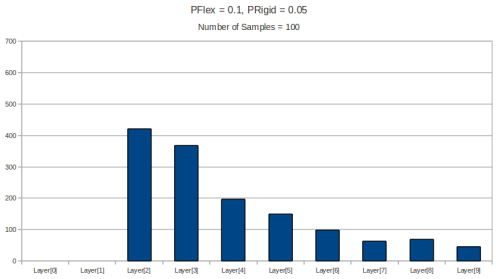
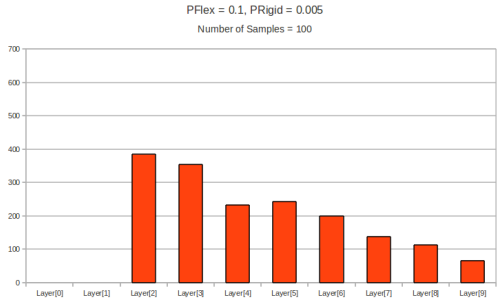


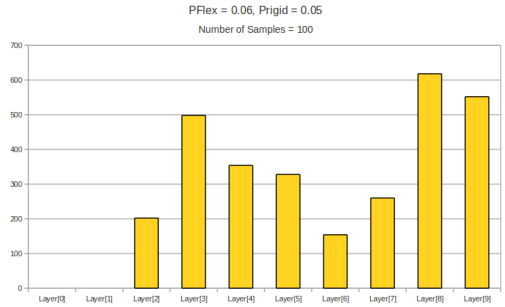
Figure 2: The number of samples = 100 for TCR. The number of rejections is graphed against the standard angle. PFlex is varied between 0.1 - 0.006, and PRigid is varied between 0.05 - 0.0005.



