Simulation of Protein Movement using Probabilistic Roadmap Methods[†]

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

1 Introduction

Major histocompatibility complex (MHC) is responsible for recognizing infectious agents, such as HIV, that enter the body. MHC can activate the immune response by binding to T-cell receptors (TCR). Previous studies of molecular binding often only considered the static shapes of the molecules; however, protein motion is critical to TCR-MHC interaction, and is often instrumental in determining whether binding will occur. In a previous study, it has been shown that MHC can have an influence on the speed of AIDS progression. HIV-1 peptides that are bound to the MHC are critical in AIDS protection [15]. An important step in vaccine development is determining which peptides will bind to MHC, which will allow the peptide-MHC complex to be presented to TCRs and start an immune response. Similar to the importance of protein movement, the interaction of a peptide and a receptor can cause physiological changes. One ligand, IgE-FcϵR, and receptor, 1RFO, have been linked to allergies. However, different clusters of IgE-FcϵR trigger very different physiological responses [17]. The challenge is determining what protein and ligand motions increase and decrease binding propensity, as well as how IgE-FcϵR and 1RFO clusters form.

This project simulates the movements of molecules in order to determine how proteins, specifically MHC and TCR, as well as IgE-FcϵR and 1RFO, bind together. Articulated linkages are used to simulate proteins and ligands because, like proteins and ligands, they are made up of chains and joints that bend and revolve independently. Also, coarse grained rigid representations are used. Probabilistic roadmap methods (PRMs) were used to build graphs corresponding to approximate maps of the molecule's energy landscape. It builds a roadmap by generating nodes, which here involves sampling valid protein conformations, and connecting selected nodes, or conformation states, with energetically feasible transitions, using a local planner [18]. The roadmaps may then be used to determine the most energetically likely transitions between conformations, using the best path between two nodes on the roadmap [6]. We built roadmaps of multiple systems in order to model protein and ligand-receptor movement.

Modeling protein and ligand movement using PRMs makes it possible to gain an understanding of the conformational changes that occur during the binding. We show dynamic views of molecular interactions with a low computational expense. This will hopefully improve vaccine development through a better understanding of protein binding propensities.

2 Related Work

2.1 Protein Motion Planning

The motion planning (MP) problem asks how to find a sequence of valid state transitions that takes an object, such as a robot, from an initial state to a goal state [29]. PRMs build roadmaps in order to search for solutions to MP problem instances. Conformations, or roadmap nodes, are sampled from the configuration space, also known as C-space, and then transitions between 'nearby' conformations are encoded as roadmap edges. A roadmap of a single rigid body is shown in Figure 1. In this figure, the green is the node generations. The large blue blocks are the obstacles that the robot must move around, and the small blue cube is the robot. The lines connecting the node generations are the possible connections that can connect the each node generation, allowing the robot to move from one configuration to another. In a previous study, PRMs were used on articulated linkages [29]. In this study, only robots with an arbitrary number of joints were considered. Each joint has limits that is dependent on the structure of the protein, which inhibit the range of motion. As each robot is moved in the C-space, all of the joints must be within their limits [12]. Articulated linkages are good representations of proteins because proteins are made up of atoms and bonds that bend and revolve independently.

When creating roadmaps for proteins, nodes are retained based on their energy level. The roadmap that is produced is an approximation of a molecule's energy landscape, and the quality is contingent on the sampling strategy. Ideally, the nodes will be placed throughout the roadmap, and not simply near the native state of the protein. This will give a roadmap that captures the true range of motion for the protein. Once the nodes are generated, they are connected to their k-closest neighbors. Thousands of paths between a start and goal confirmation are extracted, and the highest quality path is chosen. The roadmap is shown in Figure 2.

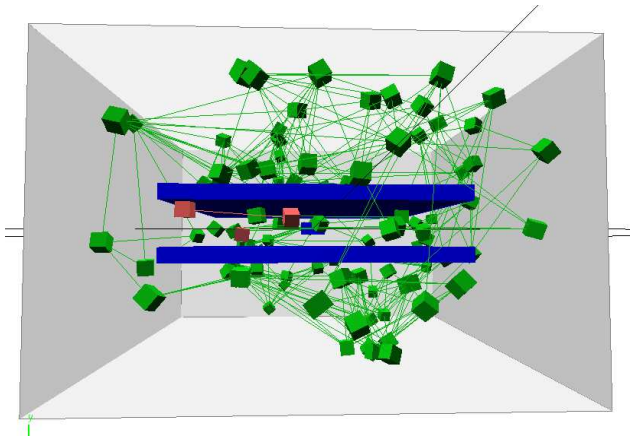


Figure 1: A sample roadmap of a rigid body, a small blue box, in an environment with two flat obstacles, large blue rectangles, with a narrow space between them. The green boxes are the node generations of the robot. The lines are the connections formed between the node generations, showing how the robot can move from one node generation to another.

