Optimizing Low Energy Pathways in Receptor-Ligand Binding with Motion Planning

Torin Adamson
Department of Computer Science
University of New Mexico
Albuquerque, USA
toriadam@cs.unm.edu

Julian Antolin Camarena
Department of Physics and Astronomy
University of New Mexico
Albuquerque, USA
jantolin@unm.edu

Lydia Tapia
Department of Computer Science
University of New Mexico
Albuquerque, USA
tapia@cs.unm.edu

Bruna Jacobson
Department of Computer Science
University of New Mexico
Albuquerque, USA
bjacobson@unm.edu

***Abstract*—Determination of ligand binding pathways is an important factor to predict drug efficacy in drug discovery. Ligand-receptor binding involves the motion of many degrees of freedom, which can make binding pathways difficult to discover with traditional methods. Interactive molecular docking tools can allow users to explore the high dimensional energy landscape of the ligand-receptor system with rigid molecular models to determine low energy ligand states and pathways to binding. To introduce the effect of ligand flexibility in molecular docking with rigid body models, we use ensembles of distinct ligand conformation states that can be swapped during exploration. Our method emulates ligand flexibility effects in rigid body docking at no extra computational cost. Our automated method simulates user search performance with a path optimization algorithm. We find that allowing the algorithm to include different ligand conformations in its search for states of lower energy can result in optimized low energy pathways with reduced search times in difficult areas near energy barriers. This method can be adapted to include molecular flexibility effects in interactive rigid body molecular docking running in commodity hardware, such as molecular docking games.**

***Index Terms*—molecular docking, receptor-ligand binding, motion planning, ligand flexibility**

I. Introduction

Ligand binding pathways are biologically feasible lowest energy routes for a free ligand to reach the ligand-receptor bound state in the high-dimensional energy landscape of a multi-molecular system. Determining ligand binding pathways is a central problem in drug discovery. The efficacy of a drug may depend not only on the affinity of the ligand-receptor interaction, but also on the time scales involved with binding events and on the molecular conformational changes that occur during binding [1], [2].

The large number of degrees of freedom of the molecular system makes the search for ligand binding pathways computationally expensive [3]. Therefore, to perform computationally efficient global explorations of the interaction energy landscape of receptor and ligand, the degrees of freedom of the problem need to be reduced. Interactive molecular docking programs achieve this by keeping ligands and receptors rigid, which reduces the computational cost of energy calculations [4]. While the problem is oversimplified by fixing the internal degrees of freedom of the molecules, much can be gained by adding a human operator who is able to use sensorial cues such as visual and haptic feedback to aid the high-dimensional search of low energy ligand states around the receptor [4]–[6]. To allow realtime visual and tactile feedback in interactive molecular docking, energy and force calculations need to be performed quickly. The data generated by users as they manipulate the ligand around the receptor can be later used to find low energy ligand pathways [7], [8].

Gamification of interactive molecular docking can enable the collection of large crowdsourced datasets of ligand conformations [7], [8]. As the number of users grow, so does the probability of finding ligand states near the binding site on the receptor, possibly leading to improved pathways. Molecular docking games incorporate game design features (eg. score, leaderboard) to help users explore the ligand-receptor interaction energy landscape to find new low energy ligand states [7]–[9]. The need for computationally efficient interactive molecular docking is greater if these games are distributed to players who mostly use commodity hardware [7], [8].

As molecules generally change conformations during binding events, pathways determined from rigid body models may miss critical information about binding. In particular, close to the binding site, rigid body docking can sample high energy states representing atomic collisions. Since such states are not physical, molecular flexibility is critical to prevent atoms from overlapping. However, implementing molecular flexibility in interactive molecular docking can require specialized hardware [5]. Moving the internal degrees of freedom of

978-1-7281-1867-3/19/$31.00 ©2019 IEEE

the molecules can slow down energy calculations, making interactive docking in commodity hardware prohibitive. Therefore, implementation of computationally expensive features in molecular docking games, such as molecular flexibility, requires novel approaches.

To optimize the search of low energy pathways of ligand binding, in this paper we demonstrate a proof-of-concept for mimicking ligand flexibility in rigid body molecular docking that can run efficiently in commodity hardware. Ligand flexibility is introduced as an ensemble of distinct rigid body ligand conformations. This ensemble includes ligand conformations that are bond rotated from the original docked ligand conformation. The ability to select different rigid body models of the ligand during docking adds complexity to the data at little computational cost. Resulting pathways can incorporate many different ligand conformations to overcome energy barriers while docking. Here we test the feasibility of this method for molecular docking. The path optimization method is summarized in Figure 1. We start with a roadmap made of simulated data consisting of thousands of random ligand samples (in random positions and orientations) around the known binding site on the receptor. Once a low energy path is found (Figure 1(a)), the roadmap is updated to exclude unused samples and to include new samples generated around the path (Figure 1(b)). A new, lower energy path is then determined from the new roadmap (Figure 1(c)), and this is done iteratively until there is no improvement in the pathways, or no new low energy ligand states are found (Figure 1(d)). This automated motion planning path optimization technique aims to emulate user exploration: Users could tend to over-explore near previously found low energy states. This iterative search and improvement on the initial low energy pathway explores more densily the region around the original path. We find indications that extending rigid body state search with an ensemble of distinct ligand conformations may allow more efficient path optimization in difficult areas of the search space while maintaining the same computational cost as searches with a single conformation. The benefits of ligand conformation ensembles in interactive molecular docking will be examined in future studies with user participation.

