Using Motion Planning to Rank Ligand Binding Affinity

Hsin-Yi (Cindy) Yeh

I. INTRODUCTION

In pharmaceuticals, the screening process rigorously tests potential drugs to select the most promising protein candidates. Since the drug discovery process is costly [2], computationally designing and screening the drug becomes important nowadays. The drug (or ligand) forms bonds with the protein yielding a complex and activating a lot of biological mechanisms. The process that a ligand attaches to a protein is called ligand binding and the binding affinity is the strength of this protein-ligand interaction. The higher the binding affinity, the more stable the protein-ligand complex.

Since ligand binding is a computationally intensive problem which studies molecular motion, motion planning has been applied on this study. The motion planning problem is to find a valid path for a movable object (robot) from a start position to a goal position. This problem has been extensively studied with randomized sampling-based planners by approximating the space in a graph (PRM) [4] or a tree (RRT) [5]. PRM constructs a roadmap by sampling randomly from configuration space ($C_{space}$) and retaining the valid ones. Since PRMs struggles in the problems with narrow passages, many variants focus sampling on certain portions of $C_{space}$, such as near obstacle surfaces or along the medial axis. However, none of them make any guarantee about the sampling distribution in these targeted regions of $C_{space}$. A new framework which can uniformly sample any region type of $C_{space}$ is developed by analyzing some fixed length line segments and then performs some simple checking along them. UOBPRM [12] and UMAPRM [11] both come from this uniform sampling framework as specific applications by their special checking along the line segments.

Since UOBPRM generates uniformly distributed samples close to the obstacle surfaces, it provides possible ligand configurations in the binding site. In this work, we model the ligand as a linkage robot and the protein as an obstacle similar to [1] and use UOBPRM to generate ligand samples to approximate binding affinity by two metrics. Our results show that UOBPRM is a potential technique for ranking ligand binding affinity and this ranking can be useful in computational drug screening.

II. PRELIMINARIES AND RELATED WORK

A. Uniform Sampling

A robot can be described by $n$ parameters (degrees of freedom, DOFs) and each parameter represents an object component, such as object position and object orientation. $C_{space}$ is an $n$-dimensional space where all feasible robot configurations form $C_{free}$ and $C_{obst}$ is the union of all infeasible configurations.

A methodology that uniformly samples any specific region type of $C_{space}$, e.g., near obstacle surfaces or along the medial axis, is developed as long as the membership can be determined by checking properties of consecutive points along a line segment. A set of uniformly distributed fixed length line segments are generated first. An analysis method is then applied on the line segment and the roadmap nodes are identified by some consecutive checking along the line segment. UOBPRM [12] uses the uniform sampling framework to generate configurations around $C_{obst}$ surfaces by checking the validity changes between neighboring configurations. UMAPRM [11] monitors the closest obstacle changes between configurations along the line segment to generate uniformly distributed configurations along the medial axis.

B. Ligand Binding

The study of protein-ligand interaction is essential to understand many biological mechanisms. Ligand is a small protein which attempts to bind to a binding site on the target protein, as shown in Figure 1. Binding is the process that ligand finds a specific transformation attached to a protein. If the ligand cannot fit in the binding site, it cannot bind to the target protein and no biochemistry process is going to happen. The interaction strength between the protein and the ligand is called binding affinity.

![Fig. 1. A protein (in sticks mode) and a ligand (in spheres mode) bound inside.](image)

There are some software which use Monte Carlo and Genetic Algorithms to predict protein-ligand binding, such as AutoDock [8] and FTDock [3]. But none of them considers the ligand as fully flexible [10]. Moreover, the exploration space is very large and the simulation time is extremely long. Recently, there are approaches which use motion planning to study protein-ligand binding [1], [9]. In [9], they first uniformly generate configurations in $C_{space}$ then sample denser in the binding site. A low-potential path is extracted from the roadmap to represent the ligand binding path. An improvement by OBPRM and haptic device is proposed in [1]. The results show that OBPRM is able to identify the binding site with the aid from the haptic device.
TABLE I
APPROXIMATED BINDING AFFINITIES AND THE PUBLISHED BINDING AFFINITIES COMPARISON FOR 3W6H.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Affinity(Ki)</th>
<th>Avg Distance</th>
<th>Min Distance</th>
<th>Avg Energy</th>
<th>Min Energy</th>
<th>Published Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5×10⁻⁴</td>
<td>13.01</td>
<td>0.51</td>
<td>328.60</td>
<td>-0.4008</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>12×10⁻⁴</td>
<td>16.94</td>
<td>0.94</td>
<td>43.64</td>
<td>-0.4037</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>200×10⁻⁴</td>
<td>29.72</td>
<td>29.72</td>
<td>7850.00</td>
<td>-0.4132</td>
<td>3</td>
</tr>
</tbody>
</table>

III. METHODS

Given a protein and a set of ligands, we rank the ligands based on how well they can bind to the protein. The protein is assumed to remain mostly static so it is treated as an obstacle and the ligand is modeled as a linkage robot. For each ligand, UOBPRM first generates samples around protein surfaces which provides samples close to the binding site. Some affinity metrics are then applied to the samples to approximate the ligand binding affinity.

We use two different affinity metrics to measure the binding affinity. One metric is the distance between the center of mass of the ligand and the center of mass of the protein. Both the average and the minimum are collected. We assume that if the binding affinity is higher for a ligand, it is more likely to be buried deeper in the protein. Thus, the smaller the distance, the higher the affinity.

Another affinity metric is the potential between the protein and the ligand. It is calculated by the following equation where $r_{ij}$ refers to the distance between protein atom $i$ and ligand atom $j$. Parameters $A$ and $B$ are taken from [6].

$$U = \sum_{\text{atom pairs}, j} \left( \frac{A}{r_{ij}^6} - \frac{B}{r_{ij}^6} \right)$$

The average and the minimum are collected again. Since the ligand searches for a stable low potential configuration during binding, the lower the energy, the higher the affinity.

IV. EXPERIMENTS

We obtain the protein-ligand pairs from BindingDB [7] with known experimentally determined binding affinities. BindingDB is a public database which provides ligands with different binding affinities for a specific target protein. For each protein-ligand pair, we approximate the binding affinity by the affinity metrics discussed in the previous section. Our simulation results are then compared to the published rank.

Table I displays the comparison between the approximated and published binding affinity for three protein-ligand pairs. Here we generate 100 UOBPRM samples for each protein-ligand pair. We find that both average and minimum distance are able to capture the affinity ranking correctly. However, there is no trend on average energy even increasing the UOBPRM sample size. When increasing the number of UOBPRM samples, we also find that the same configuration is always captured for the minimum energy. We approximate the potential by the van der Waals interaction only which excludes many components such as the indirect solvent effects. This can be one reason that the potential here is not a good metric to rank the affinity.

V. CONCLUSION

In this work, we use a novel obstacle-based sampling method, UOBPRM, to study the binding affinity problem. We experiment on 3 protein-ligand pairs by 2 affinity metrics and find that UOBPRM is able to approximate the binding affinity by the distance between the ligand and the protein. In the future, we plan to study a better potential calculation and more binding affinity metrics, e.g., the contact between the ligand and the protein. We also want to combine several metrics with weights to rank the affinity.

ACKNOWLEDGMENT

Many thanks to Dr. Nancy M. Amato, my advisor, and Dr. Shawna Thomas for their guidance and supervision.

REFERENCES