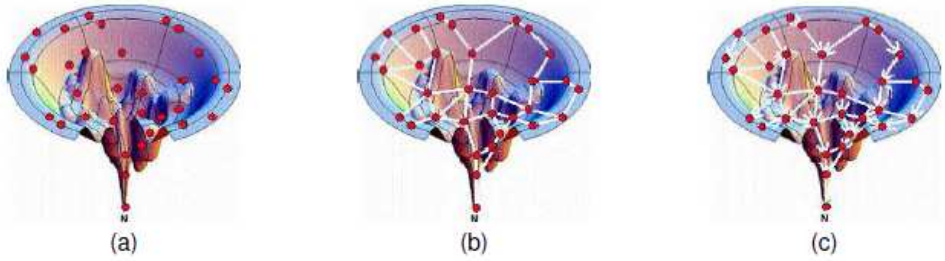


Figure 2: A PRM roadmap for molecular folding shown imposed on a visualization of the molecules energy landscape: (a) after node generation (note sampling is denser around N, the known native structure), (b) after the connection phase, and (c) using it to extract folding paths to the known native state.

Our PRM framework, which was originally developed for robotic motion planning, has been successfully applied for molecular motions to study protein folding and motion [27, 28, 31]. Our strategy prefers low energy confirmations and transitions. During the sampling phase, lower energy samples have a higher retention probability, and during the node connection, each connection is determined by the energies of all the intermediate conformations along the transitions. Therefore, the shortest paths in the roadmap correspond to the most energetically feasible paths, or the least-weight paths between configurations, and thousands of feasible pathways are encoded by roadmaps. PRM-based approaches have also been applied to several other molecular domains [1, 2, 5, 21, 22, 23, 24, 30]. Singh, Latombe, and Brutlag first applied PRMs to protein/ligand binding [21]. Another PRM variant later explored this problem with additional success [5]. We have applied PRMs to model protein folding pathways [27, 28, 31] and RNA folding kinetics [25, 26]. PRMs have also been used by other groups to study molecular motions [3, 4, 7, 8].

Potential Energy Calculation. To calculate the energy function, a step function approximation of the van der Walls component is used. If two side chains are too close, a potential energy above the disqualification threshold is returned. 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 kcal/mol and $d_0 = d_c = 2$Å [14]. Full details can be found in [2].

Biased Sampling. All of our samples are based on their energy. A sample q, with potential energy E_q, is accepted with probability:

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

where E_{min} is the potential energy of the open chain and E_{max} is $2E_{min}$. The quality of the roadmap is dependent on the sampling strategy. Generally, we are most interested in regions 'near' the target conformation, so we concentrate sampling there.

Connecting the Roadmap. We attempt to connect each node in the roadmap with its k closest neighbors. To connect the roadmap, the two nodes, q_1 and q_2, are labeled with edge weights that express the viability of transitioning between the two. First, all the intermediate nodes that connect q_1 to q_2, $q_1 = c_0, c_1, ..., c_{n-1}, c_n = q_2$, are identified. Then, for each pair of consecutive confirmations c_i and c_{i+1}, the probability P_1 of transitioning from c_i and c_{i+1} depends on the difference between their potential energies $\Delta E_i = E(c_{i+1}) - E(c_i)$:

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

This helps keep the balance between the two adjacent states and allows the edge weight to be calculated by summing the logarithms of the probabilities for all of the pairs of consecutive conformations. Since this defines the edge weight, when extracting a path, we can use graph search algorithms to discover the most energetically feasible pathways.

In previous work [2], there are provided methods for building an approximate map of a protein's potential energy landscape. These roadmaps give an approximate view of a protein's folding landscape. In the past, low-energy pathways and validated secondary structure formation order have been successfully extracted.