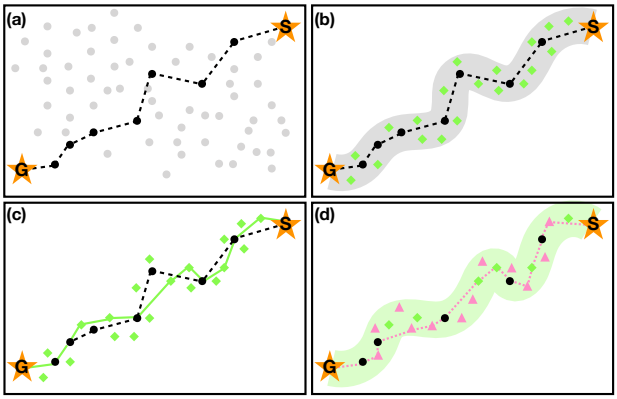


Fig. 1: Schematics of our pathway optimization method. (a) Initial ligand samples (gray circles) and start and goal states (stars) are added to the roadmap (edges not shown). A low energy path (dashed black line) is found that connects the start state to goal state. Ligand states along the path are shown in black. (b) First optimization iteration: Ligand states not included in the initial path are discarded. New ligand states (green diamonds) are sampled in a volume around the original path (shaded gray area) and a new roadmap is created. (c) A new low energy path is found (green solid line) for the updated roadmap that may include samples in the original path (black circles) and new samples (green diamonds). (d) The procedure depicted in steps (b) and (c) is repeated: unused states are discarded, new ligand states are created around the green path (pink triangles in shaded green area), the roadmap is updated and and a new path is found (dotted pink line). This is done until there is no improvement on the path.

II. Related Work

A. Simulation of Ligand-Receptor Kinetics

Proteins rely on molecular flexibility to accomplish their function and to react to their environment [10]. Receptor proteins and ligands may change conformation during binding events. In principle, their interactions and subsequent motions can be simulated directly in molecular dynamics (MD) simulations. Molecular dynamics is a physics-based mechanical simulation of atomic systems [11]. It uses molecular structure and atom bond connectivity information from experimental data to simulate dynamics of individual atoms, and molecular flexibility stems directly from such simulations. Since MD takes into account the dynamic motion of each atom, which is influenced by all other atoms in both the ligand and the receptor, simulation of receptor-ligand binding comes at a large computational cost, and pathways analysis that require long time scales become prohibitive.

However, classical MD simulations can still be useful in understanding ligand-receptor kinetics. Recent work has used ensembles of MD trajectories to extract Markov states that may be involved in ligand binding pathways [12]. Brownian dynamics coupled with MD methods have also been used to estimate ligand-receptor kinetics [2]. Other methods have used metadynamics techniques that apply MD force fields to explore receptor-ligand kinetics [13]. For some receptor-ligand systems, metadynamics has been able to not only confirm the existence of known pathways but also can find new ones [1].

In drug discovery, ensemble-based virtual screening can combine multiple receptor conformations to make docking site predictions [14], [15]. While in this paper we use a similar approach with an ensemble of ligand conformations, here the docking site is known but pathways to binding are unknown.

B. Interactive Molecular Docking

Interactive molecular docking combines various modes of feedback including realtime potential energy scoring to guide a human operator towards finding potentially docked states.

These docking tools can respond with haptic feedback, allowing the operator to feel the potential energy interactions between molecules [6], [16], [17]. Users can also be immersed in 3D visual feedback [18]. To enable further exploration of molecular interactions, high end graphics hardware can be utilized for realtime receptor flexibility [5]. Interactive molecular docking can be gamified as a puzzle game. A docking game can be crowdsourced to expand ligand state exploration and find better pathways [7]–[9]. This avenue has also seen success in the context of protein folding [19].

C. Motion Planning with Roadmaps

Molecular docking can be expressed as a motion planning problem where the task is to find a series of valid state transitions from an initial state to a goal state. This study uses a version of the Probabilistic Roadmap method (PRM) [20] that has been extended to support a rigid body molecular model. In a PRM, the possible state space (also known as configuration space) is sampled at random for states that are collision-free (also known as valid states). Edges are then added if they evaluate to valid state transitions. Once a roadmap has been constructed, it can be efficiently queried repeatedly to produce motion paths. PRMs have been applied to molecular motions before, notably in the context of protein folding prediction [21]. Input from a human operator can be combined with this method to inform the motion planner [22] and, more specifically, within rigid body molecular docking [16]. Preliminary studies were also done with roadmaps built from human contributed data in an interactive game environment [7]. The quality of paths resulting from roadmap methods will depend on the state samples being a good approximation of the space and useful for planning in narrow spaces full of obstacles.

D. Path Optimization

Motion paths produced by roadmap methods can be of low quality in a discrete robotic environment [23], let alone with complex potential fields as obstacles. Paths can be optimized to overcome these drawbacks according to various criteria. In robotics, this may mean smoothing out sharp angular turns [24], optimizing for sensor coverage [25] or maximizing clearance from obstacles [23]. Higher quality paths can also be obtained from PRM methods by using another more specialized sampling based method after a query has been performed [26]. Our approach is a similar two-stage method that uses a second PRM with a dense sampling focused around the path obtained from the global planner.

