Using Motion Planning to Evaluate Protein Binding Site Accessibility

Diane Uwacu, Everett S. Yang, Shawna Thomas, and Nancy M. Amato

Technical Report TR18-001
Texas A&M University, College Station TX 77840, USA,
http://www.parasol.tamu.edu

Abstract. Despite much effort and considerable breakthroughs in ligand binding prediction, the best predictors still produce many false positives and a reliable, fully automated prediction framework has yet to be developed. Binding site accessibility is an important feature ignored by methods that classify binding based solely on the energetic or geometric properties of the final bound protein-ligand complex. To evaluate this necessity, we transform the ligand accessibility problem into a robot motion planning problem where the ligand is modeled as a flexible agent whose task is to travel from outside the protein to its binding site. We use Rapidly-exploring Random Graphs coupled with Mean Curve workspace skeletons to quickly and thoroughly explore a protein environment in order to produce valid paths for ligand motion. Path weights reflect the influences of intermolecular forces on the given ligand. Low weight paths are extracted and analyzed for characteristics of accessibility. In this paper, we show that our algorithm provides a mechanism to evaluate binding site accessibility for a ligand.

Keywords: ligand binding, motion planning, computational biology, accessibility

1 Introduction

Ligand binding is the process in which a ligand (e.g., drug molecule) binds to a specific site on a protein, allowing the ligand to create a stable compound with the protein. Understanding protein-ligand interactions is important in the analysis of drug molecules and can provide invaluable insight for many biological phenomena.

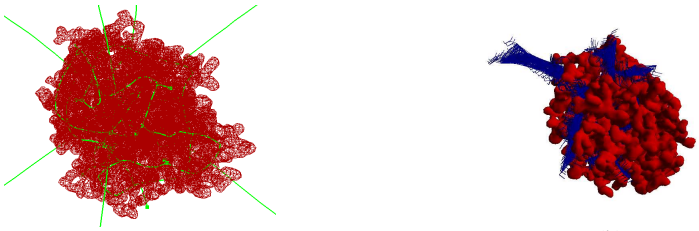
Several binding site prediction methods have been created based on geometric and energetic considerations [19]. However, they typically only consider the state of the final bound complex. In this way, such methods would not consider the dynamics of the ligand as it approaches the protein, ignoring the effects of intermolecular (e.g., van der Waals) forces on a ligand body moving along its trajectory. These methods also often make the improper assumption of rigidity of protein and ligand structures.

Recently, attention has been turned towards protein tunnels that connect buried binding sites to the convex hull surface of the protein. It has been found that these tunnels influence binding site activity and stability by regulating accessibility of certain ligands [7]. This has motivated us to use motion planning to study the accessibility of such sites by the ligand.

In this work, we present a motion planning based approach to model the accessibility of a binding site to a ligand. The idea of using motion planning to model ligand binding was first presented in [22]. Since then, there have been ongoing attempts to apply motion planning to ligand binding, including [1] which combines Obstacle-based PRMs with input from the user. Although this approach has been shown to be effective and has benefits of its own, the ultimate goal was to provide a fully automated framework for the study of protein-ligand interactions.

Our work uses Rapidly-Exploring Random Graphs (RRG) [6] to explore protein free space and find paths that represent the accessibility of the binding site by the ligand. We chose to use RRGs because they allow for multiple queries, efficient navigation of narrow passages, and inherit the benefits of an RRT. To generate valid ligand configurations in the protein, we restrict our sampling to dynamic regions along the workspace skeleton and use an Obstacle-based sampler to produce samples on the protein surface in these regions.

Contribution. In this paper, we provide a new motion planning approach to ligand binding that assesses the accessibility of a ligand to a binding site. To analyze our method, we perform a case study involving two ligands that bind to distinct mutants of Haloalkane dehalogenase (DhaA). We show that considering accessibility provides a significant improvement in the prediction of reactivity when compared to existing methods.



(a) (b)

Fig. 1: (a) shows a Mean Curve workspace skeleton of the free space of the protein 4WCV (wireframe). Planning along the skeleton gives the well explored roadmap (b) in blue.