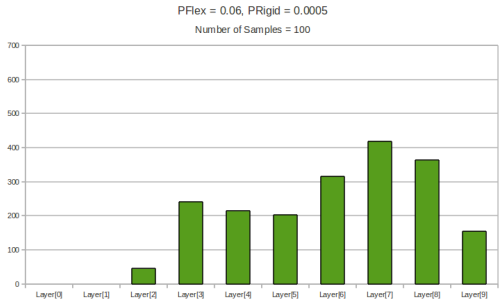
(a) PFlex = 0.1, PRigid = 0.05



(b) PFlex = 0.1, PRigid = 0.0005

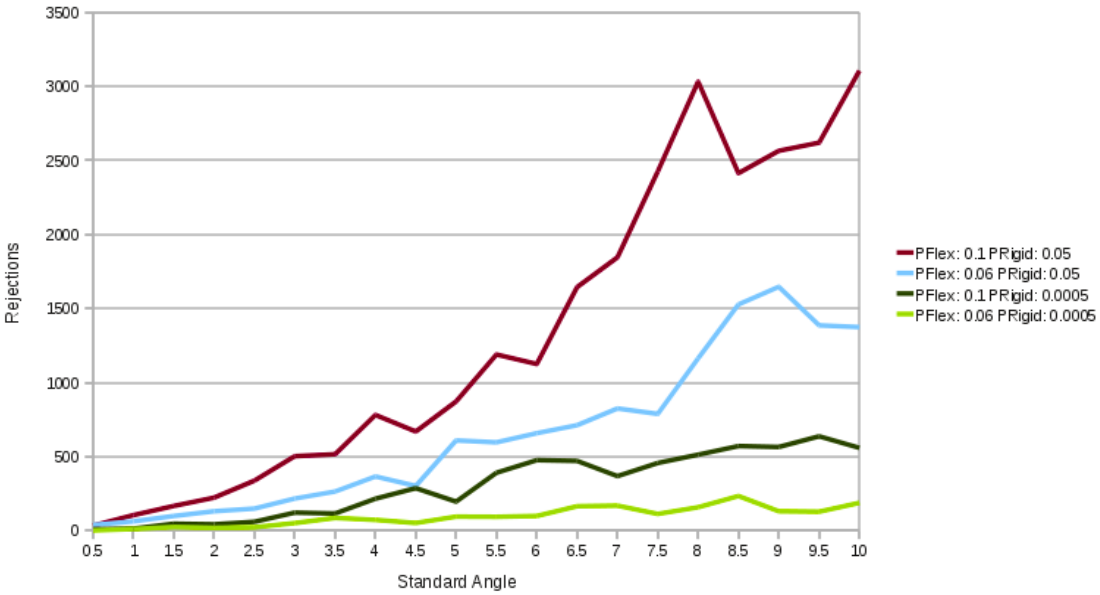


(c) PFlex = 0.06, PRigid = 0.05



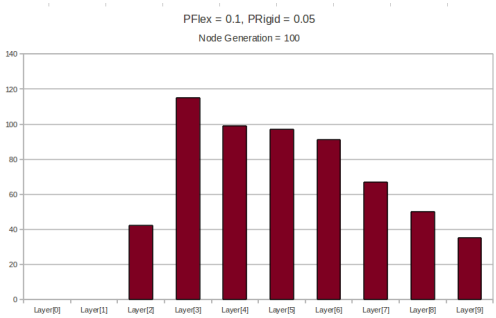
(d) PFlex = 0.06, PRigid = 0.0005

Figure 3: TCR: The number of nodes produced in each layer for every run is graphed. In (a), the number of nodes generated in every layer is graphed for PFlex = 0.1 and PRigid = 0.05. In (b), the number of nodes generated in every layer is graphed for PFlex = 0.1 and PRigid = 0.0005. In (c), the number of nodes generated in every layer is graphed for PFlex = 0.06 and PRigid = 0.05. In (d), the number of nodes generated in every layer is graphed for PFlex = 0.06 and PRigid = 0.0005.

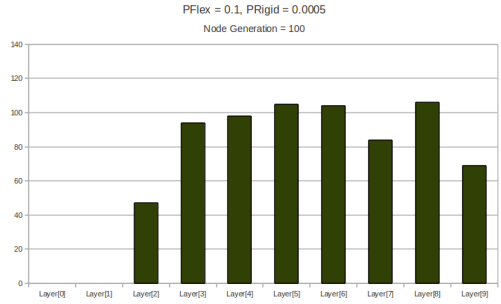


TCR: Node Generation = 100

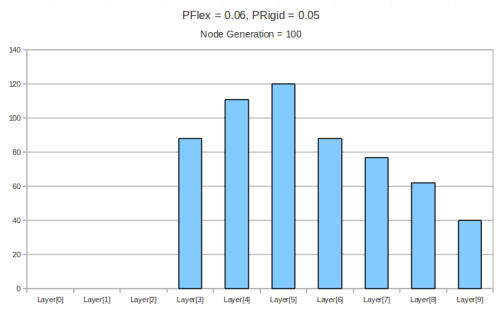
Figure 4: TCR: The number of nodes generated = 100 for TCR. PFlex is varied between 0.1 - 0.006, and PRigid is varied between 0.05 - 0.0005.