2.2 Feature-Sensitive Motion Planning

There are many different solutions to the MP problem, because they are high dimensional problems that make complete MP solutions computationally intractable, with each solution having pros and cons. Many environments exist which can be mapped out using PRMs, including proteins. One such solution to the MP problem is Feature-Sensitive motion planning [18], which has been used in previous work. In Feature-Sensitive motion planning, the C-space is divided into regions that are suited to a method in the library of motion planners. Then, depending on the type of region, the best-suited method is applied to each region, and all of the region roadmaps are combined to create a roadmap of the entire C-space [18]. While this gives a detailed roadmap of varied environments, in order for each region to be correctly identified, many examples of each type of C-space must be implemented first, to allow a machine learning algorithm to learn how to classify regions.

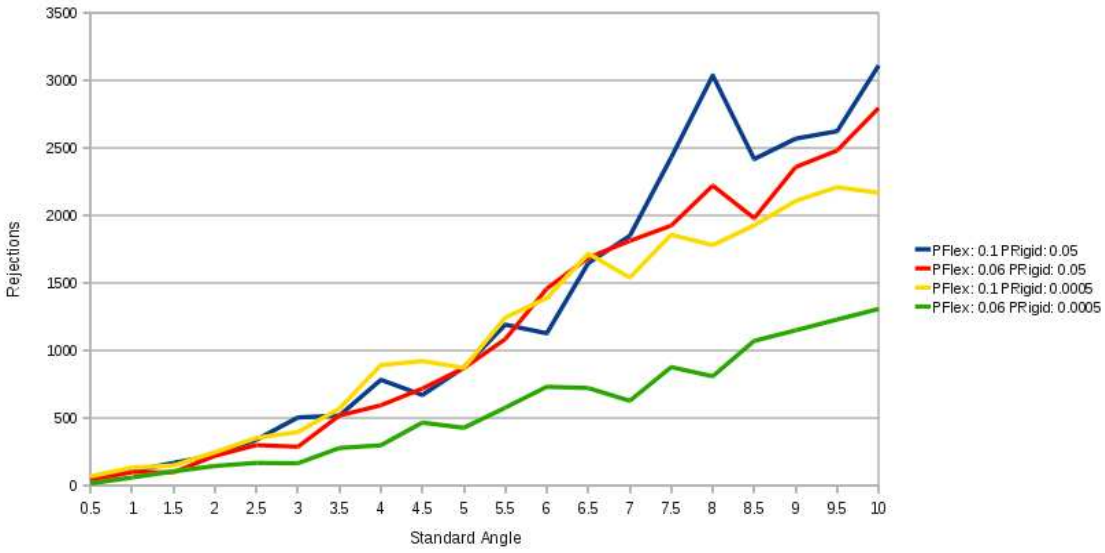
2.3 Ligand-Receptor Interactions

When a ligand binds to a receptor, the shape of the three dimensional receptor is altered. This can cause physiological responses; for example, when the antigenspecific IgE antibody binds to FcϵR, which then establishes antigen recognition by mast cells, which play a role in allergy response, wound healing, and defense against pathogens [13]. When IgE-FcϵR interacts with an antigen, a clustering of FcϵR is formed, which initiates a signal that causes the release of histamines and other allergic reaction mediators [16]. In previous studies, various reagents have been used to induce the clustering of FcϵR, showing that the signaling

and cellular responses are dependent on the properties of the FcϵR accumulations [9, 10, 11, 20, 32]. It has been reported that the synthesis of a trivalent antigen can originate a cellular secretory response, but a structurally similar bivalent antigen did not cause detectable levels of secretion [19]. Previous studies have also shown the clustering of IgE-FcϵR induced by trivalent antigens differs greatly from clustering of IgE-FcϵR induced by bivalent antigen [17]. Because it triggers different responses, it is important to investigate the clustering of IgE-FcϵR complexes caused by trivalent antigens [17].

3 Application: Parameter study

In this section we are studying the parameters for node generation of proteins. This work is particularly useful to our feature-sensitive MP work since the best parameters we find can be used as starting points for automated parameter searches.