2 Related Work

In this section, we review existing methods for predicting binding sites and explain concepts and terms that we will use in the rest of the paper.

2.1 Predicting protein-ligand binding sites

Geometry-based methods find cavities on the protein's surface and return those that are more likely to physically fit a ligand. In general, these methods return the largest pocket

among the identified ones [5, 19, 30, 25]. Some methods in this category also check for energy stability once they have a small set of cavity candidates. By considering more information, the accuracy is improved.

Energy-based methods use molecular probes to sample the surface of a protein in order to find areas that offer the most favorable energy for the protein-ligand reaction to take place [21].

Since low energy levels indicate stability of the ligand-protein bond, residues that are likely to minimize energy are included in the binding site. Methods in this category also consider the geometric factors (that is, if the ligand can fit in the cavity) by using grid based algorithms to identify cavities.

Template-based methods first find a protein with known binding sites that is structurally similar to the protein of interest. Then, binding sites are found for the new protein by comparing to the identified template through sequence alignment [16]. These methods tend to be highly accurate, but ther accuracy depends on the choice of the template protein. In addition, structurally similar proteins can have different binding preferences and proteins that are not identical in structure can have the same binding preferences.

Statistically-based methods identify features that are known to differentiate binding and non-binding sites. These features are then studied using statistical methods and used to score protein regions. Regions with high scores are returned as binding sites. The most used features are cavity size, energy level and electrostatic forces [3, 29]. Recently, such methods have gained popularity, sparking new research in the related areas of machine learning.

There are some studies regarding accessibility. CAVER [17] is a tool that helps analyze the dynamics of protein tunnels. It is used in many methods to study the binding process. For example, Kaushik et al. [7] recently showed the importance of focusing on ligand-specific accessibility for tunnel engineering using CAVER to predict protein tunnel, and molecular dynamics to explore different configurations of the protein as the tunnels open and close.

2.2 Motion Planning Preliminaries

Motion planning consists of finding a valid path between a start and a goal for a movable object. Although this problem seems simple, it becomes increasingly difficult as the robot becomes more constrained and the environment more complex [12]. Except for robots with few degrees of freedom, the problem is computationally intractable [13].

Sampling-based motion planning algorithms have been particularly successful in solving these problems in a relatively short time. The trade-off for this faster computation is probabilistically completeness. That is, as more time is spent exploring the problem space, the probability of not finding a solution, if one exists, approaches zero [12]. As we will discuss next, this property of sampling-based algorithms leads us to assign ratings to binding sites in order to represent uncertainty.

Rapidly-exploring Random Trees (RRTs) have been shown to be useful in narrow passage environments [14]. They construct a roadmap by repeatedly extending from an existing one. Specifically, they sample a random node q_{rand} in the state space and attempts to connect it to the closest configuration q_{near} on an existing roadmap by taking intermediate configurations on the straight line edge joining the two. The furthest

valid configuration from q_{near} becomes a new node in the tree. However, RRTs lack multiquery capabilities.

Another fundamental method, the Probabilistic Roadmap (PRMs) [8], constructs roadmaps by generating random free configurations of the robot and connecting these configurations using a local planner [8]. Once a roadmap is constructed, a path connecting the start to the goal can be found using any searching algorithm (A*, Dijkstra's algorithm, etc.). PRMs are capable of conducting multiple queries. However, they are not suited for planning in narrow passages.

A recently developed method, the Rapidly-Exploring Random Graph [6], unifies PRMs and RRTs, inheriting the benefits of both methods.

There has also been previous work using motion planning to study ligand binding. Specifically, gradient descent methods have been used [22] where a set number of nodes are uniformly sampled in the protein environment and dense sample clusters are created around those configurations that exhibit the lowest energy. Eventually they construct a roadmap representation of the protein that approximates the protein environment. In most cases, they were able to detect the true binding site. However, they failed to detect the true binding site for one example due to the problem of narrow passages.