E. Robotics Methods in Molecular Binding and Unbinding Simulations

Molecular docking prediction problems can be investigated with motion planning methods originally designed for robotics by representing molecules as semi-flexible bodies, with any flexibility of the molecule represented as articulate joint degrees of freedom [27]. One approach views ligand motion pathway planning as a disassembly problem, using tree-based search methods to determine which flexible degrees of freedom are important in finding collision-free states away from the bound state [28], or representing atomic structures as deformable mesh to reduce dimensionality while minimizing energy [29]. Individual docked states for challenging structures can be found using an incremental assembly method [30]. These methods take advantage of algorithms originally designed to plan in high-dimensional spaces to find motions in the similarly difficult space of molecular interactions.

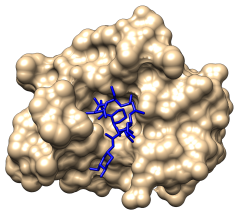


Fig. 2: Model of immunophilin-immunosuppressant molecular complex used in this study (taken from entry 1FKF from RCSB). The ligand molecule is seen in blue in its native (docked) state with the receptor (in tan).

III. Methods

A. Molecular Models

A model of the molecular immunophilin and immunosuppressant complex (PDB ID 1FKF) [31] is used as an example in this study (Figure 2). Hydrogen atoms were added to the model via the "AddH" tool in the molecular visualization and analysis software Chimera [32]. The final model of the receptor contains 1663 atoms and the ligand has 126 atoms.

We generate 9 ligand conformations by performing bond rotations in Chimera on the original native (docked) ligand conformation extracted from the PDB file. Bond rotations are performed using the "Adjust Torsions" setting. The ligand contains 129 bonds, but only up to 9 bonds were selected for rotation. Each selected bond i can be rotated by a random angle $\Delta\alpha_i$. Table I shows the atomic bond ID and $\Delta\alpha_i$ values for each of the non-native ligand conformations (labeled '2' to '10'; The native state is labeled '1'). The bottom line in Table I shows the Root Mean Squared Deviation (RMSD) values for each non-native conformation as compared to the native ligand state. The native state is not shown on the table since it has no bond rotations, and its RMSD is zero. The ensemble of ligand conformations used in this work contains 10 states, including all 9 bond-rotated ligands from Table I plus the native ligand state. By adding states with conformal flexibility, we hypothesize that rigid body collisions near the binding site can be mostly avoided.

B. All-Atom Energy Calculation

To score the quality of individual rigid body states, an all-atom intermolecular potential energy function is used. Since only electrostatic and van der Waals potentials are calculated, scoring can be done fast enough to support even interactive applications on commodity hardware [7]. The potential energy is calculated as the sum of energy between all atoms i in a

TABLE I: Identification of individual rotated bonds for each non-native conformation. Each value in the table is the angle $\Delta\alpha_i$ by which the original bond i was rotated. Dashes mean that the bond was not rotated in relation to the native ligand state. The RMSD is calculated relative to the native conformation.

Non-Native Ligand Conformations									
Bond ID	2	3	4	5	6	7	8	9	10
C27—C28	110°	110°	-	-	-	110°	-	-	-70°
C21—C38	-235°	-235°	-10°	20°	-	-235°	-	-110°	-
C15—O8	-150°	-150°	-	-	-	-150°	-	-	-
C10—O6	-33°	-33°	-	-	-	-33°	-	-	-
C31—O11	-126°	-126°	-	-	-	-126°	-	-	-
C38—C39	-60°	-60°	-	-	-	-60°	-	-	-
C26—C27	-	-190°	-	-	-	114°	-	-	-
C28—C29	-	-	-	-15°	-	-	60°	-	-
O11—C45	-	-	-	-	20°	-	-	-	-
RMSD (Å)	2.03	2.88	0.17	0.32	0.06	3.11	1.00	0.81	1.37

receptor molecule R and all atoms j in a ligand molecule L, as shown in Equations 1, 2, 3.

$$U_{esp}(i,j) = C\frac{q_i q_j}{r_{ij}} \tag{1}$$

$$U_{vdw}(i,j) = \sqrt{\epsilon_i \epsilon_j}\left[\left(\frac{\rho_i+\rho_j}{r_{ij}}\right)^{12} - 2\left(\frac{\rho_i+\rho_j}{r_{ij}}\right)^{6}\right] \tag{2}$$

$$U = \sum_{i}^{R}\sum_{j}^{L}\left[U_{esp}(i,j) + U_{vdw}(i,j)\right] \tag{3}$$

C is the electrostatic constant, $q_{i(j)}$ is the atomic charge of atom i (or j), r_{ij} is the distance between atoms i and j, $\epsilon_{i(j)}$ is the van der Waals well depth parameter of atom i (or j) and $\rho_{i(j)}$ is the van der Waals radius parameter of atom i (or j). All the amino acid parameters are given by the Amber99 force field [33]. Ligand parameters were obtained from Antechamber [34] calculations.