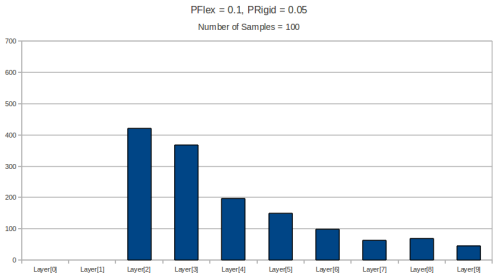


TCR: Number of Samples = 100

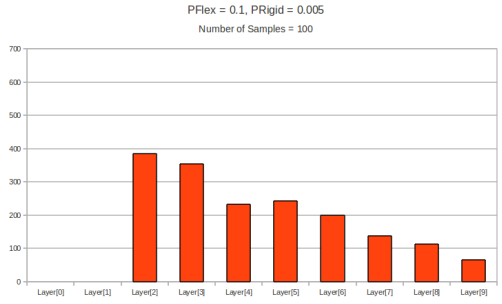
Figure 3: The number of samples = 100 for TCR. The number of rejections is graphed against the standard angle. P_{Flex} is varied between 0.1 - 0.006, and P_{Rigid} is varied between 0.05 - 0.0005.

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 P_{Flex}, and the probability that a portion of the structure that is labeled as rigid is perturbed, known as P_{Rigid}. 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 numbers of native contacts, 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 higher the quality of the nodes that are produced.

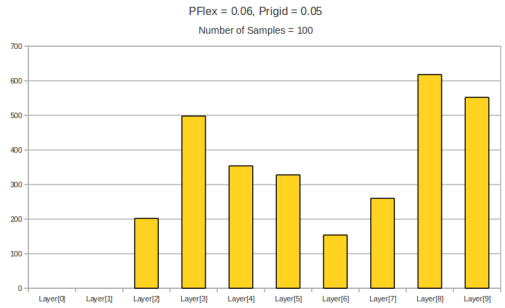
For the parameter study, I focused on TCR and Alpha-1-Antitrypsin (1QLP). TCR was studied because when molecules such as MHC bind to it, an immune response is activated. 1QLP was studied because it is larger than TCR and MHC. 1QLP has 372 residues, while TCR only has 200 residues. Also, 1QLP is highly studied, and is a mixed structure protein. It can aggregate and has been linked to diseases such as chronic obstructive pulmonary disease, or COPD.



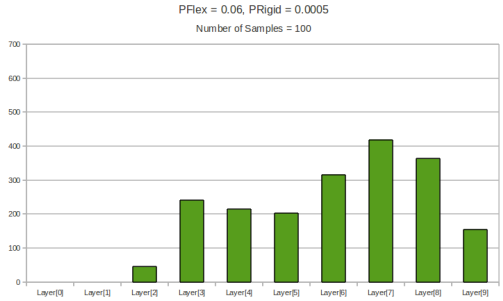
(a) $P_{Flex} = 0.1$, $P_{Rigid} = 0.05$



(b) $P_{Flex} = 0.1$, $P_{Rigid} = 0.0005$



(c) $P_{Flex} = 0.06$, $P_{Rigid} = 0.05$



(d) P_{Flex}, $P_{Rigid} = 0.0005$

Figure 4: The protein studied is 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 $P_{Flex} = 0.1$ and $P_{Rigid} = 0.05$. In (b), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.1$ and $P_{Rigid} = 0.0005$. In (c), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.06$ and $P_{Rigid} = 0.05$. In (d), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.06$ and $P_{Rigid} = 0.0005$.

