Ligand Binding with OBPRM and Haptic User Input: Enhancing Automatic Motion Planning with Virtual Touch*

O. Burchan Bayazit — burchanb@cs.tamu.edu
Guang Song — gsong@cs.tamu.edu
Nancy M. Amato — amato@cs.tamu.edu

Technical Report TR00-025
Department of Computer Science
Texas A&M University
College Station, Texas, 77843-3112
October 9, 2000

Abstract

In this paper, we present a framework for studying ligand binding which is based on techniques recently developed in the robotics motion planning community. We are especially interested in locating binding sites on the protein for a ligand molecule. Our work investigates the performance of a fully automated motion planner, as well improvements obtained when supplementary user input is collected using a haptic device. Our results applying an obstacle-based probabilistic roadmap motion planning algorithm (OBPRM) to some known protein-ligand pairs are very encouraging. In particular, we were able to automatically generate configurations close to, and correctly identify, the true binding site in the three protein-ligand complexes we tested. We find that user input helps the planner, and a haptic device helps the user to understand the protein structure by enabling them to feel the forces which are hard to visualize.

*This research supported in part by NSF CAREER Award CCR-9624315 (with REU Supplement), NSF Grants IIS-9619850 (with REU Supplement), ACI-9872126, EIA-9975018, EIA-9805823, and EIA-9810937, and by the Texas Higher Education Coordinating Board under grant ARP-036327-017. Bayazit is supported in part by the Turkish Ministry of Education.

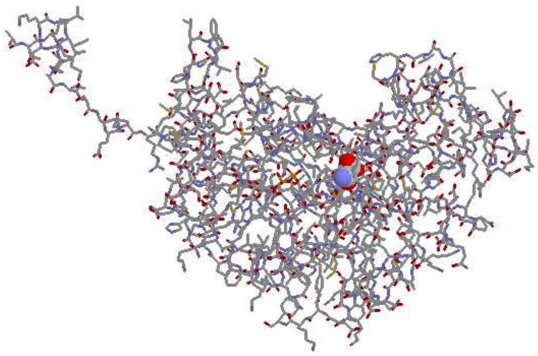


Figure 1: A protein in wireframe with a space-fill ligand. PDB file: 1LDM

1 Introduction

Efficiency of a drug molecule is measured by its ability to find a specific position and orientation inside a protein. This process is called binding (sometimes called docking) and the drug molecule is often referred as a ligand. The binding configuration should satisfy some constraints based on geometry, electrostatic, and chemical reactions between the ligand and protein atoms. A good binding site should also be accessible to the ligand which must reach it from an outside location. This makes the path to the binding site important, and motivates the use of motion planning to study this problem.

Most researchers investigating automated docking methods simplify the problem by treating the ligand and protein molecules as rigid bodies. Our experiments show that this simplification negatively impacts our ability to identify the binding site. However, if the molecules are flexible, i.e., the molecules are articulated bodies, the problem becomes very high-dimensional and a deterministic search of the configuration space is almost impossible. In motion planning terms, the configuration space, or C-space, of a movable object is the space consisting of all possible positions and orientations of the object in the environment. In terms of ligand binding, a configuration is a conformation of the ligand in a particular position and orientation.

In this work we will apply *probabilistic roadmap* (PRM) motion planning methods to protein-ligand binding. PRMs have been extremely successful for problems with high-dimensional configuration spaces, and moreover, are very simple to apply, requiring only the ability to randomly generate points in C-space, and then test them for feasibility. As discussed below, the potential applicability of PRMs to protein-ligand binding was first noted by Singh *et al.* [29]. The configuration of the ligand in the binding site has low potential energy, and so the usual PRM feasibility test (collision) is replaced by a test for low potential energy. In this study, we use an obstacle-based PRM called OBPRM [1] that generates configurations close to C-space obstacles which results in configurations close to the protein's surface. Since our goal, the binding site, is a local minimum point of the potential, we used a gradient-descent like algorithm to drive the roadmap nodes towards local minimum. To recognize potential binding sites, we construct 'local roadmaps' around (most of) the roadmap nodes and give each candidate node a score based on an analysis of its local roadmap. Our approach was able to generate configurations close to, and correctly identify, the true binding site in the three tested protein-ligand complexes. Two of these were studied in [29], where in one case their approach failed to generate a configuration in the binding site.

While our results are very encouraging, a fully automated approach suffers from the known problems of PRMs. The most significant of which is the *narrow passage* problem [8], so called because it is difficult for the planner to sample important configurations in a narrow C-space region. In molecular docking, the narrow passage becomes a passage through high potential areas. Another challenge with PRMs is that it is hard for the user to visualize and understand the progress made by the planner. In molecular docking, this problem becomes even more significant since the familiar three-dimensional workspace is replaced with an energy space. To address these problems, we add a human operator equipped with a haptic device to the system. A haptic device enables the user to touch and feel a virtual object. The user adds human insight and the haptic device enables the user to explore the energy space. In our approach, the user selects some important configurations of the ligand with the haptic device and they are passed to the planner for further processing. Since feeling the potential energy surface is not sufficient to understand the planner's progress, roadmap visualization methods similar to [2] are also employed.