C. Roadmap Construction

This study constructs state transition roadmaps using PRMs, producing roadmaps that can be efficiently queried for motion paths (Algorithm 1). Ligand states are generated at random in a Gaussian distribution $\mathcal{N}$ and evaluated according to their potential energy, sampling a 6-dimensional space of states x, y, z, p, t, r. Then, edges are formed between states if the RMSD length is within a threshold, representing a state transition. Finally, these transitions are weighted according to the difference in potential energies (Equation 4), allowing a shortest weighted path algorithm to query for a path between any two states in the roadmap.

$$W_{ij}(\Delta E) = \begin{cases} 1/\ln(-\Delta E), & \text{if } \Delta E \leq -2 \text{ kcal/mol} \\ c_1\Delta E + c_2, & \text{if } \Delta E > -2 \text{ kcal/mol} \end{cases} \tag{4}$$

The edge weight function (Equation 4) is used to penalize transitions into higher potential energy states while still differentiating between transitions into lower ones. In this equation, the constants are $c_1 = 0.1858$ and $c_2 = 1.8142$, and $\Delta E = E_j - E_i$ is the energy difference between the final (j) and initial (i) states connected by an edge. This expression for the edge weight function guarantees that all $W_{ij}(\Delta E) > 0$.

```
Construct a roadmap from new samples
Given a set of states in M and edge limit d_limit;
for Each state i in M do
    for Each other state j in M do
        Let d_rmsd be the RMSD between i and j;
        if d_rmsd ≤ d_limit then
            Find potential energies x_i and x_j;
            Calculate weight W_ij using Equation 4;
            Add edge E_ij with weight W_ij to M;
        end
    end
end
Result: M is now a roadmap of edges and states
```

Algorithm 1: Roadmap Construction

Paths obtained from the initial roadmap are optimized according to Algorithm 2 by constructing additional roadmaps of higher density, as illustrated in Figure 1. States that do not belong to the path are excluded. A new roadmap is created that includes the states along the path and a new set of states. These new states are sampled with Gaussian distributions centered at each original state of the path. In this path optimization step, the values for μ and σ^2 in the Gaussian distribution $\mathcal{N}$ should be smaller than those used for the initial roadmap. To ensure more optimal states are considered in the roadmap construction step, states are only kept if their potential energy is lower than the original state energy x_i. Next, a new roadmap is constructed (Algorithm 1) and the original query is repeated to find an lower energy path. Roadmap creation and path optimization steps are iterated as necessary to continue lowering the pathway energy.

To analyze the effect of multiple ligand states in path optimization, we compare four different scenarios for the addition of ligand states in roadmap construction. Figure 3 shows a flowchart representation of the types of roadmaps constructed in our method. In Figure 3, *Top*, the two original roadmaps

```
Initialization step
Let P be input path;
Let N be samples per iteration;
Let F be the available ligand conformation states;
Let c be number of conformations in F;
Let d_limit be the RMSD edge length limit;

for I iterations do
    Sample new lower energy states
    Let m be number of states in P;
    Let M be a new empty roadmap;
    for Each state i in P do
        Let x_i be the potential energy of state P_i;
        Add P_i to M;
        while M < N/m samples do
            Choose S ∈ x, y, z, p, t, r from N(μ, σ^2);
            Choose f ∈ F from U(1, c);
            Calculate potential energy x_s of state S, f;
            if x_s ≤ x_i then
                Add S, f to M;
            end
        end
    end

    Construct roadmap with M and d_limit (Algorithm 1);

    Find the new path and repeat
    Query P_1 (start) to P_m (goal) in M;
    Set P to the new path yielded;
end
Result: P is now the optimized path
```

Algorithm 2: Path Optimization

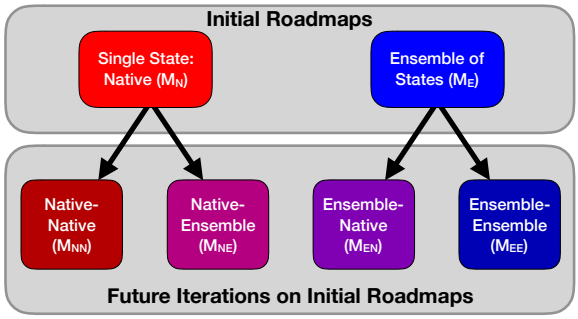


Fig. 3: Roadmaps used in the pathway optimization method. *Top*: Two original roadmaps are created. One roadmap contains only native conformations of the ligand (M_N, left), and another contains any ligand state from the ensemble of 10 states (M_E, right). A low energy path (original path) is found in each of these roadmaps as shown in Figure 1(a). *Bottom*: Future iterations on the initial roadmaps. New samples around the original pathways of M_N and M_E are created as shown in Figure 1(b-d). From the initial roadmaps, 4 scenarios of future iterations exist: Native-native (M_{NN}), starting from roadmap M_N and adding new samples that are only in the native conformations for all iterations; Native-Ensemble (M_{NE}), starting from roadmap M_N and adding new samples that belong to the ensemble of 10 ligand states for all iterations; Ensemble-Native (M_{EN}), starting from roadmap M_E and adding new samples that are only in the native conformations for all iterations; Ensemble-Ensemble (M_{EE}), starting from roadmap M_E and adding new samples that belong to the ensemble of 10 ligand states for all iterations.

used in our study: One generated with only native ligand states, M_N, and another generated with all 10 ligand states in the ensemble, M_E. Once low energy paths are determined from these roadmaps (Figure 1(a), (b)), future roadmap construction in the optimization iterations (Algorithms 1, 2) will: (1) add only native states ($c = 1$, $F = \{1\}$) to the original paths from initial roadmaps, and to future iterations on the optimized paths (M_{NN} and M_{EN} in Figure 3, *Bottom*); or (2) add any ligand state from the ensemble ($c = 10$, $F = \{1, ..., 10\}$) to roadmaps created from the original path, and to future iterations on the optimized paths (M_{NE} and M_{EE} in Figure 3, *Bottom*). The four scenarios we compare are the iterations represented by M_{NN}, M_{NE}, M_{EN}, and M_{EE}.