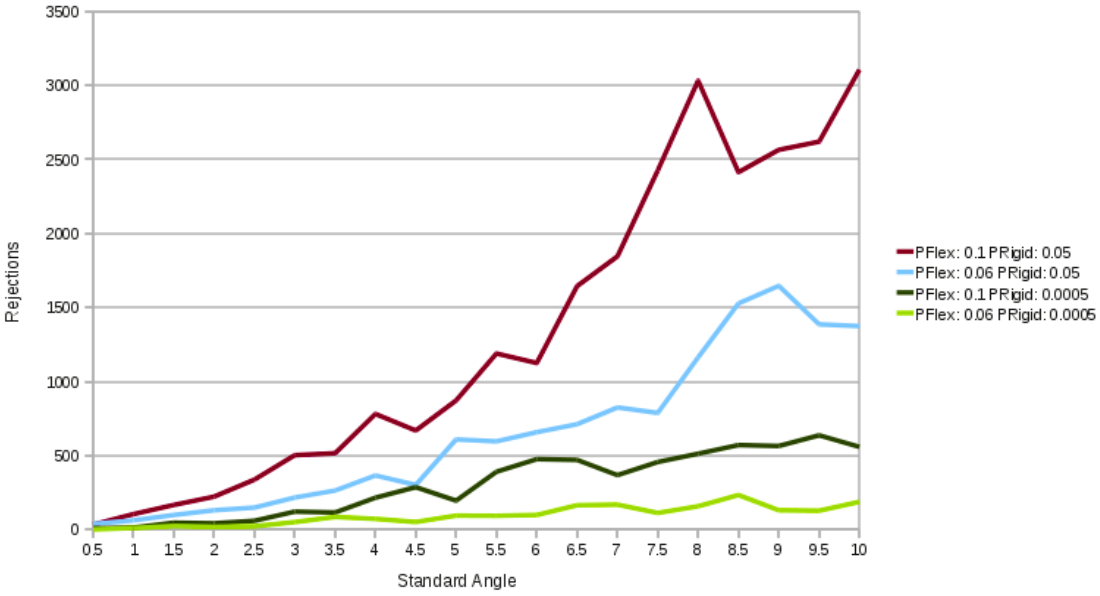
3.1 Methods

The parameters that were changed were P_{Rigid} and P_{Flex}. For large proteins, such as TCR and 1QLP, the P_{Rigid} and P_{Flex} values should be fairly low. Also, it is logical to keep the P_{Rigid} value below the P_{Flex} value. For TCR, P_{Rigid} was kept between 0.05-0.0005, while P_{Flex} was kept between 0.1-0.06. For 1QLP, only the P_{Flex} values were adjusted, and were kept between 0.1-0.02, with the P_{Rigid} value kept at a constant 0.01. The standard angles for the proteins were kept low, between 0.5 and 10.0. The data for each layer were captured in every run to see the quality of the node placement, as well as the time each run took to complete. The number of rejections for every angle were graphed, as well as the distribution of the nodes for every protein. Then, the best parameters were chosen based on the least number of rejections, the quality of the nodes produced, and the time each run took to complete.

3.2 Results

For the first tests for TCR, the number of samples was set at 100. P_{Flex} was kept between 0.1-0.06, and P_{Rigid} was kept between 0.05-0.0005, as shown in Figure 3. The lower the P_{Flex} and P_{Rigid} values, the lower the number of rejections that occurred at each standard angle, as shown in Figure 3. The distribution of the nodes across the layers is also shown in Figure 4, where layer 9 is the native state. When $P_{Flex} = 0.06$ and $P_{Rigid} = 0.0005$, the number of rejections are at their lowest, and the nodes are distributed across 8 layers. When $P_{Flex} = 0.1$ and $P_{Rigid} = 0.05$ or 0.0005, the nodes are also distributed across 8 layers, but the number of rejections is much higher.

For the second tests for TCR, the number of nodes generated was set at 100. P_{Flex} was kept between 0.1-0.06 and P_{Rigid} was kept between 0.05-0.0005, as shown in Figure 5. The distribution of nodes across



TCR: Node Generation = 100

Figure 5: The number of nodes generated = 100 for TCR. P_{Flex} is varied between 0.1 - 0.006, and P_{Rigid} is varied between 0.05 - 0.0005.

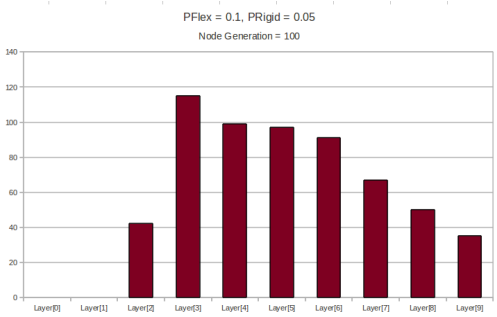
the layers is shown in Figure 6. When $P_{Flex} = 0.06$ and $P_{Rigid} = 0.0005$, the number of rejections are the lowest, however the distribution of nodes is not as diverse, which means the nodes are not at their highest quality. When $P_{Flex} = 0.06$ and $P_{Rigid} = 0.05$, the number of rejections is fairly low, and the nodes are highly distributed.