Although there has been much work on automated solutions of the molecular docking problem, they have mainly concentrated on fully automated solution of a simplified version of the problem (e.g., they have assumed rigid ligands). There has been less research on man-machine interaction for this problem and there has been even less work using novel technologies like haptic interfaces. We discuss relevant previous work in Section 2. We briefly describe probabilistic roadmap planning methods and our potential energy computations in Sections 3 and 4, respectively. Section 5 describes our haptic interaction system for molecular docking. We give details of our approach in Sections 6 and 7, and present experimental results in Section 8. Section 9 concludes our paper.

2 Previous Work

Many automated docking algorithms have been proposed, including, for example, AutoDock [21], Dock [12, 4], FlexX [28], FLOG [20], FTDock [10], and Gold [9]. While many successes have been achieved by such algorithms, it has been stated that the success ratio is often relatively low (2-20%) [32]. Also, in most cases these algorithms make the simplifying assumption that the ligand is rigid to reduce the computation cost.

Recently, a new approach which makes use of the PRM paradigm for protein-ligand bind was proposed in [29]. Their results were very promising, and are in large part the motivation for this work. They uniformly sample configurations, and resample more densely around a small subset of low potential configurations from the original sample. In essence, they identified randomly sampled configurations with low potential energies and further explored those regions. Since the goal of the obstacle-based PRM used in our work is to generate, directly, configurations near the protein's surface, we believe that it will provide a more representative sample than the uniform sampling approach of [29]. As will be seen in Section 8, some evidence in favor of our approach is our success in generating configurations in, and identifying, the binding site for a test case in which this previous method failed.

Because of the relatively low success rates for fully automated algorithms, some researchers have added a human operator into the decision process. For example, [3] discusses new strategies for man-machine interaction in molecular modeling. STALK [15] enables the user to interact with the genetic algorithms in a virtual reality environment. An operator can move the ligand with a mouse in SCULPT [30, 31]. In SMD [14], biochemists can "tag" molecules into different shapes by specifying external forces on the graphical interface. A revolutionary approach taken in [24, 23] was to include a haptic interface and let the user feel the force applied on the ligand during the docking process. In [25] it is shown that force feedback performed better than visual display in the docking process. In [26], an interactive molecular docking simulator was used for interactive molecular design. It is shown in [33] that SCULPT can be improved by adding a haptic device. [19] discusses using a haptic device in molecular modeling and drug design. Finally, [7] haptically visualizes the probability density functions of individual atomic orbitals.

Unfortunately, there has not been much work on improving automated planners with human input. In [2], we used a haptic device together with a PRM planner on motion planning problems typical of maintainability studies in mechanical CAD designs. Our results showed that the human insights improved the planner's ability to solve problems involving narrow passages in C-space.

3 Probabilistic Roadmap Methods

Given a description of the environment and a movable object (the 'robot'), the motion planning problem is to find a feasible path that takes the movable object from a given start to a given goal configuration [13]. Since there is strong evidence that any complete planner (one that is guaranteed to find a solution, or determine that none exists) requires time exponential in the number of degrees of freedom (dof) of the movable object [13], attention has focussed on randomized or probabilistic methods.

As mentioned in Section 1, our approach is based on the probabilistic roadmap (PRM) approach to motion planning [11]. Briefly, PRMs work by sampling points 'randomly' from C-space, and retaining those that satisfy certain feasibility requirements (e.g., they must correspond to collision-free configurations of the movable object). Then, these points are connected to form a graph, or roadmap, using some simple planning method to connect 'nearby' points. During query processing, paths connecting the start and goal configurations are extracted from the roadmap using standard graph search techniques. (See Figure 2.)

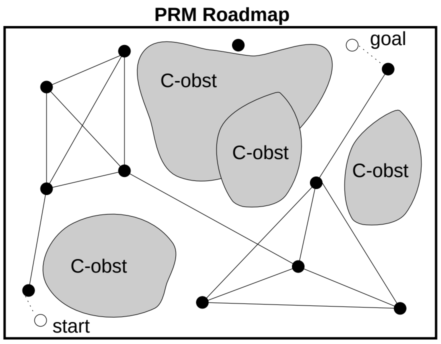


Figure 2: Querying a PRM roadmap (C-space).

We will use a particular variant of the PRM called the obstacle-based PRM or OBPRM [1]. In OBPRM, instead of generating the nodes uniformly at random in C-space, they are generated on or near constraint surfaces (e.g., contact configurations in traditional motion planning applications). As we will see, OBPRM is particularly well suited to the ligand binding problem.