IV. Results

To examine the effects of selecting distinct ligand conformation states while exploring the state space of rigid body molecular docking, a path optimization scheme was applied to low energy paths obtained from querying two initial roadmaps.

The two orginal roadmaps M_N and M_E were created by constructing edges of length no greater than 5Å RMSD between states among 50,000 Gaussian distributed state samples with a mean around the native state, a translation deviation of 10Å and a rotational deviation of 180° as seen in Figure 4. States from the roadmap containing only states in the native conformation (referred to as the "Single-State (Native)", or M_N Roadmap) are shown in Figure 4, red. States in the other roadmap comprised of the 10 ligand conformations chosen uniformly at random (herein called the "Ensemble of States", or M_E Roadmap) is shown in Figure 4, blue. Most states obtained with RMSD $<$ 10 Å are high energy states due to atomic collisions. Edges were weighted according to Equation 4. Shortest weighted path queries were performed on the roadmaps to obtain two paths to be analyzed for optimization. Figure 5(a) shows the original low energy path in the M_N roadmap (black line), and Figure 5(b) shows the original low energy path in the M_E roadmap (black line).

From the original paths, we resampled new ligand states from a Gaussian distribution using a mean centered around the states of the original path with a translation standard deviation of 2.5Å and a rotational standard deviation of 5°. New states are accepted when they are a potential energy equal to or lesser than the original state's energy, and this was repeated until 500 states were found per iteration. A new roadmap is constructed from these states using an edge limit of 5Å (Algorithm 1) and

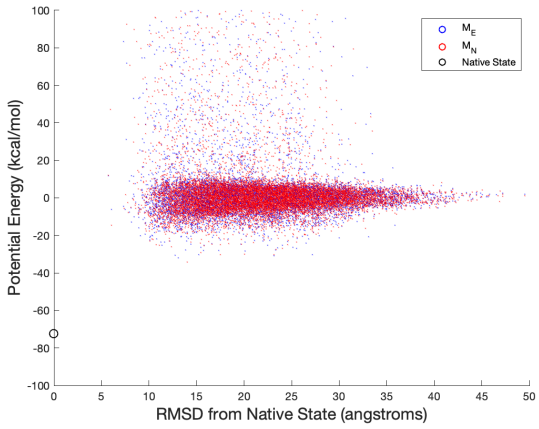
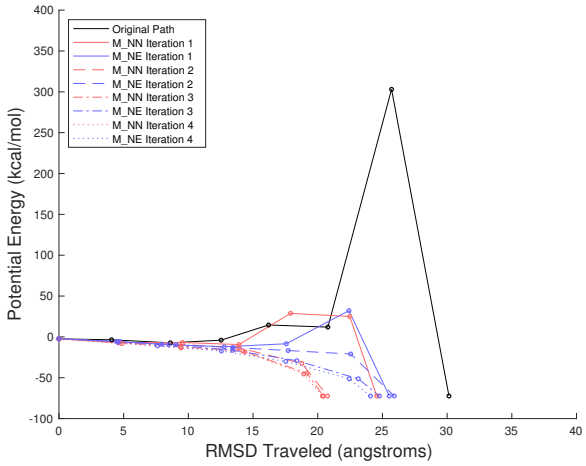
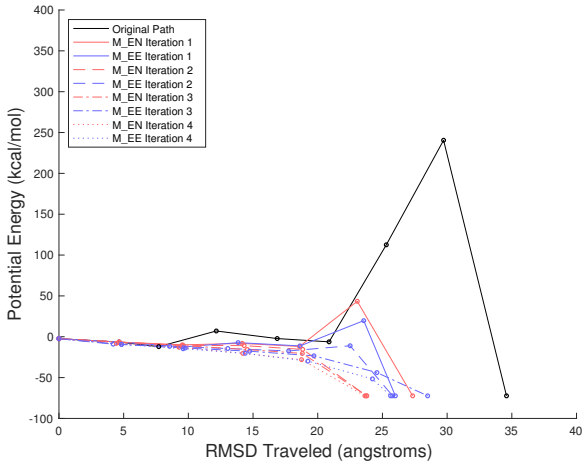


Fig. 4: The initial roadmaps (M_N and M_E in Figure 3, *Top*) of our analysis (only states are shown, not edges). Each roadmap contains 50,000 Gaussian distributed rigid body ligand states. Ligand states are represented by dots. The intermolecular potential energy is plotted against the RMSD from known native state. States were generated with a mean centered around the native state, 10.0Å in translation deviation and 180° in angular deviation. Single state (native, M_N) roadmap (red) can be seen against the roadmap made of an ensemble of conformations (M_E), in blue. Most states obtained in RMSD $\in [0, 10]$ Å are high energy collision states and are not shown in the plot.



(a) Initial Roadmap: Single state (Native), M_N



(b) Initial Roadmap: Ensemble of states, M_E

Fig. 5: Queries performed on a roadmap built with Gaussian distributed ligand states. Edges between states are formed if they are closer than 5.0Å RMSD. The query is performed from a start state 14.0Å RMSD from the native state.

the original query is performed iteratively on this new roadmap to obtain a possibly more optimal path (Algorithm 2).