More recently, there has been work on applying Obstacle-based PRMs to Ligand Binding [1]. OBPRM is used to sample on the surface of the protein in order to detect the true binding site. Their work also incorporates haptic user input to allow users to generate nodes in the local minima of a protein. This approach leverages human intelligence to determine the optimal minima based on haptic feedback.

2.3 Dynamic Region-biasing and Skeleton-guided Motion Planning

Sampling-based motion planning algorithms are challenged by environments with narrow passages and clutter. Workspace skeletons intend to alleviate this issue by constructing an abstract representation of the workspace. Here we describe a few types of such skeletons and briefly introduce a relevant sampling scheme.

Recently, Denny et al. [4] emphasized the importance of using the environment topology to guide planning in narrow and cluttered environments. Their Dynamic Region-biased RRTs allows for the efficient exploration of workspace by constructing a Flow Graph skeleton that compactly encodes the changes in the topology of the workspace. Representative configurations are generated by confining sampling to dynamic sampling regions (any bounded volume) along this skeleton. Once such a region contains a sample, it is advanced along the Flow Graph until the sample is no longer contained, thus preventing redundant sampling.

Mean Curve Skeletons [24] are constructed using a mesh-based algorithm that incorporates mean curvature flow in order to compute a skeletal representation of some surface. Specifically, the input mesh of an object is compressed and a skeleton is extracted. During skeleton construction, the geometry is locally remeshed via edge collapses, while preserving shape topology, allowing for fast computation and a precise skeleton.

3 Overall Approach

Our approach is inspired by the Dynamic Region-biasing RRT method mentioned in Section 2.3. The idea is that this algorithm, augmented with the Mean Curve skeleton, provides a framework that can accurately model protein-ligand interactions, including ligand accessibility. At a high level, we are building a geometric model of a protein and using a topological representation of the model for guidance when planning ligand motion. Once we finish the roadmap construction, we extract favorable paths from the graph. Detailed steps of the approach are described in the following subsections.

For simplicity, we will use a 3D tunnel maze (Figure 3) environment to explain the algorithm. This environment is composed of eight C shaped tunnels inside a box. Four of the tunnels break through the surface of the box, while the other four are blocked. Tunnels also have different radii.

Algorithm 1 Topology-Guided Binding Site RRG

Input. $roots$: configurations that represent the bound set, env: environment with protein and a bounding box, P: Protein pdb file
Output. Roadmap G

$S \leftarrow$ `GetWorkspaceSkeleton`(env) : $Pre - computation$
for each $root \in roots$ **do**
  $G \leftarrow root$
end for
while !$done$ **do**
  $region \leftarrow SelectRegion(S.regions)$
  $samples \leftarrow$ `Sample`$(region, RRTSampler)$
  for each $sample \in samples$ **do**
    if `Extend`$(sample, G)$ **then**
      $G \leftarrow sample$
      $k \leftarrow$ `ConnectClosest`$(sample, G)$
    end if
  end for
  if `IsOutsideConvexHull`$(samples)$ **then**
    $region.Probability \leftarrow 0.00$
  end if
end while

3.1 Modeling Intermolecular Forces

In order to understand how ligands behave around the protein, we must take into account the interacting forces between the two.

In this section, we present an appropriate distance metric that models van der Waals and electrostatic forces.

We discretize our problem space into grid cells of a certain resolution. For every cell coordinate i and every protein atom j, the influence of intermolecular forces on the given ligand at i is measured by:

$$V_{tot} = \sum_{i,j} \frac{332 * q_i q_j}{d_{i,j}} + \frac{A}{{d_{i,j}}^{12}} - \frac{B}{{d_{i,j}}^{6}}$$

where $d_{i,j}$ is the distance between the ligand atom and the protein atom, q_i and q_j are atom charges and A and B are van der waals constants. This is a standard representation found in [15].

Precomputing an energy grid in this manner saves a substantial amount of computation time during planning. A ligand's energy value at any position in the problem space is simply given by the grid cell closest to it.