A major strength of PRMs is that they are quite simple to apply, even for problems with high-dimensional configuration spaces, requiring only the ability to randomly generate points in C-space, and then test them for feasibility (the local connection can often be effected using multiple applications of the feasibility test).

The ligand binding problem has a notable difference from usual PRM applications, in that the traditional collision-free constraint is replaced by a preference for low energy conformations.

3.1 C-spaces of Ligands

Proteins usually consist of thousands of atoms, while the ligand typically consists of tens. While both protein and ligand are in fact flexible, we will model the protein as a rigid body and the ligand as a flexible body as was done in [29]. In particular, the ligand is modeled as an articulated body with a free base, where the torsional movement of atomic bonds is modeled by one degree of freedom (dof) revolute joints (bond angles and lengths are fixed). Thus, the base atom requires 6 dof and each additional torsional movement requires 1 dof.

The joint angle of a revolute joint takes on values in $[0, 2\pi)$, with the angle 2π equated to 0, which is naturally associated with a unit circle in the plane, denoted by S^1. Assuming some position and orientation for one of the links (the base), the positions of each of the remaining links can be specified by the *joint angle* between the link and some adjacent link. Thus, a *configuration* of an n joint tree-like articulated object can be specified by 6 dof for the base atom, and $n-1$ joint angles for the remaining atoms. That is, the configuration space of interest for our ligands can be expressed as:

$$\mathcal{C} = \{q \mid q \in R^3 \times S^3 \times S^{n-1}\}. \tag{1}$$

Note that $\mathcal{C}$ simply denotes the set of all possible configurations, but says nothing about their feasibility. The validity of a point in $\mathcal{C}$ will be determined by potential energy computations.

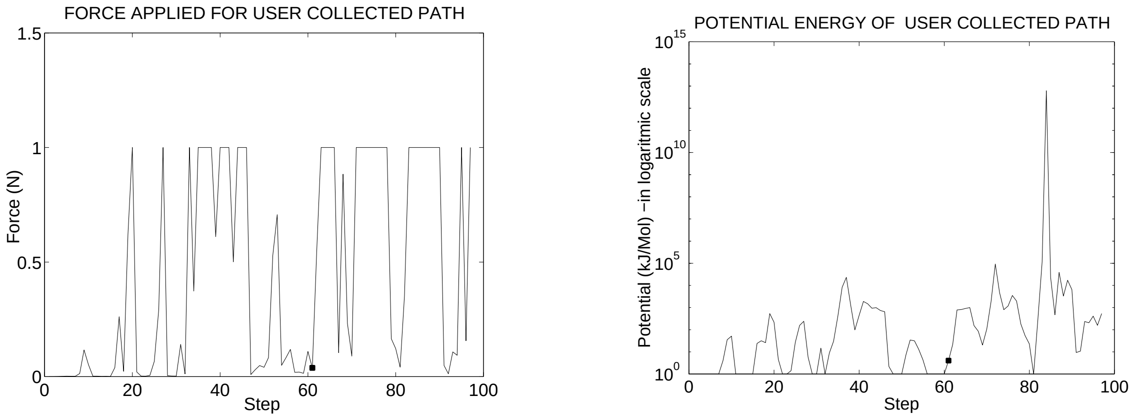
4 Potential Energy

A key requirement in simulating the ligand binding process is to have a good model of the interaction between a ligand and its receptor, a protein. During binding, the ligand can be viewed as moving in the potential field created by the protein's atoms, searching for a stable low potential configuration (often called conformations). The potential has many components such as van der Waals potential, Coulomb potential, etc. Things are further complicated by the solvent effect due to the surrounding fluid (mainly water), which is difficult to model efficiently. Thus, a complete, accurate model of the potential is not feasible. Instead, we approximate the potential by the van der Waals terms only. While this is an extreme simplification, the results obtained are quite impressive indicating that perhaps the van der Waals interaction plays the main role in the binding process. In particular, the potential we use is

$$U = \sum_{atom\ pair\ i,j} (A/r_{ij}^{12} - B/r_{ij}^{6}),$$

where r_{ij} is the distance between protein atom i and ligand atom j, and the parameters are taken from [16].

Even with our simplified potential, computing the potential energy of a ligand configuration is expensive since it includes a contribution for each atom pair. To optimize this computation we use a grid-based energy calculation. In this approach, the protein is covered by a three-dimensional grid, and each grid point stores the potential applied by all protein atoms at that point [27]. During potential energy calculations, the value stored at the closest grid point to each ligand atom is used as an approximation of potential for that atom. The accuracy of this method directly depends on the size of the grid. Our experiments indicate that a 0.5 A grid is sufficient for realistic force feedback for our haptic device.