In this study path optimization was able to find more energetically feasible paths regardless of the roadmap (4^{th} iterations shown as dotted lines in Figure 5). The path obtained for the 4^{th} iteration of M_{NN} and M_{EN} (dotted red lines) is shorter than those obtained in the 4^{th} iteration of M_{NE} and M_{EE} (dotted blue lines), but with slightly higher potential energies. This is because non-native conformations can increase the RMSD distance (up to 3.11Å in this study with conformation state '7'), resulting in fewer edges meeting the 5 Å limit in the iterative roadmaps as shown in Table II, which yield fewer shortcuts in the shortest path algorithm.

Path optimizations on M_{NE} and M_{EE} resulted in a slight reduction of potential energy over optimizations on M_{NN} and M_{EN}, and the number of samples that had to be evaluated for potential energy was reduced. The number of evaluations that had to be performed to obtain 500 acceptable state samples per iteration are shown in Table II with M_{NE}, M_{EE} optimization only requiring about half as many samples by the 4^{th} iteration, as compared to M_{NN}, M_{EN} respectively (values shown in bold). This reduces the computation time by almost 6 hours (in the worst case), eliminating a significant amount of computational cost (on a single core of an Intel Xeon E3-1240 running at 3.7GHz). Lower potential energies in difficult locations were more likely to be found when exploring paths in M_{NE}, M_{EE}.

Even though ligand conformations in M_{NE}, M_{EE} optimizations are chosen in an uniformly random distribution, the histograms in Figure 6 show that the distribution of states with potential energies low enough to be accepted by Algorithm 2 during the sampling step is not uniform, especially towards later iterations. States '2', '3', and '7' were common in iterations 3 and 4 of the path optimization. States '2' and '7' were commonly chosen for later iterations over the M_{NE} roadmaps (Figure 6a), while '3' and '7' were common in the M_{NE} roadmaps (Figure 6b). Due to continual path refinement over subsequent iterations, these later iterations had lower potential energy barriers. Therefore, in these particular roadmaps, non-native conformation states with RMSD to native greater than 2 Å aided exploration of challenging state space.

V. Conclusions

We have shown that providing an ensemble of conformations to support ligand flexibility in rigid body molec-

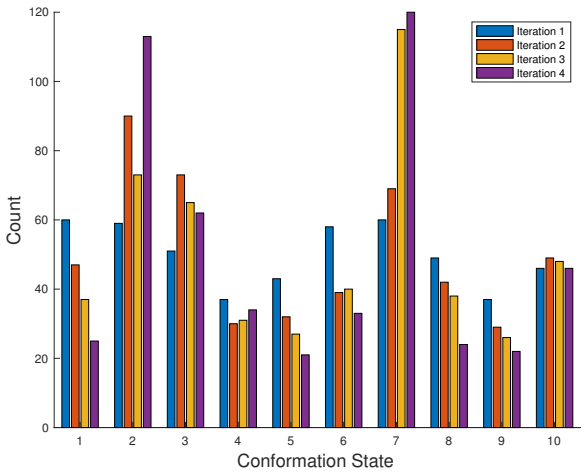
TABLE II: Number of potential energy samples calculated while refining paths taken from both roadmaps. These samples are taken around the states in the original paths in a Gaussian distribution with 2.5Å deviation for translation and 5.0° deviation for rotation. Samples are taken until there are 500 states that satisfy potential energy limits that are equal to or lesser than the original path states, per iteration. Computation time was recorded on a DELL T3620 MT Precision Workstation with an Intel Xeon E3-1240 (3.7GHz).

Roadmap	Iteration	Samples Evaluated	Computation Time (min)	Avg. Edges per State
M_{NN}	1	2372	1.0	110
	2	11840	5.1	175
	3	304526	131.3	192
	4	**1660565**	**716.0**	171
M_{NE}	1	2461	1.1	77
	2	7536	3.2	113
	3	273516	117.9	121
	4	**838231**	**361.4**	107
M_{EN}	1	2593	1.1	107
	2	19260	8.3	126
	3	146895	63.3	179
	4	**649615**	**280.1**	194
M_{EE}	1	2150	0.9	78
	2	8955	3.9	99
	3	95097	41.0	96
	4	**366425**	**158.0**	112

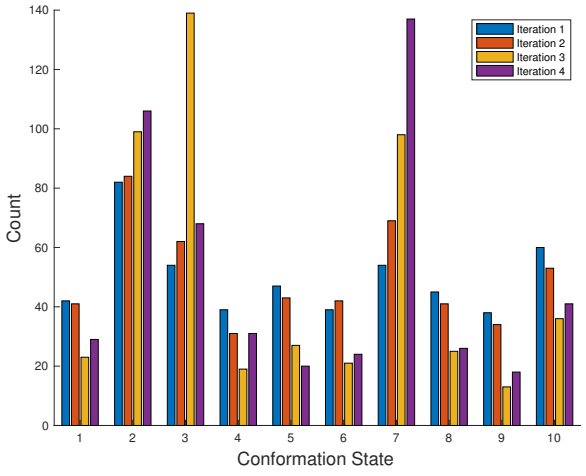
ular docking can explore regions in the high-dimensional energy landscape that may be initially unreachable with a single ligand rigid body conformation. Even though paths were improved regardless of state selection, the ensemble of ligand conformations improves performance of the sampler as it finds new ligand states within decreasing potential energy constraints. This indicates the possibility of improved exploration of the energy landscape by users of interactive molecular docking within the limited computational resources of commonly available mobile devices.