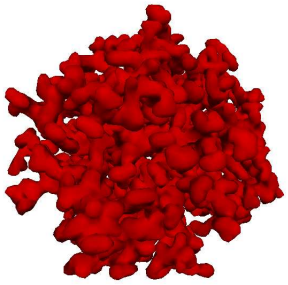
3.2 Modeling Molecular Structures

The environment is composed of a protein body and a mobile ligand structure. Since proteins are largely static, their flexibility is not accounted for and thus we model them as rigid bodies. A ligand is represented as a flexible linkage where bonds between atoms are linkages connected by revolute joints that keep them at a fixed distance from each other.

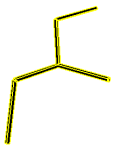
There are a few rules that constrain the flexibility of a ligand:

- Ring structures have strong bonds and thus should remain fixed.
- Adjacent atoms are kept at a fixed distance and angle from each other to preserve the chemical integrity of the molecule.
- Since links are revolute, connections to leaf atoms do not need to be flexible since their revolution does not affect the state of the ligand.

Figure 2 shows an example of a protein and a ligand in our environment.



(a) 4FWB



(b) TCP

Fig. 2: Geometric model of a protein (a) and a ligand (b).

3.3 Protein Skeletalization

Proteins exhibit a highly complex structure that impedes the efficacy of many motion planning techniques. Sampling biased towards a workspace skeleton of the free space

mesh alleviates this problem by constraining samples to be placed in representative regions of the protein.

To construct the mesh, we decompose the free space into tetrahedra that we then pass as input to a skeleton generator. In our study, we use the Mean Curve skeleton described in Section 2.3 as we found it to provide good coverage of the protein free space mesh.

Figure 3a shows the mean curve skeletal structure of the 3D maze tunnels. Note that the skeleton goes through all tunnels because a point robot is used to create its nodes and edges in the workspace. As we will see in the next section, not all tunnels will be sampled in C-space.

3.4 Motion Planning and Path Extraction

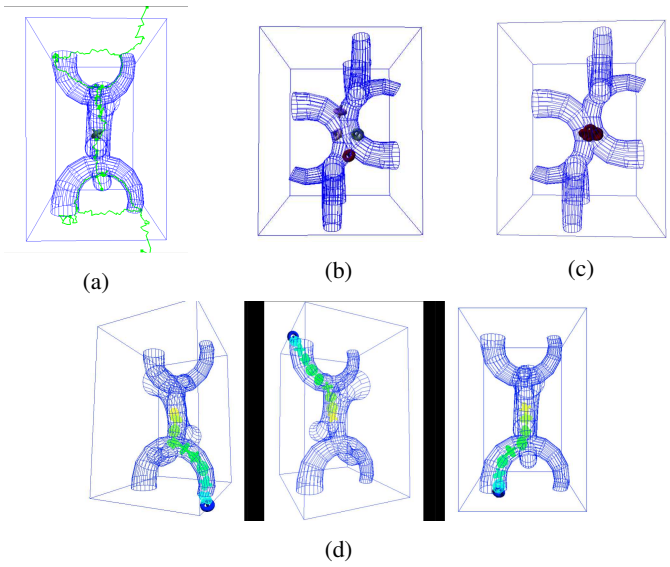


Fig. 3: A mean Curvature Skeleton of the free space structure is created (a). Witness points are extracted around the center of the body (b). Root configurations are created close to witness points (c), and three tunnels are discovered following the skeleton (d).

Our planning algorithm leverages information about known binding site residues.

Binding site residues are protein amino acids, located in the protein (obstacle), that have influence on binding between a protein and a ligand. We obtain a collection of these residues from the Protein Data Bank [2], and use these as witness points for generating valid ligand configurations that will serve as roots of the RRG. Figures 3b and 3c show how witness points on the surface of a body are used to generate free samples in the center area.

We use Algorithm 1 to build a Mean Curve skeleton-guided RRG rooted in the center of the binding site. During planning, the graph exits the protein. Every valid sample generated in the selected region is added to the roadmap. A region is selected based on its relative success rate as described in [4].

In order to ensure that all available tunnels are explored fairly, once a sample crosses the convex hull, sampling in its corresponding region is prohibited, allowing sampling in unexplored tunnels (Line 14 of Algorithm 1). Growth, and therefore planning, terminates once a certain number of nodes in the graph reach the convex hull.