(a) (b)

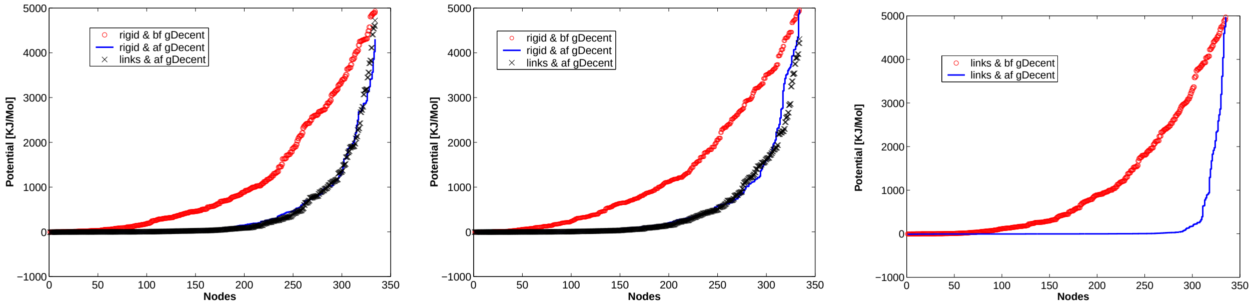
Figure 3: Correlation between (a) user-felt force and (b) potential energy

5 Haptic Interaction for Molecular Docking

Our prototype system consists of a PHANToM haptic device (6 dof) output) [18], and a Xeon 550Mhz NT Workstation. The operator uses the PHANToM to manipulate a *rigid* ligand around a virtual protein.

A challenge with the molecular docking is to compute the potential energies of the configurations fast enough for haptic interaction. To achieve this, we used a grid-based force calculation as described above, which was also used in [23]. Figure 3(a) shows the force (based on a grid) felt by the user while collecting a path of ligand configurations. The binding configuration is represented by a small rectangle. To stay in the PHANToM's force feedback limits, any computed force greater than 1N was reduced to 1N without changing its direction, which resulted in a repulsive force in the correct direction. Figure 3(b) shows the real potential energy at the sampled configurations (y-axis is logarithmic). The correlation between the figures shows the user feels low repulsive forces when the real potential energy is low. Note that at the binding configuration, the potential energy is very small and the user felt almost no force, hence with visual and haptic feedback the user can estimate the position of a binding site. A notable observation is that there are some low potential configurations near the binding configuration which actually represent a corridor which the ligand must pass through to reach the binding site. Most other configurations with small potentials were outside the protein.

Our haptic system also lets the user investigate a path or roadmap generated by the automated planner. The user can visualize a path, select a configuration in the path, and investigate the potential field around that configuration. Similarly, the user can visualize a roadmap. Roadmap configurations are represented by points in the three-dimensional workspace and roadmap edges by lines connecting them [2]. The user can investigate a specific roadmap configuration by moving the haptic pointer close to a roadmap configuration causing the configuration to pop up visually. To aid location of roadmap configurations, there is an option which pulls the user to the nearest roadmap configuration.



(a) (b) (c)

Figure 4: The (sorted) potential distribution of all roadmap nodes before and after gradient decent for three protein-ligand complexes (given by pdb identifier): (a) 1LDM, (b) 1A5Z and (c) 1STP.

6 Generating Binding Site Candidates

Good binding configurations are thought to be configurations at local minima in the potential landscape, and moreover, should be surrounded by high potentials so that the ligand will tend to stay in the site once it gets there [29].

Our goal is to automatically generate such configurations of the ligand. Our strategy is to first generate nodes near the protein's surface, and then push the generated nodes to their local minima. Our approach follows the PRM paradigm as described in Section 3. Nodes can be generated automatically with our OBPRM planner, or manually by the user with the haptic device. Next, these nodes are connected into a roadmap that will be used to determine the accessibility of a site (this is very useful for eliminating 'decoy' nodes that are inaccessible, but otherwise appear to be good candidates). Finally, we will compute a weight for all low potential configurations contained in large connected components of the roadmap (see Section 7). Details of each step are discussed below.

Automatic Node Generation. OBPRM, ostacle-based PRM, is designed to generate contact and surface nodes. We model each protein atom as a unit cube (1 Å), and each ligand atom as a sphere with 1 Å radius. Considering only the base atoms of the articulated ligand model, we generate a collision-free (geometrically) node for the base. Then, we assign random values for the joint angles. Next, we evaluate the potential of the ligand's configuration and keep the node only if its potential is less than some fixed value $E_{\max}$. A fairly high value of 5,000 KJ/Mol was chosen for $E_{\max}$. However, after applying gradient descent to the nodes (described below), most remaining nodes have potentials below zero, as can be seen in Figure 4.

User Generated Nodes. We also experiment with operator collected nodes using a visual and haptic interface. Our previous work with CAD models has shown significant improvement when user input is given to the automatic planner [2]. A user would collect a node when he/she felt like it was local minimum in the potential and also saw it was in some sort of pocket on the protein. The ligand was restricted to a rigid body for the haptic interaction.