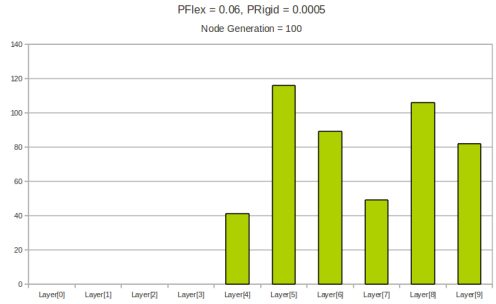
(b)PFlex = 0.1, PRigid = 0.05



(c)PFlex = 0.1, PRigid = 0.0005



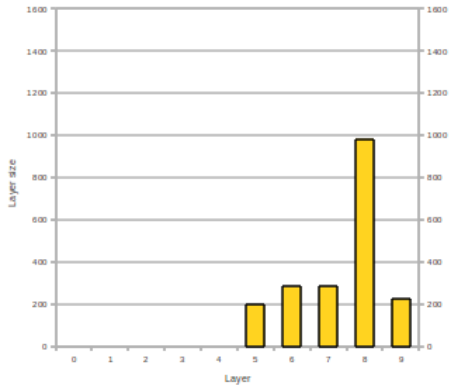
(d) PFlex = 0.06, PRigid = 0.05



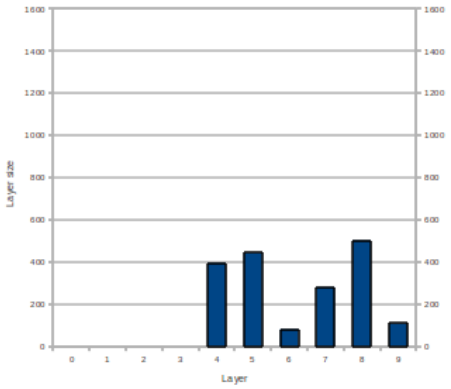
(d)PFlex = 0.06, PRigid = 0.0005

Figure 5: TCR: The number of nodes produced in each layer for every run is graphed. In (a), the number of nodes generated in every layer is graphed for PFlex = 0.1 and PRigid = 0.05. In (b), the number of nodes generated in every layer is graphed for PFlex = 0.1 and PRigid = 0.0005. In (c), the number of nodes generated in every layer for PFlex = 0.06 and PRigid = 0.05. In (d), the number of nodes generated in every layer is graphed for PFlex = 0.06 and PRigid = 0.0005.

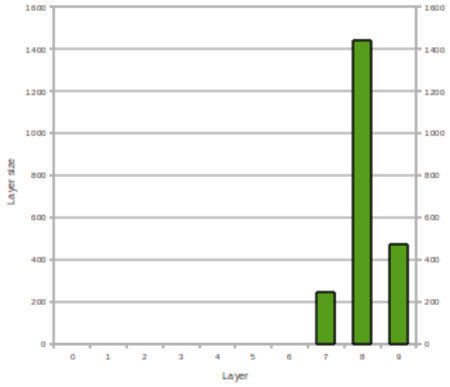
Figures 6 and 7 show the data collected using MHC with number samples set to 100 per iteration. Figure 6 shows that the best node populations and distribution over layers occurred when the PRigid and PFlex values were set to 0.05 and 0.1(a), respectively or to 0.0005 and 0.1 (c) respectively. 7 shows the best cost (i.e; fewest rejected nodes) occurred when the PRigid and PFlex values were set to 0.05 and 0.1 (c), respectively or to 0.0005 and 0.06 respectively (d).



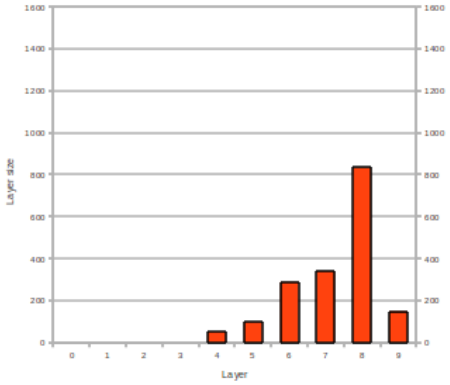
(a): PRigid:0.05 PFlex:0.1



(b): PRigid:0.05 PFlex:0.06



(c): PRigid:0.0005 PFlex:0.1



(d): PRigid:0.0005 PFlex:0.06

Figure 6: MHC (number samples=100)