When testing 1QLP, the number of samples and the number of nodes generated were not changed. P_{Rigid} was kept at 0.01, and P_{Flex} varied between 0.1-0.02, as shown in Figure 7. The distribution of nodes across the layers is shown in Figure 8. When $P_{Flex} = 0.1$, the layers have the best distribution, however the number of rejections are quite high. When $P_{Flex} = 0.02$, the number of rejections is lower than for other tested values of P_{Flex}, but the nodes are not generated across as many layers as when $P_{Flex} = 0.1$.

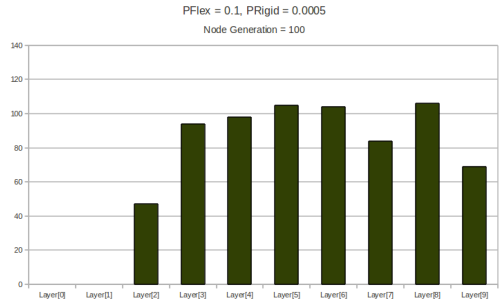
3.3 Conclusion

TCR. When the number of samples is set at 100, the number of rejections tends to be higher, especially when the values of P_{Flex} and P_{Rigid} are low. The number of nodes generated in each layer tends to be higher, and the nodes are distributed across more layers. However, this is to be expected, as when the number of nodes generated is set at 100, the number of nodes in each layer should be lower. When the number of samples is set at 100, the parameters with the smallest number of rejections and largest distribution of nodes is when $P_{Flex} = 0.06$ and $P_{Rigid} = 0.0005$. When the number of nodes generated is set at 100, the parameters with the smallest number of rejections and the largest distribution of nodes is when $P_{Flex} = 0.1$ and $P_{Rigid} = 0.0005$.

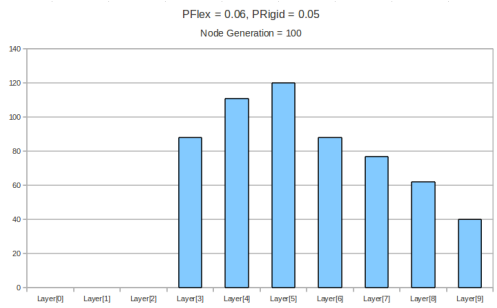
1QLP. The results of 1QLP were much more varied than the results for TCR, but since 1QLP is a larger protein, this is not surprising. The number of rejections decrease as P_{Flex} decreases, with the notable exception being when $P_{Flex} = 0.08$. The nodes are not distributed across as many layers as the nodes for TCR. The parameters with the smallest number of rejections and biggest distribution of nodes is when $P_{Flex} = 0.1$.



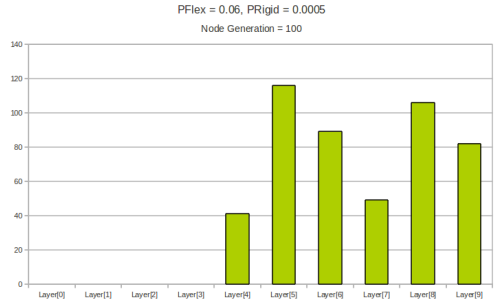
(a) $P_{Flex} = 0.1$, $P_{Rigid} = 0.05$