Approximate Gradient Decent. Since a binding configuration is commonly thought to be a local minima surronded by higher potentials, it is natural to push our generated nodes to local minima. In particular, we perform an approximate gradient decent as follows. For each generated node, we sample uniformly n nearby nodes, and select the lowest potential new node for iteration; we chose $n = 2d$, where d is the degree of freedom for the ligand. This process is repeated for K iterations, or until a local minima is reached (when all sampled nodes have higher potential); a local minima was usually obtained in 2-4 iterations.

Node Connection. After gradient decent is performed, all local minima nodes are added to the roadmap (with the previously generated nodes). Then, we perform node connection as is normally done for PRMs, i.e., we attempt to connect each node to its k nearest neighbors along the straight-line in C-space between them (we use $k = 10$). An edge between the node and its neighbor is added to the roadmap if it passes some validity check. The only difference here is that our validity check uses potential evaluation instead of collision detection.

7 Recognizing Binding Sites

Automated methods, including ours, need to characterize and rank (score) configurations in terms of their potential as binding configurations. This is needed both to identify configurations in the binding site, and to determine promising configurations for further processing during the computation. There exist as many scoring functions as docking algorithms, including empirical functions [5, 21, 12], knowledge based functions [22], and force field functions [17, 6].

In the PRM based method presented in [29], they first chose the top 10 candidates as the 10 roadmap nodes with with lowest potential, and then further evaluated them with a score function called average path weight. They successfully found the binding sites for two of the three examples they studied.

Our approach in this paper is similar to theirs, but has two significant differences. First, we cannot restrict our attention to such a small set of candidates that are sorted based on potential. This is because, as shown in Figure 4, after the gradient decent, our roadmaps contain many nodes with low potential energy. Since the approximation used for the potential can easily change the relative order of the nodes, selecting the nodes with lowest potential can easily cause us to miss configurations in the true binding sites. Therefore, we must further evaluate *all* the nodes in the flat region of the potential. However, one factor that can be used to filter nodes is accessibility. As mentioned above, we use the roadmap's connectivity to filter out inaccessible configurations, and in particular, only evaluate nodes contained in the largest connected component of the roadmap.

Second, we propose a new weight function. Since the binding site is a *local* feature, we observe there is no need to evaluate path weights globally as was done in [29]. Instead, we propose building a 'local roadmap' for each candidate node and evaluating path weights in it. In particular, we first uniformly sample n nodes within a fixed distance r from the candidate node, and then define the weight, or score, of the candidate as the average potential of all the sampled nodes, i.e.,

$$Weight(c) = \frac{\sum_{i=0}^{n} P(c_i)}{n},$$

where c_i is a random configuration sampled near c. The potential, $P(c_i)$, is truncated to be 5,000 KJ/Mol if it has a larger value. We believe this is a better indication of a configuration's fitness as a binding configuration since it densely samples a site and should be a better measure of its local properties. Figure 5 shows that the weight function is much sharper than the potential for ranking the nodes.

8 Experiments

We tested our approach on three different protein-ligand complexes obtained from the Protein Data Bank at `http://www.rcsb.org/pdb/`, which provides atom positions for the protein and the

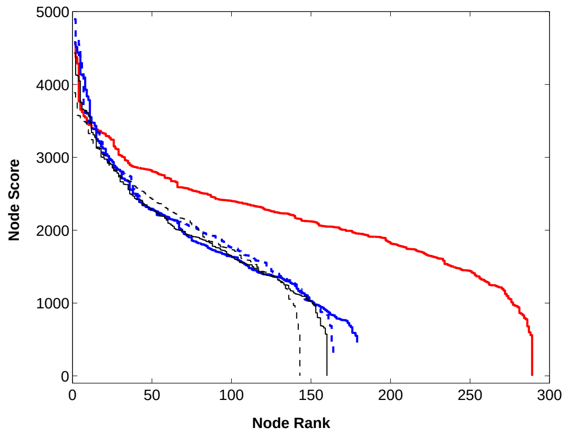


Figure 5: The (sorted) weight/score distribution of all the nodes with potential below 50 KJ/Mol. The five curves, from left to right, are the weight for: 1LDM rigid model, 1LDM articulated model, 1A5Z rigid, 1A5Z articulated, and 1STP articulated. 1LDM, 1A5Z and 1STP are the three protein pdb files used in our experiments, see section 8 for more details.

binding configuration of the ligand. Two of these complexes, 1LDM and 1STP, were studied in [29]. The third, 1A5Z, was chosen based on the complexity of protein's potential field which created many isolated areas. Some details of these complexes are shown in Table 2.

In our experiments we investigate the following questions:

- Is the most widely used simplification of the problem, i.e., the approximation of the ligand molecule as a rigid object sufficient?
- Can an OBPRM based automated planner identify binding sites?
- Can a user provide helpful information to an automated planner?