From the constructed roadmap, we identify tunnels from the biology site to the protein surface. A tunnel is defined to be a passage that a ligand traverses to reach the active site. A tunnel is detected whenever there is a connected component of ligand configurations that includes configurations in the active site and on the surface. For each tunnel, a start and goal pair is created from configurations in the unbound (outside the protein) and bound (inside the binding site) sets. The k energetically favorable paths, or the paths with the lowest energy (weight) values, from each tunnel are extracted. Dijkstra's single-source shortest path algorithm is used to determine this.

Generally, motion planning algorithms find the most efficient (fastest or shortest) path connecting a start and a goal. However, we are interested in finding all the possible ways that a ligand can access the binding site region. It is then important to maximally explore all available tunnels. When a tunnel is found (i.e., a branch of the roadmap breaks though the convex hull surface of the protein), we lower the probability of selecting the region for expansion. In this way, we force sampling into previously unexplored regions.

In addition, we stop expansion attempts in a regions after its failure rate has gone above a certain threshold. This allows us to spend more time constructing a roadmap that expands through feasible tunnels and one that does not get stuck in areas that are too small.

In Figure 3d, our algorithm was able to go through three out of the four tunnels proposed by the skeleton. The fourth tunnel was left unexplored due to its radius being smaller than the robot's.

4 Experiments

In this section, we compare our results to the expected results from molecular biology labs as well as results predicted by ProBis-CHARMMing [9].

4.1 Environment Description and Expectations

Rhodococcus rhodochrous is a bacterium used for soil inoculation. It has diverse biodegradation mechanisms. Haloalkane dehalogenase (DhaA) is an enzyme that catalyzes the removal of halogens from a substrate. Many people are interested in engineering mutants of this enzyme to increase binding and thus the degradation of harmful molecules including 1,2,3-Trichloropropane (TCP) and 1,2-Dibromoethane (EDB) that often have no known natural degradation processes.

1,2,3-Trichloropropane(TCP) is an industrial pollutant with no known natural biodegradation pathway. **1,2-Dibromoethane (EDB)** is a synthetic chemical that was used as a fumigant until it was banned in the 1980s [28]. They pose a risk to agriculture and water supplies. DhaA often removes chlorine from compounds that are structurally similar to TCP. It can degrade TCP in the laboratory into dichlropropane (DCL). DhaA can also degrade EDB. However, DhaA's activity is too low to sustain, and many DhaA mutants have been engineered in laboratories in an attempt to increase their affinity to TCP and EDB.

All DhaA mutants have one site and several tunnels leading to it. Some residues in those tunnels play a role in the accessibility of the site. The engineering of DhaA's mutants mainly targets residues in tunnels that influence the accessibility of the binding site.

Our experiments compare DhaA's wild type (found in nature) **4HZG**, against four modified mutants:

- **4FWB**[11]: Has high affinity with TCP. It achieves 32 times more activity towards TCP than the wild type protein.
- **3FBW**[23]: Has moderate affinity with TCP. It achieves 3 times more activity with TCP than the wild type protein.
- **4WCV**[27]: Has high affinity with EDB. It achieves 32 times more activity with EDB than the wild type protein.
- **4F60**[26]: Has low affinity with EDB. Its activity is two times less than that of the wild type protein. However, it achieves more stability, i.e. it is not easily modified by the solvent around it.

4.2 Experimental Setup

We obtain information about molecular structures used in our experiments from PDB files taken from the Protein Data Bank (PDB) [2]. UCSF Chimera [18] is subsequently used to obtain an appropriate geometric representation of the protein from its corresponding PDB file. Similarly, we use a custom script to construct a flexible ligand by parsing its corresponding PDB file.