Figure 7 shows four out of seven states that were chosen in the 4^{th} iteration of the path optimization of M_{EE}. Even though the algorithm found low energy paths, these may not correspond to the correct biological pathways of binding. The ligand states selected for this work were generated randomly, as our goal was not to determine biological pathways for drug discovery, but to demonstrate feasibility of the method. To guarantee that the paths obtained by the algorithm are biologically feasible, ligand states should be generated by a physics-based method such as molecular dynamics, with individual states identified with Markov state models such as in [12]. Conversely, large RMSD differences in the final ligand pathway can be smoothed using methods such as targeted molecular dynamics.

The performance of exploration with an ensemble of ligand conformation states is limited by the ability to sample the conformation space of molecules, and requires a more robust method to be applicable to general interactive molecular docking environments. This study only considered ligand flexibility,



(a) M_{NE}



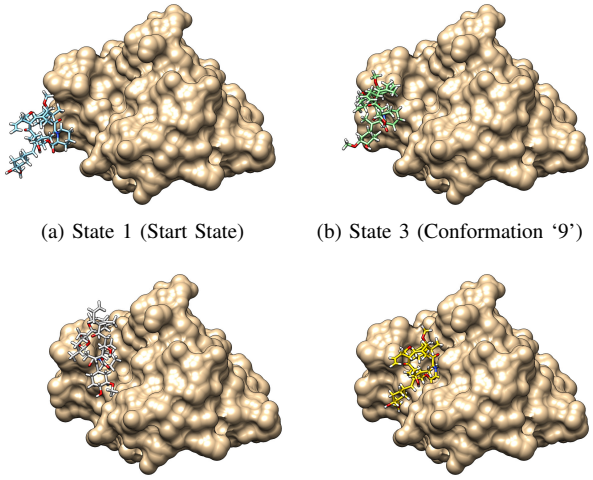
(b) M_{EE}

Fig. 6: Composition of ligand conformation states accepted for each iteration of (a) M_{NE} and (b) M_{EE} path optimizations (500 each iteration). State '1' is the native conformation state and '2'–'10' are the non-native ligand conformation states. Some iterations showed a heavy preference towards particular conformation states, particularly states '2', '3', and '7'.

but receptor flexibility also has a large impact on docking pathways. The number of flexible states a human operator can effectively utilize is also unknown. This work will be applied to future user studies of interactive docking to understand these limits and test interfaces for conformation selection.

ACKNOWLEDGMENTS

The authors thank Selina Bauernfeind for helpful comments. This material is based upon work supported by the National Science Foundation under Grant Numbers IIS-1716195, IIS-1553266, and IIS-1528047. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



(a) State 1 (Start State) (b) State 3 (Conformation '9')

(c) State 5 (Conformation '2') (d) State 7 (Native Conformation)

Fig. 7: States along the path obtained in the 4^{th} iteration of M_{EE} (7 in total). All queries started with the same state (a) and ended with the known bound state (d). Intermediate states along the path are shown in between along with the conformation state.

REFERENCES

[1] R. Capelli, P. Carloni, and M. Parrinello, "Exhaustive search of ligand binding pathways via volume-based metadynamics," *The Journal of Physical Chemistry Letters*, vol. 10, no. 12, pp. 3495–3499.

[2] F. Zeller, M. P. Luitz, R. Bomblies, and M. Zacharias, "Multiscale simulation of receptor–drug association kinetics: Application to neuraminidase inhibitors," *Journal of Chemical Theory and Computation*, vol. 13, no. 10, pp. 5097–5105, 2017.

[3] S. Grinter and X. Zou, "Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design," *Molecules*, vol. 19, no. 7, pp. 10 150–10 176, 2014.

[4] G. Iakovou, S. Hayward, and S. D. Laycock, "Virtual environment for studying the docking interactions of rigid biomolecules with haptics," *Journal of Chemical Information and Modeling*, vol. 57, no. 5, pp. 1142–1152, 2017.

[5] N. Matthews, A. Kitao, S. Laycock, and S. Hayward, "Haptic-assisted interactive molecular docking incorporating receptor flexibility," *J. Chem. Inf. Model.*, vol. 59, no. 6, pp. 2900–2912, 2019.

[6] X. Hou and O. Sourina, "Six degree-of-freedom haptic rendering for biomolecular docking," in *Trans. Comput. Sci. XII*, ser. Lect. Notes Comput. Sci., vol. 6670. Springer Berlin Heidelberg, 2011, pp. 98–117.

[7] T. Adamson, J. Baxter, K. Manavi, A. Suknot, B. Jacobson, P. G. Kelley, and L. Tapia, "Molecular Tetris: Crowdsourcing molecular docking using path-planning and haptic devices," in *Proc. Intl. Conf. Motion Games (MIG)*. Los Angeles, California, USA: ACM, 2014, pp. 133–138.

[8] A. Chávez, T. Adamson, L. Tapia, and B. Jacobson, "A mobile game for crowdsourced molecular docking pathways," in *Proc. Intl. Conf. Motion Games (MIG)*. Newcastle upon Tyne, United Kingdom: ACM, 2019.