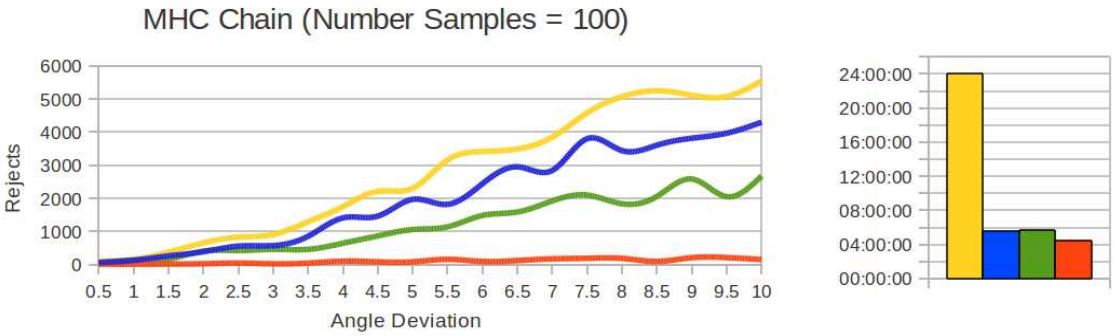


Figure 7: MHC (number samples=100)
Yellow (a): PRigid:0.05 PFlex:0.1 Blue (b): PRigid:0.05 PFlex:0.06
Green (c): PRigid:0.0005 PFlex:0.1 Orange (d): PRigid:0.0005 PFlex:0.06

Figures 8 and 9 show the data collected using MHC with number of nodes generated set to 100 per iteration. Figure 8 shows that the best node populations and distribution over layers occurred for the parameters used in (a) PRigid 0.05 and PFlex 0.08, (b) PRigid 0.05 and PFlex 0.1, (c) PRigid 0.05 and PFlex 0.06, and (d) PRigid 0.0005 and PFlex 0.01. Figure 9 shows the best cost (i.e.; fewest rejected nodes) occurred for the parameters used in (b) PRigid 0.05 and PFlex 0.1, (c) PRigid 0.05 and PFlex 0.06, or (d) PRigid 0.0005 and PFlex 0.01.