In every discovered tunnel, we extract five paths with the lowest overall energy profile. We rank protein-ligand activity as $High$, $Moderate$ and Low depending on the quality of the energy profile of paths reaching the binding site. A $High$ rank is assigned to a protein-ligand complex if the energy along the ligand path decreases monotonically as the ligand approaches the binding site cavity. A $Moderate$ rank is assigned to a complex if there are energy barriers along the paths found. An energy barrier is any stark increase in slope along the energy profile, as we shall observe in Figure 5. A Low rank is assigned to a complex if the energy along a path consistently increases as the ligand approaches the binding site residue.

Expected activity is recorded in molecular biology publications referenced in Section 4.1.

Probis-CHARMMing is one of the state-of-the-art binding site prediction methods. **ProBis**, a template-based binding site prediction method [10], is combined with **CHARMMing** [20] to optimize protein-ligand interaction energy computation. Given

a protein, ProBis-CHARMMing returns a list of ligands that are expected to bind with it, with a scoring value that reflects the confindence in the prediction. The maximum score is 5 while the minimum reaches 0.5. In our experiments, all protein-ligand complexes were predicted by Probis-CHARMMing with a confidence above 4.0, which is considered $High$.

4.3 Results

As shown in Table 1, our motion planning rankings are consistent with the observations made in wet lab experiments. As mentioned in 4.2, we base our rankings on the resulting energy profile produced from our algorithm. The wild type protein 4HZG was ranked Low because, in all its discovered tunnels, energy increases as it approaches the binding site, as shown in Figures 4a and 4b. This means that, although the ligand can geometrically fit through the tunnel, the energy conditions do not help the ligand smoothly move toward the binding site.

3FBW was ranked $Moderate$ because of the energy barrier that is midway between the ligand start configuration and the binding site. Although conditions get better around the binding site, the spike in energy that we see in Figures 4c and 4d hinders the activity of the protein with TCP.

4FWB was ranked $High$ because in all discovered tunnels, the energy profile decreases monotonically as the path leads to the binding site, as shown in Figures 4e and 4f.

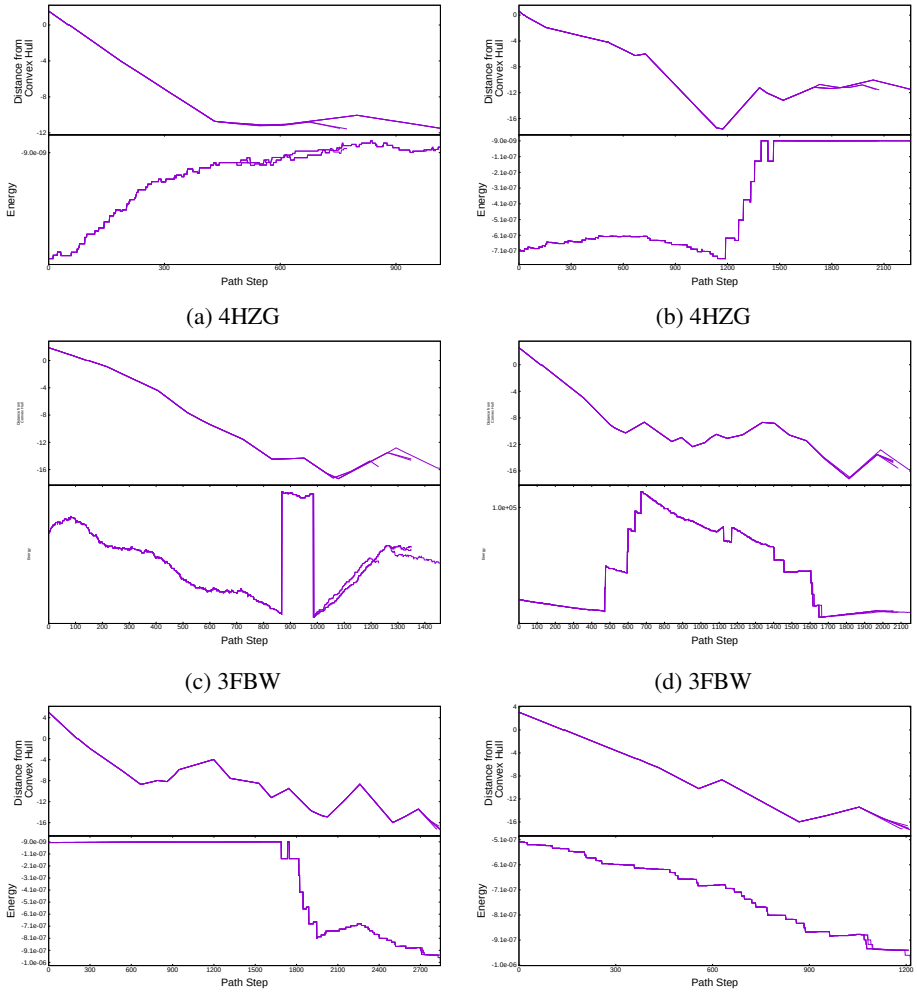
	Expected Activity	**Probis Predicted Activity**	**Motion Planning Accessibility Rating**
4HZG	Low	High	Low
3FBW	Moderate	High	Moderate
4FWB	High	High	High