(b) $P_{Flex} = 0.1$, $P_{Rigid} = 0.0005$

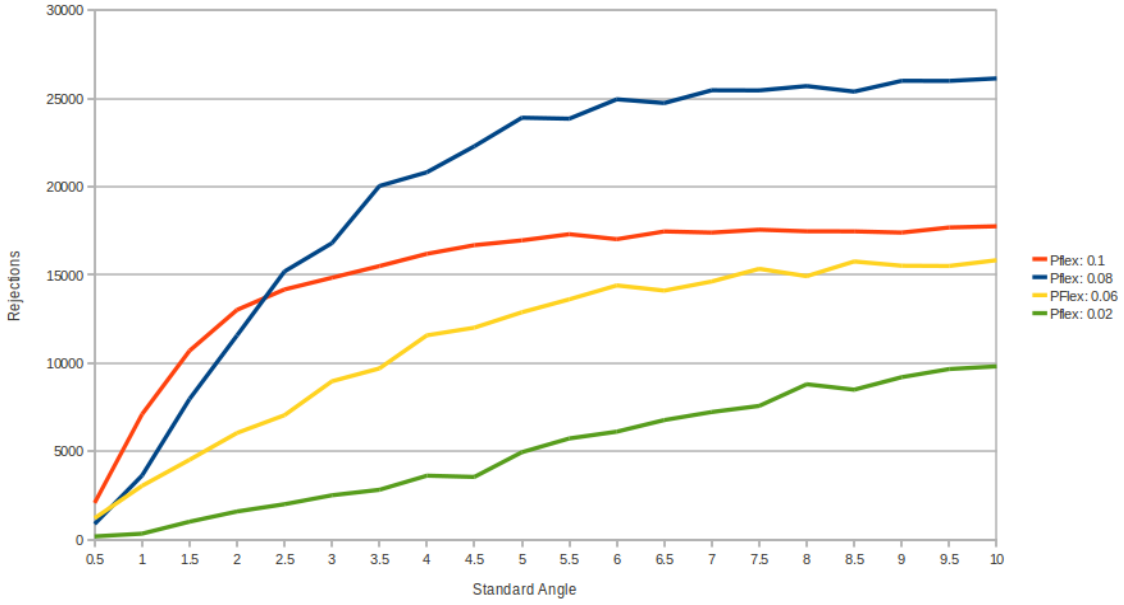


(c) $P_{Flex} = 0.06$, $P_{Rigid} = 0.05$



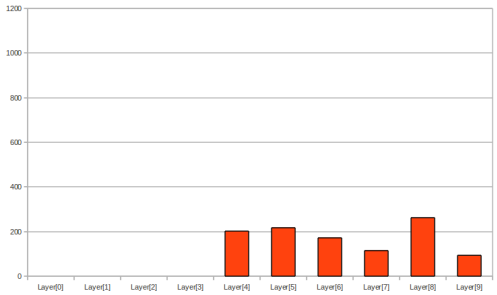
(d) $P_{Flex} = 0.06$, $P_{Rigid} = 0.0005$

Figure 6: The protein studied is 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 $P_{Flex} = 0.1$ and $P_{Rigid} = 0.05$. In (b), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.1$ and $P_{Rigid} = 0.0005$. In (c), the number of nodes generated in every layer for $P_{Flex} = 0.06$ and $P_{Rigid} = 0.05$. In (d), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.06$ and $P_{Rigid} = 0.0005$.

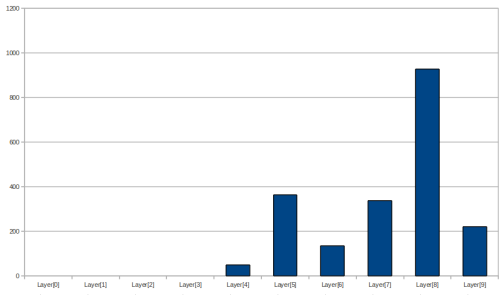


1QLP

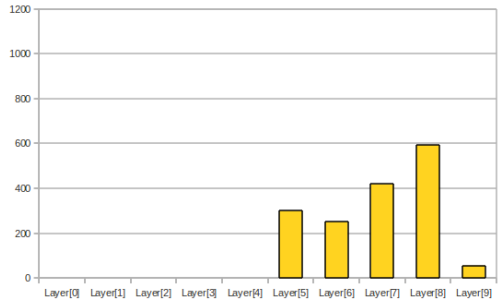
Figure 7: The rejections are graphed against the standard angle for 1QLP. P_{Flex} is varied between 0.1 - 0.02, while P_{Rigid} is kept constant at 0.01.



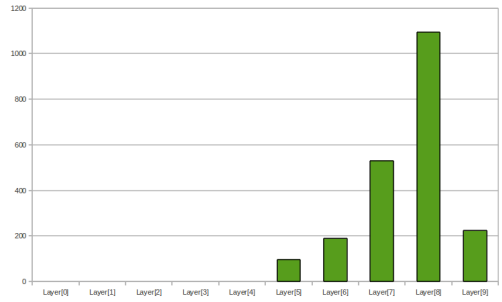
(a) $P_{Flex} = 0.1$, $P_{Rigid} = 0.01$



(b) $P_{Flex} = 0.08$, $P_{Rigid} = 0.01$



(c) $P_{Flex} = 0.06$, $P_{Rigid} = 0.01$



(d) $P_{Flex} = 0.02$, $P_{Rigid} = 0.01$

Figure 8: The protein studied is 1QLP. 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 $P_{Flex} = 0.1$. In (b), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.08$. In (c), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.06$. In (d), the number of nodes generated in every layer is graphed for $P_{Flex} = 0.02$.