[9] G. Levieux, G. Tiger, S. Mader, J.-F. Zagury, S. Natkin, and M. Montes, "Udock, the interactive docking entertainment system," *Faraday Discussions*, vol. 169, pp. 425–441, 2014.

[10] K. Teilum, J. G. Olsen, and B. B. Kragelund, "Protein stability, flexibility and function," *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, vol. 1814, no. 8, pp. 969–976, 2011.

[11] B. Leimkuhler and C. Matthews, *Molecular Dynamics*. Springer, 2016.

[12] N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models," *Nature Communications*, vol. 6, p. 7653, 2015.

[13] R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni, and M. Parrinello, "Unbinding kinetics of a p38 map kinase type ii inhibitor from metadynamics simulations," *Journal of the American Chemical Society*, vol. 139, no. 13, pp. 4780–4788, 2017.

[14] B. Xie, J. D. Clark, and D. D. L. Minh, "Efficiency of stratification for ensemble docking using reduced ensembles," *J. Comput. Infect. Model.*, vol. 58, no. 9, pp. 1915–1925, 2018.

[15] R. E. Amaro, J. Baudry, J. Chodera, Ö. Demir, J. A. McCammon, Y. Miao, and J. C. Smith, "Ensemble docking in drug discovery," *Biophysical Journal*, vol. 114, no. 10, pp. 2271–2278, 2018.

[16] O. B. Bayazit, G. Song, and N. M. Amato, "Ligand binding with OBPRM and user input," in *Proc. IEEE Intl. Conf. Rob. Auto. (ICRA)*, vol. 1, Seoul, Korea, 2001, pp. 954–959.

[17] A. Bolopion, B. Cagneau, S. Redon, and S. Régnier, "Haptic feedback for molecular simulation," in *Proc. IEEE Intl. Conf. Intel. Rob. Syst. (IROS)*, St. Louis, MO, USA, 2009, pp. 237–242.

[18] S. Cakici, S. Sumengen, U. Sezerman, and S. Balciso, "DockPro: A VR-based tool for protein-protein docking problem," *Intl. J. Virtual Reality (IJVR)*, vol. 8, no. 2, pp. 19–23, 2009.

[19] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic, and F. Players, "Predicting protein structures with a multiplayer online game," *Nature*, vol. 466, pp. 756–760, 2010.

[20] N. M. Amato, O. B. Bayazit, L. K. Dale, C. Jones, and D. Vallejo, "OBPRM: An obstacle-based PRM for 3d workspaces," in *Proc. Intl. Workshop Alg. Found. Rob. (WAFR)*, 1998, pp. 155–168.

[21] N. M. Amato and G. Song, "Using motion planning to study protein folding pathways," *J. Comput. Biol.*, vol. 9, no. 2, pp. 149–168, 2002.

[22] O. B. Bayazit, G. Song, and N. M. Amato, "Enhancing randomized motion planners: Exploring with haptic hints," *Autonomous Robots*, vol. 10, no. 2, pp. 163–174, 2001.

[23] R. Geraerts and M. H. Overmars, "Clearance based path optimization for motion planning," *Proc. IEEE Intl. Conf. Rob. Auto. (ICRA)*, vol. 3, pp. 2386–2392, 2004.

[24] A. Ravankar, A. A. Ravankar, Y. Kobayashi, Y. Hoshino, and C.-C. Peng, "Path smoothing techniques in robot navigation: State-of-the-art, current and future challenges," *Sensors (Basel)*, vol. 18, no. 9, p. 3170, 2018.

[25] B. Bogaerts, S. Sels, S. Vanlanduit, and R. Penne, "A gradient-based inspection path optimization approach," *IEEE Rob. Auto. Letters*, vol. 3, no. 3, pp. 2646–2653, 2018.

[26] K. E. Bekris, B. Y. Chen, A. M. Ladd, E. Plaku, and L. E. Kavraki, "Multiple query probabilistic roadmap planning using single query planning primitives," *Proc. IEEE Intl. Conf. Intel. Rob. Syst. (IROS)*, vol. 1, pp. 656–661, 2003.

[27] M. L. Teodoro, G. N. P. Jr., and L. E. Kavraki, "Molecular docking: a problem with thousands of degrees of freedom," *Proc. IEEE Intl. Conf. Rob. Auto. (ICRA)*, p. 960, 2001.

[28] J. Cortes, L. Jaillet, and T. Simeon, "Molecular disassembly with RRT-like algorithms," *Proc. IEEE Intl. Conf. Rob. Auto. (ICRA)*, pp. 3301–3306, 2007.

[29] M. K. Nguyen, L. Jaillet, and S. Redon, "ART-RRT: As-rigid-as-possible exploration of ligand unbinding pathways," *J. Comput. Chem.*, vol. 39, no. 11, pp. 665–678, 2019.

[30] D. A. Antunes, D. Devaurs, M. Moll, G. Lizée, and L. E. Kavraki, "General prediction of peptide-MHC binding modes using incremental docking: A proof of concept," *Sci. Reports*, vol. 8, no. 1, 2018.

[31] G. D. Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, and J. Clardy, "Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex," *Science*, vol. 252, no. 5007, pp. 839–842, 1991.

[32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera - a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[33] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cielplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *J. Comput. Biol.*, vol. 24, no. 16, pp. 1999–2013, 2003.

[34] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *J. Mol. Graph. Model.*, vol. 25, pp. 247–60, 2006.