Table 1: TCP results

Table 2 shows results for EDB. The wild type protein 4HZG was ranked Low because, in all its discovered tunnels, the energy profile increases as it approaches the binding site, as shown in Figures 5a and 5b. 4WCV was ranked $Moderate$ because of the energy barriers that are midway towards the binding site. Figures 5c and 5d show that the spike in energy hinders the activity of the protein with EDB. 4F60 was ranked Low because, in all discovered tunnels, the energy profile consistently increases as it approaches the binding site, as shown in Figures 5e and 5f.

5 Conclusion

In this paper, we have demonstrated the role of motion planning in assessing the accessibility of a binding site on a protein to a ligand. Specifically, we have developed a fully-automated framework for studying the dynamics of ligand binding by considering the accessibility of the ligand to known binding residues on the protein. Our algorithm



(a) 4HZG (b) 4HZG

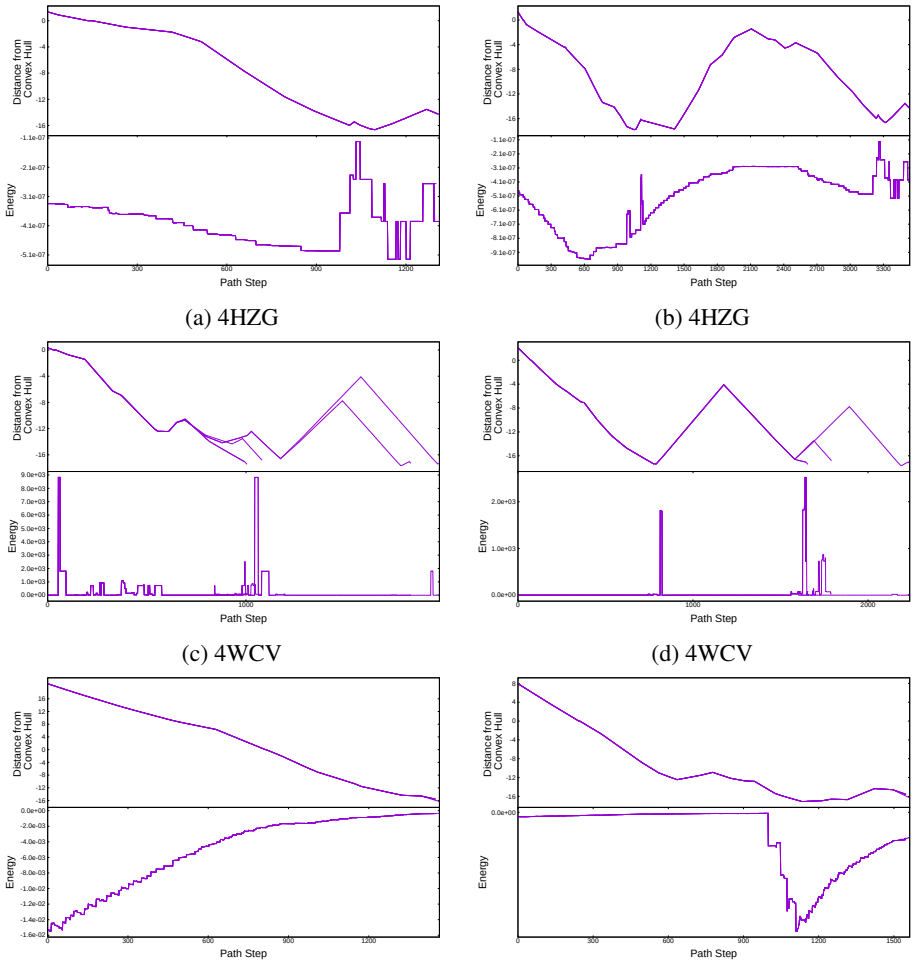
(c) 3FBW (d) 3FBW

(e) 4FWB (f) 4FWB

Fig. 4: TCP energy profiles for 3FBW, 4FWB, and 4HZG, showing the ligands energy value and its distance from the protein convex hull as it travels towards the binding site. The path steps are used to indicate a position in the path. $Note$: Negative distance values indicate that the ligand is inside the convex hull of the protein.

	Expected Activity	**Probis Predicted Activity**	**Motion Planning Accessibility Rating**
4HZG	Low	High	Low
4WCV	High	High	Moderate
4F60	Low	High	Low

Table 2: EDB results



(a) 4HZG (b) 4HZG

(c) 4WCV (d) 4WCV

(e) 4F60 (f) 4F60

Fig. 5: EDB energy profiles for 4HZG, 4WCV, and 4F60, showing the ligands energy value and its distance from the protein convex hull as it travels towards the binding site.

allows us to rapidly explore all pathways connecting these residues to a ligand placed outside of the protein by biasing sampling towards a skeleton representation of the protein free space mesh. In addition, we sample in unique dynamic regions so to disallow redundancies in sampling. Results gathered from five binding scenarios show that there is a high correlation between accessibility and protein-ligand affinity as measured by our method.