To answer these questions we perform three different experiments.

In the first experiment, the ligand is treated as a rigid body. The planner generates configurations, and finds and ranks potential binding sites using these configurations.

In the second experiment, the user collects configurations using the PHANToM by manipulating a rigid ligand. The automated planner then fine-tunes these configurations, now treating them as articulated bodies, and evaluates the possible binding sites.

In the third experiment, we investigate the behavior of the automated planner when the ligand is flexible, i.e., articulated.

The running time and roadmap statistics are shown in Table 1. The number of configurations the user collected varies and can be found in Table 2.

8.1 Results

A summary of our results is shown in Table 3. For each experiment for a protein-ligand complex, we list the top five scoring configurations identified. The first column represents the rank of the configuration. The 0 ranked configuration is the binding configuration for the respective complex. The RMSD column shows the Root Mean Square Deviation with respect to the known binding site. The score is the score of that configuration computed as described in Section 7. The last column represents the potential energy of the complex at that configuration. Based on the results of the

Roadmap Statistics						
Model	Gen. (sec)	Conn. (sec)	Total (sec)	# node(init #, # local min added)	# edge	bigCC
1A5Z Rigid	55	70	230	1429 (1000, 429)	3784	1182
1A5Z User	75	101	199	1684 (1666, 18)	3045	1304
1A5Z Artic.	75	203	658	2417 (1666, 751)	6253	2071
1LDM User	853	141	279	2011 (2000, 11)	4141	1643
1LDM Artic.	85	299	1089	2912 (2000, 912)	8472	2579
1STP Rigid	104	369	822	1771(1000, 771)	5589	1738
1STP User	162	163	343	1674 (1666, 8)	3115	1504
1STP Artic.	162	447	2095	3020 (1666, 1354)	7971	2909

Table 1: Roadmap statistics. Total time includes node generation time, connection time, and node score calculation time. 1000-2000 nodes are generated for each case to ensure a potential binding site node is accessible from outside. Then all nodes are pushed to local minina and those with potential less than 50 KJ/Mol are added in roadmap, the numbers are shown. bigCC indicates the number of nodes in the largest connected component. All times are in seconds.

Pair ID	***Protein***		***Ligand***			**# Cfgs user collected**
	Name	**# Atom**	**Name**	**#Atom**	**dof**	
1LDM	M_4-Lactate Dehydrogenase	2544	Oxamate	6	7	37
1A5Z	L-Lactate Dehydrogenase	2416	Oxamate	6	7	129
1STP	Streptavidin	1001	Biotin	16	11	92

Table 2: Statistics for different protein-ligand complexes. The table also shows number of configurations that are collected by the user during the node generation phase.

rigid body experiments for 1A5Z and 1STP (Table 3(a) and 3(c)), we did not consider the rigid body representation for 1LDM.

As can be seen, in almost all cases, at least one the top five scoring configurations is very close (2-3 Å) to the true binding configuration. This suggests that not only is our automated planner effective at finding configurations close to the binding site, but also that our scoring function works well, giving close configurations high scores. In addition, our success in identifying the binding site for 1STP shows an advantage of our OBPRM based approach over the results presented in [29], where the closest configuration identified was 13 Å from the true binding site.

When we compare the automated results for the rigid body and articulated representations, for 1A5Z and 1STP in Table 3, it is clear that better results were obtained using the articulated representation, which did incur some additional computational costs as seen in Table 1. The articulated representation was slightly better for 1A5Z, where the difference in the degrees of freedom between the rigid and articulated versions is small. When these differences increase the performance gap increases, as with 1STP. In general, since there are many local minima in the protein's potential field, movement is hard for a rigid ligand. However, an articulated ligand can move easily in these areas by taking on the shape of the surrounding potential surface. Our comparison of the rigid and articulated representation is potentially relevant for other docking methods, since, as mentioned in Section 2, representing the ligand as a rigid body is a commonly used simplification.

The results for the fully automated articulated model and those starting from user generated nodes are comparable. In 1A5Z and 1LDM the user was able to find the closest configurations to the binding site. However, in 1STP the automated planner reached a closer configuration. This may be related to the fact that the user has to work with a rigid body version of the ligand which decreases the efficiency, while the automated planer takes advantage of the high-dimensional space. The large difference in the performance of the rigid body and the articulated body representations for the 1STP complex supports this idea.

An advantage of user input is that, as shown in Table 1, it was noticably faster than the fully automated method because the user provided a much smaller sample of nodes. Fewer nodes were provided since the user only collected nodes in/near pockets. A disadvantage is that the user input suffers the users' bias. It is very hard (tedious) for the user to explore every corner of the protein to find a binding site, therefore he/she may miss the true binding site. In addition, a user may be misled by some 'decoy' sites based on only visual and/or force feedback. This is shown is the central column of 1A5Z, where several nodes the user collected seem to lie in a site 15-16 Å away from the true binding site, even though the user also identified configurations in the binding site.