C

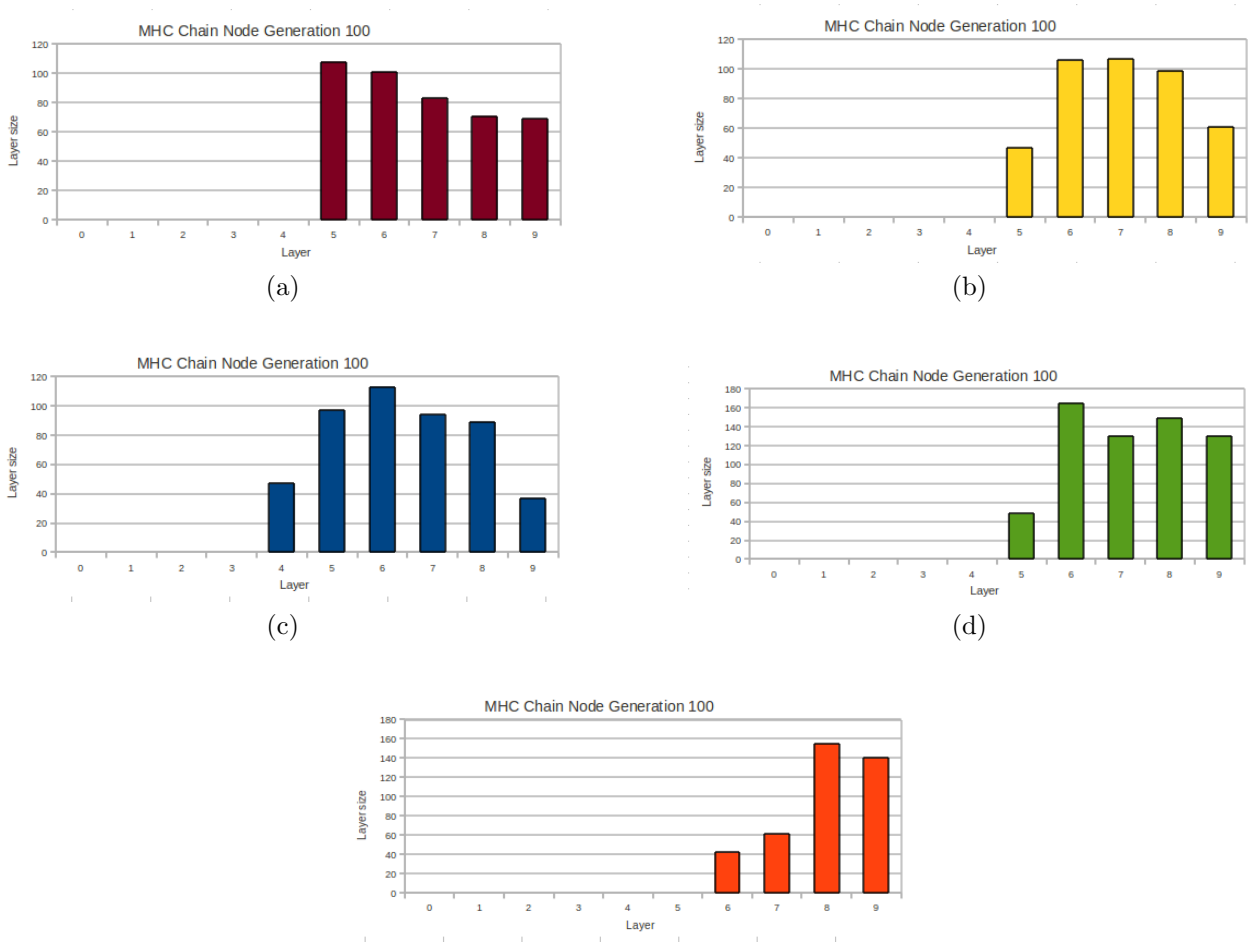


Figure 8: MHC (nodes generated = 100)