4 Application: Ligand-Receptor Binding

This section is awaiting publication by the University of New Mexico Medical School.

5 Conclusion

In this paper, we presented a way to model the movements of proteins, ligands, and receptors using PRMs. The advantage to using PRMs is that it can allow dynamic views of molecular interactions to be simulated with a low computational expense, which makes thousands of simulations possible. By using articulated linkages to model proteins, ligands, and receptors, it possible to observe the conformational changes that occur during binding. The parameter study, shown in Section 3, demonstrated that changing the P_{Flex} and P_{Rigid} of a protein can affect the number of rejections that are generated, as well as how far away from the native state nodes are placed. The Receptor-Ligand binding study, shown in Section 4, demonstrated that it is possible to use articulated linkages as accurate representations of ligands and receptors. It also showed that, as we suspected, the more robots or nodes that are generated in each roadmap, the longer it takes to generate the robot. Hopefully, both will be used in the future to study ligand-receptor and protein binding, ideally to aid the development of new vaccines.

6 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 programming 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.

We also wish to thank the Spaciotemporal Modeling Center at UNM and Los Alamos National Labs for information on ligands and the IgE and FcϵR pdb files and images.

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] H. Choset, K. M. Lynch, S. Hutchinson, G. A. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations.* MIT Press, Cambridge, MA, June 2005.

[7] 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.

[8] 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.

[9] C. Fewtrell and H. Metzger. Larger olilgomers of IgE are more effective than dimers in stimulating rat basophilic leukemia cells. *J. Immunol.*, 125:701–710, 1980.

[10] W. S. Hlavacek, A. S. Perelson, B. Sulzer, J. Bold, J. Paar, W. Gorman, and R. G. Posner. Quantifying aggregation of IgE-FcRI by multivalent agent. *Biophys. J.*, 76:2421–2431, 1999.

[11] D. Holowka, D. Sil, C. Torigoe, and B. Baird. Insights into immunoglobulin E receptor signaling from structurally defined ligands. *Immunol. Rev.*, 217:269–279, 2007.

[12] L. E. Kavraki, P. Švestka, J.-C. Latombe, and M. H. Overmars. Probabilistic roadmaps for path planning in high dimensional configuration spaces. *IEEE Trans. Robot. Autom.*, 12:566–580, 1996.

[13] J. Kinet. The high-affinity IgE receptor (FcRI): from physiology to pathology. *Annu. Rev. Immunol.*, 17:931–972, 1999.

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

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

[16] H. Metzger. Transmembrane signaling: the joy of aggregation. *J. Immunol.*, 149:1477–1487, 1992.

[17] M. I. Monine, R. G. Posner, P. B. Savage, J. R. Faeder, and W. S. Hlavacek. Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys. J.*, 98:48–56, 2010.

[18] 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.

[19] R. G. Posner, P. B. Savage, A. S. Peters, A. Macias, J. DelGado, G. Zwartz, L. A. Sklar, and W. S. Hlavacek. A quantitative approach for studying IgE-FcRI aggregation. *Mol. Immunol.*, 38:1221–1228, 2002.

[20] D. M. Segal, J. D. Taurog, and H. Metzger. Dimeric immunoglobulin E serves as a unit signal for mast cell degranulation. *Proc. Natl. Acad. Sci. USA*, 74(7):2993–2997, 1977.

[21] 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.

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

[23] 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.

[24] X. Tang, B. Kirkpatrick, S. Thomas, G. Song, and N. M. Amato. Using motion planning to study RNA folding kinetics. *J. Comput. Biol.*, 12(6):862–881, 2005. Special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2004.

[25] 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.

[26] 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.

[27] L. Tapia, X. Tang, S. Thomas, and N. M. Amato. Kinetics analysis methods for approximate folding landscapes. In *Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, pages 539–548, 2007.

[28] 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.

[29] L. Tapia, S. Thomas, B. Boyd, and N. M. Amato. An unsupervised adaptive strategy for constructing probabilistic roadmaps. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 4037–4044, May 2009.

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

[31] 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.

[32] K. Xu, B. Goldestein, D. Holowka, and B. Baird. Kinetics of multivalent antigen DNP-BSA binding to IgE-FcRI in relatinoship to the stimulated tyrosine phosphorylation of FcRI. *J. Immunol.*, 160:3225–3235, 1998.