1A5Z									
	RIGID			**USER INPUT (RIGID)**			**ARTICULATED (7dof)**		
Rank	RMSD(Å)	Score	Pot.	RMSD(Å)	Score	Potential	RMSD(Å)	Score	Pot.
0	**0.00**	**3743**	**35.51**	**0.00**	**3743**	**35.51**	**0.00**	**3743**	**35.51**
1	5.74	4694	0.09	2.08	4138	-8.783	5.82	4658	19.92
2	5.20	4462	24.81	1.94	3670	-5.19	7.31	4488	41.43
3	4.10	4422	0.81	15.90	3316	-5.209	2.31	4459	13.67
4	9.27	4376	30.36	16.89	2682	22.207	5.63	4393	-9.59
5	4.35	4290	-7.79	16.11	2413	-5.95	2.42	4342	19.88

(a)

1LDM						
	USER INPUT (RIGID)			**ARTICULATED (7dof)**		
Rank	RMSD(Å)	Score	Pot.	RMSD(Å)	Score	Potential
0	**0.00**	**4570**	**5.00**	**0.00**	**4570**	**5.00**
1	1.52	4652	-6.06	5.03	4634	30.96
2	2.93	4564	-6.73	2.03	4593	23.35
3	3.79	4541	-7.04	3.07	4577	48.12
4	2.35	4499	-6.88	4.14	4475	20.95
5	2.11	4463	-8.70	1.66	4418	43.34

(b)

1STP									
	RIGID			**USER INPUT (RIGID)**			**ARTICULATED (11dof)**		
Rank	RMSD(Å)	Score	Pot.	RMSD(Å)	Score	Pot.	RMSD(Å	Score	Pot.
0	**0.00**	**4552**	**-5.21**	**0.00**	**4552**	**-5.21**	**0.00**	**4552**	**-5.21**
1	4.21	4127	48.79	5.50	3431	49.22	2.56	4609	0.37
2	11.40	3934	-12.73	3.77	3426	11.10	5.05	4284	36.44
3	10.63	3921	7.94	7.32	3425	-6.47	3.44	4151	-5.79
4	12.42	3894	-10.51	4.99	3287	48.80	4.11	4014	18.31
5	11.29	3876	38.19	6.90	3211	44.05	7.41	3952	22.68

(c)

Table 3: Performance of the scoring function on each protein. The tables show the results of three different kind of experiments; OBPRM with the ligand as a rigid body, OBPRM with the ligand as an articulated body and user collected configurations. For 1LDM, we only experimented on OBPRM with the articulated ligand and user collected configurations.

An interesting observation was that, in the vicinity of the binding site the user could feel a cavity area and there were tunnels that the user could feel leading toward the binding site.

Finally we remark that our energy filtering keeps many nodes that are in the flat region of the potential distribution plot. Therefore, we avoid throwing away nodes near binding sites even though their potential may be slightly higher, which could very easily be caused by inaccuracy of our potential function.

9 Conclusions and Future Work

Our results show that our approach to molecular docking is promising. In the examples we studied, we were able to find and recognize configurations in the true binding site. However, further processing, perhaps with a more accurate potential function, will be needed to find an exact binding configuration. For example, candidates identified by our our method might be used as input for other docking programs that perform detailed simulations, such as molecular dynamics methods. We also saw that better results were obtained using an articulated representation for the ligand, as opposed to the commonly used rigid body simplification. User input was seen to improve efficiency, and moreover, haptic feedback was observed to help the user better understand the problem.

Our future work includes finding a better representation for potential energy formulations, concentrating on the binding sites and reaching the binding configurations, and improving the haptic interface by letting the user collect configurations for a flexible ligand.

Acknowledgements

We would like to thank Stan Swanson for his help.

References

[1] N. M. Amato, O. B. Bayazit, L. K. Dale, C. V. Jones, and D. Vallejo. OBPRM: An obstacle-based PRM for 3D workspaces. In *Proc. Int. Workshop on Algorithmic Foundations of Robotics (WAFR)*, pages 155–168, 1998.

[2] O. B. Bayazit, G. Song, and N. M. Amato. Enhancing randomized motion planners: Exploring with haptic hints. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 529–536, 2000.

[3] Jurgen Brickmann, Wolfgang Heiden, Horst Vollhardt, and Carl-Dieter Zachmann. New man-machine communication strategies in molecular modeling. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 273–282, 1995.

[4] R.L. DesJarlais, R.P. Sheridan, R. Nilakatan, K.S. Haraki, N. Bauman, and R. Venkataraghavan. A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. *J. Computer-Aided Molecular Design*, 8:231–242, 1994.

[5] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, and R.P. Mee. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Computer-Aided Molecular Design*, 11:425–445, 1997.