Yellow (a): PRigid:0.05 PFlex:0.1 Blue (b): PRigid:0.05 PFlex:0.06

Green (c): PRigid:0.0005 PFlex:0.1 Orange (d): PRigid:0.0005 PFlex:0.06

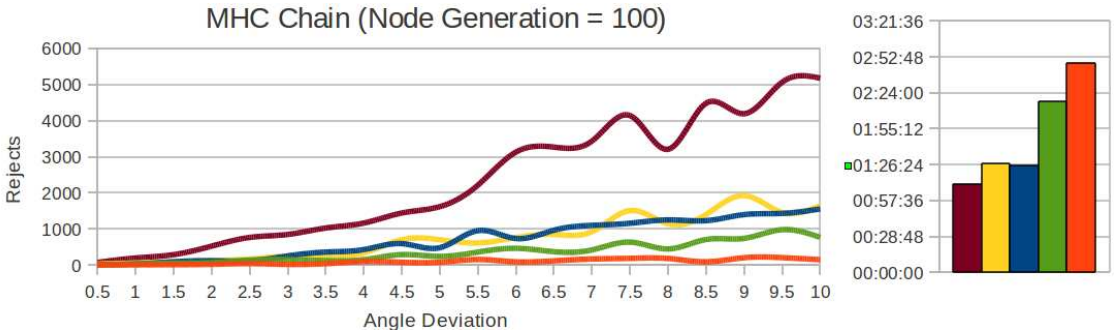


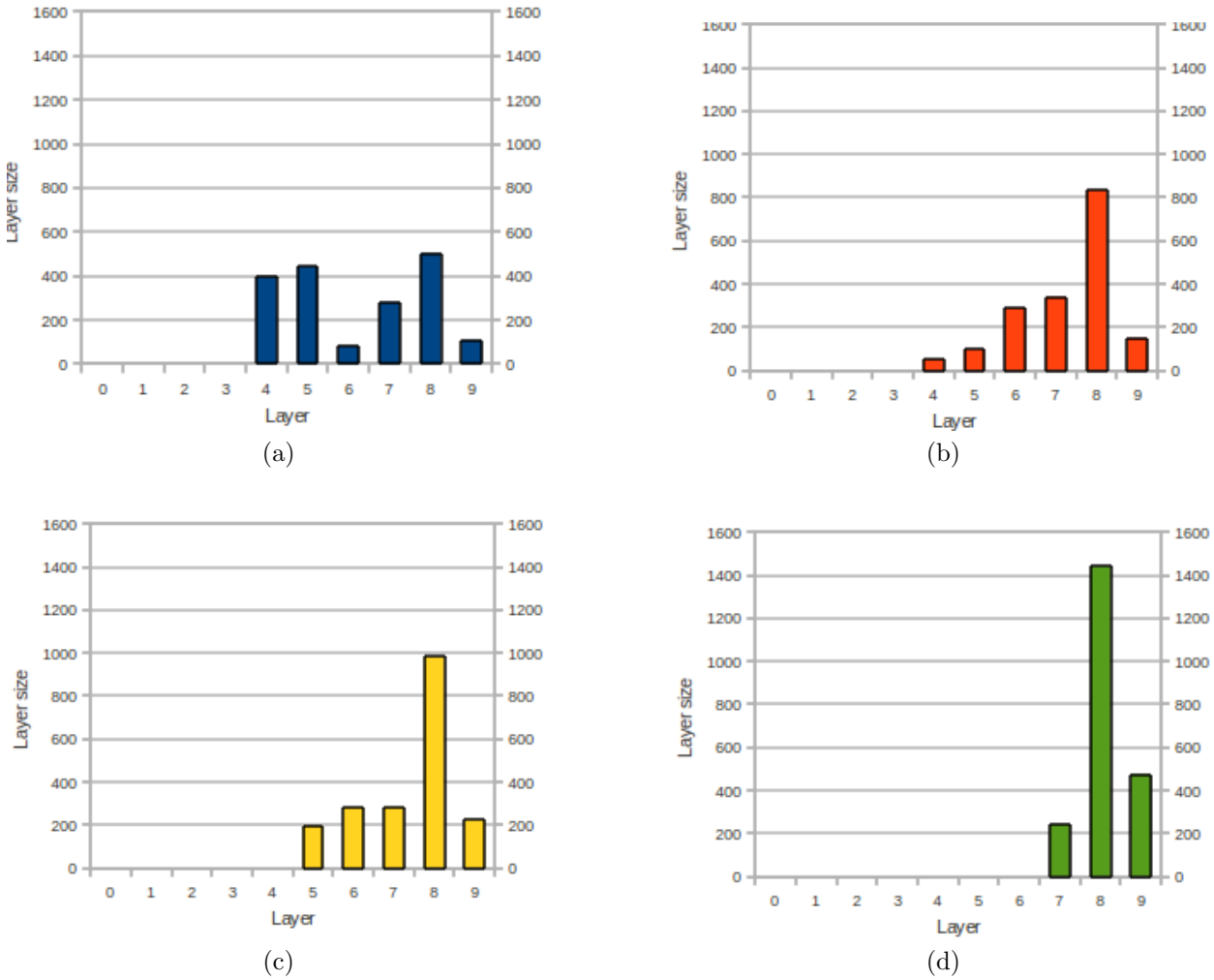
Figure 9: MHC (nodes generated = 100)

Red: PRigid: 0.05 PFlex: 0.08 Yellow: PRigid 0.05 PFlex 0.1

Blue: PRigid: 0.05 PFlex: 0.06 Green: PRigid: 0.0005 PFlex: 0.1

Orange: PRigid: 0.0005 PFlex: 0.06

Figures 10 and 11 show the data collected for the protein 1QLP. Figure 11 shows that the best node populations and distribution over layers occurred for the parameters used in (a) PFlex 0.06 and PRigid 0.05, however Figure 11 shows that the parameters used in (a) also are associated with a high number of rejected nodes.



(a) (b)

(c) (d)

Figure 10: 1QLP (1 Alpha Antitrypsin)
Blue(a): PFlex: 0.06 PRigid:0.05 Orange(b): PFlex: 0.06 PRigid:0.005
Yellow(c): PFlex: 0.06 PRigid:0.0005 Green(d): PFlex: 0.01 PRigid:0.0005

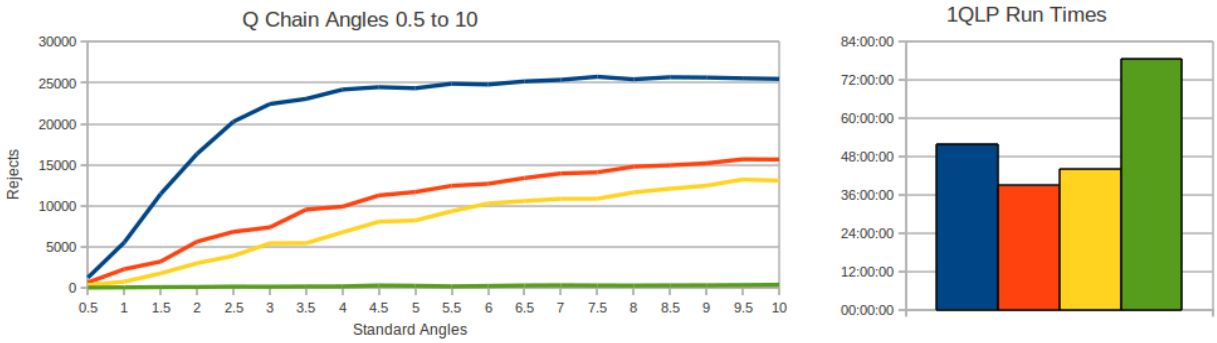


Figure 11: 1QLP (1 Alpha Antitrypsin)
Blue(a): PFlex: 0.06 PRigid:0.05 Orange(b): PFlex: 0.06 PRigid:0.005
Yellow(c): PFlex: 0.06 PRigid:0.0005 Green(d): PFlex: 0.01 PRigid:0.0005

5 Application: Ligand-receptor binding

Section temporarily removed, pending publication of data used for this study.

6 Conclusion

When studying layer generation for MHC, TCR and 1QLP, we found that smaller PFlex and PRigid values and larger standard angle setting produced more nodes and edges in more layers of the protein energy landscape. However, this also resulted in a longer time needed to generate the landscape. The tradeoff of time must be considered against the increased coverage of the landscape.

A portion of this section has been temporarily removed, pending publication of data used for this study.

7 Acknowledgments

Special thanks to our coworkers; Latha Doddikadi (Computer Science graduate student, UNM), John Baxter (Computer Science PhD student, UNM), and Anthony Lee Hickerson (Computer Science undergraduate student, UNM). Latha provided programing assistance, John provided a review of this paper and svn support, and Anthony provided assistance with NX. Thank you also to Bill McCabe for assistance with debugging our programs.
A portion of this section has been temporarily removed, pending publication of data used for this study.

References

[1] 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. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 2–11, 2002.

[2] 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.

[3] 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.

[4] 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.

[5] 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.*

[6] J. Cortés and T. Siméon. Sampling-based motion planning under kinematic loop-closure constraints. In *Algorithmic Foundations of Robotics VI*, pages 75–90. Springer, Berlin/Heidelberg, 2005. book contains the proceedings of the International Workshop on the Algorithmic Foundations of Robotics (WAFR), Utrecht/Zeist, The Netherlands, 2004.

[7] J. Cortés, T. Siméon, M. Remaud-Siméon, and V. Tran. Geometric algorithms for the conformational analysis of long protein loops. *J. Computat. Chem.*, 25(7):956–967, 2004.

[8] P. Elliott, X. Pei, T. Dafforn, and D.A.Lomas. Topography of a 2.0 a structure of alpha1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci.*, 9:1274, 2000.

[9] D. Jacobs. Generic rigidity in three-dimensional bond-bending networks. *J. Phys. A: Math. Gen.*, 31:6653–6668, 1998.

[10] D. Jacobs and M. Thorpe. Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.*, 75(22):4051–4054, 1995.

[11] D. Jacobs and M. Thorpe. Generic rigidity percolation in two dimensions. *Phys. Rev. E*, 53(4):3682–3693, 1996.