[6] T.A. Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of Computational Chemistry*, 17:490–519, 1996.

[7] Erica Harvey and Chaim Gingold. Haptic representation of the atom. In *Proceedings of IEEE International Conference on Information Visualization*, pages 232–235, 2000.

[8] D. Hsu, L. Kavraki, J-C. Latombe, R. Motwani, and S. Sorkin. On finding narrow passages with probabilistic roadmap planners. In *Proc. Int. Workshop on Algorithmic Foundations of Robotics (WAFR)*, 1998.

[9] G. Jones, P. Willett, R. C. Glen, A. R. L. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. of Molecular Biology*, 245:43–53, 1985.

[10] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, and C. Aflalo I.A. Vakser. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. In *Natl. Acad. Sci. USA*, volume 89, pages 2195–2199, 1992.

[11] L. Kavraki, P. Svestka, J. C. Latombe, and M. Overmars. Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Trans. Robot. Automat.*, 12(4):566–580, August 1996.

[12] I.D. Kuntz. Structure-based strategies for drug design and discovery. *Science*, 257:1078–1082, 1992.

[13] J. C. Latombe. *Robot Motion Planning*. Kluwer Academic Publishers, Boston, MA, 1991.

[14] Jonathan Leech, Jan F. Prins, and Jan Hermans. Smd: Virtual steering of molecular dynamics for protein design. *IEEE Computational Science and Engineering*, pages 38–45, 1996.

[15] D. Levine, M. Facello, P. Hallstorm, G. Reeder, B. Walenz, and F. Stevens. Stalk: An interactive system for virtual molecular docking. *IEEE Computational Science & Engineering*, April-June:55–66, 1997.

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

[17] M. Liu and S.M. Wang. Mcdock: A monte carlo simulation approach to the molecular docking problem. *Journal of Computer-Aided Molecular Design*, 13(5):435–451, 1999.

[18] T. H. Massie and J. K. Salisbury. The PHANToM haptic interface: A device for probing virtual objects. In *Int. Mechanical Engineering Exposition and Congress, DSC 55-1*, pages 295–302, Chicago, 1994. C. J. Radcliffe, ed., ASME.

[19] E.F. Meyer, S.M. Swansan, and J.A. Williams. Molecular modelling and drug design. *Pharmacology & Therapeutics*, 85:113–121, 2000.

[20] M.D. Miller, S.L. Kearsley, D.J. Underwood, and R.P. Sheridan. Flog: A system to select 'quasi flexible' ligands complementary to a receptor if known three-dimensional structure. *J. Computer-Aided Molecular Design*, 8:153–174, 1994.

[21] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. *J. Computational Chemistry*, 19:1639–1662, 1998.

[22] I. Muegge, Y.C. Martins, P.J. Hajduk, and S.W. Fesik. Evaluation of pmf scoring in docking weak ligands to the fk506 binding protein. *J. Medicinal Chemistry*, 42(14):2498–2503, 1999.

[23] M. Ouh-young. *Force Display in Molecular Docking, TR90-004*. PhD thesis, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1990.

[24] M. Ouh-Young, M. Pique, J. Hughes, N. Srinivisan, and F. P. Brooks, Jr. Using a manipulator for force display in molecular docking. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, volume 3, pages 1824–1829, 1988.

[25] Ming Ouh-young, David V. Beard, and Frederick P. Brooks, Jr. Force display performs better than visual display in a simple 6-d docking task. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, pages 1462–1466, 1989.

[26] Ruth Pachter and James A. Lupo. Virtual reality for materials design. In *The 14th Annual Dayton Section Symposium AESS/IEEE*, pages 51–56, 1997.

[27] N. Pattabiraman, M. Levitt, T. E. Ferrin, and R. Langridge. Computer graphics in real time docking with energy calculation and minimization. *J. of Computational Chemistry*, 6:432–436, 1985.

[28] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Molecular Biology*, 261:470–489, 1996.

[29] A.P. Singh, J.C. Latombe, and D.L. Brutlag. A motion planning approach to flexible ligand binding. In *7th Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*, pages 252–261, 1999.

[30] Mark C. Surles. *Techniques for Interactive Manipulation of Graphical Protein Models*. PhD thesis, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1992.

[31] M.C. Surles, J.S. Richardson, D.C. Richardson, and F.P. Brooks, Jr. Sculpting proteins interactively: Continual energy minimization embedded in a graphical modeling system. *Protein Science*, 3:198–210, 1994.

[32] Maxim Totrov and Ruben Abagyan. Derivation of sensitive discrimination potential for virtual ligand screening. In *Third Annual International Conference on Computational Molecular Biology*, pages 312–320, 1999.

[33] Len Wanger. Haptically enhanced molecular modeling: A case study. In *Proceedings of the Third PHANTOM Users Group Workshop*, 1998.