[12] D. J. Jacobs and B. Hendrickson. An algorithm for two dimensional rigidity percolation: The pebble game. *J. Comp. Phys*, 137:346–368, 1997.

[13] W. Jorgensen and J. Tirado-Rives. The opls potential functions for proteins. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[14] J.-P. Kinet. The high-affinity IgE receptor (FcRI): From physiology to pathology. *ARI*, 17:931–972, 1999.

[15] A. Kulczycki, Jr. and H. Metzger. The interaction of IgE with rat basophilic leukemia cells. II. quantitative aspects of the binding reaction. *JEM*, 140:1676–1695, 1974.

[16] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins*, 35:133–152, 1999. http://mingus.sci.ccny.cuny.edu/server/.

[17] A. Lee and I. Streinu. Pebble game algorithms and sparse graphs. *European Conference on Combinatorics, Graph Theory and Applications*, 2005.

[18] M. Levitt. Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.*, 170:723–764, 1983.

[19] A. J. McMichael and E. Y. Jones. First-class control of HIV-1. *Science*, 330:1488–1490, 2010.

[20] H. Metzger. Transmembrane signaling: The joy of aggregation. *J. Immunol.*, 5:1477–1487, 1992.

[21] M. Morales, L. Tapia, R. Pearce, S. Rodriguez, and N. M. Amato. A machine learning approach for feature-sensitive motion planning. In *Algorithmic Foundations of Robotics VI*, pages 361–376. Springer, Berlin/Heidelberg, 2005. book contains the proceedings of the International Workshop on the Algorithmic Foundations of Robotics (WAFR), Utrecht/Zeist, The Netherlands, 2004.

[22] A. P. Singh, J.-C. Latombe, and D. L. Brutlag. A motion planning approach to flexible ligand binding. In *Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, pages 252–261, 1999.

[23] G. Song. *A Motion Planning Approach to Protein Folding.* Ph.D. dissertation, Dept. of Computer Science, Texas A&M University, December 2004.

[24] 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.

[25] M. J. Sternberg. *Protein Structure Prediction.* OIRL Press at Oxford University Press, 1996.

[26] 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.

[27] X. Tang, S. Thomas, L. Tapia, and N. M. Amato. Tools for simulating and analyzing RNA folding kinetics. In *Proc. Int. Conf. Comput. Molecular Biology (RECOMB)*, pages 268–282, 2007.

[28] X. Tang, S. Thomas, L. Tapia, D. P. Giedroc, and N. M. Amato. Simulating RNA folding kinetics on approximated energy landscapes. *J. Mol. Biol.*, 381:1055–1067, 2008.

[29] L. Tapia, N. M. Amato, and R. Elber. Late and parallel events in the folding of protein G. *under submission*, 2011.

[30] L. Tapia, X. Tang, S. Thomas, and N. M. Amato. Kinetics analysis methods for approximate folding landscapes. *Bioinformatics*, 23(13):539–548, 2007. Special issue of Int. Conf. on Intelligent Systems for Molecular Biology (ISMB) & European Conf. on Computational Biology (ECCB) 2007.

[31] L. Tapia, S. Thomas, and N. M. Amato. A motion planning approach to studying molecular motions. *Communications in Information and Systems*, 10(1):53–68, 2010. special issue in honor of Michael Waterman.

[32] S. Thomas, G. Song, and N. Amato. Protein folding by motion planning. *Physical Biology*, 2:S148–S155, 2005.

[33] S. Thomas, X. Tang, L. Tapia, and N. M. Amato. Simulating protein motions with rigidity analysis. *J. Comput. Biol.*, 14(6):839–855, 2007. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2006.

[34] R. Webster, L. Tapia, and A. Howells. Simulation of protein movement using multiple articulated linkages and probabilistic roadmap methods. Technical report, Computer Science Department, University of New Mexico, Aug 2011.

[35] S. Weiner, P. Kollman, D. Case, U. Singh, C. Ghio, G. Alagona, S. P. Jr., and P. Weiner. A new force field for the molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.