References

1. Bayazit, O.B., Song, G., Amato, N.M.: Ligand binding with OBPRM and haptic user input: Enhancing automatic motion planning with virtual touch. In: Proc. IEEE Int. Conf. Robot. Autom. (ICRA). pp. 954–959 (2001), this work was also presented as a poster at *RECOMB 2001*.
2. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P.: The protein data bank. Nucleic Acids Research 28(1), 235–242 (2000)
3. Chen, B.Y.: Vasp-e: Specificity annotation with a volumetric analysis of electrostatic isopotentials. PLOS Computational Biology 10(8) (2014), https://doi.org/10.1371/journal.pcbi.1003792
4. Denny, J., Sandstrom, R., Bregger, A., Amato, N.M.: Dynamic region-biased exploring random trees. In: Proc. Int. Workshop on Algorithmic Foundations of Robotics (WAFR). San Francisco, CA (December 2016)
5. Huang, B., Schroeder, M.: Ligsite cs: credicting ligand binding sites using the connolly surface and degree of conservation. BMC Structural Biology 6, 1472–6807 (2006), http://dx.doi.org/10.1186/1472-6807-6-19
6. Kala, R.: Rapidly exploring random graphs: motion planning of multiple mobile robots. Advanced Robotics 27(14), 1113–1122 (2013), https://doi.org/10.1080/01691864.2013.805472
7. Kaushik, S., Marques, S.M., Khirsariya, P., Paruch, K., Libichova, L., Brezovsky, J., Prokop, Z., Chaloupkova, R., Damborsky, J.: Impact of the access tunnel engineering on catalysis is strictly ligand specific. Federation of European Biochemical Societies Journal 285(8), 1456–1476 (2018)
8. Kavraki, L.E., Švestka, P., Latombe, J.C., Overmars, M.H.: Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE Trans. Robot. Automat. 12(4), 566–580 (August 1996)
9. Konc, J., Miller, B.T., Štular, T., Lešnik, S., Woodcock, H.L., Brooks, B.R., Janežič, D.: ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. Journal of Chemical Information and Modeling 55(11), 2308–2314 (2015), https://doi.org/10.1021/acs.jcim.5b00534, pMID: 26509288
10. Konc J, J.D.: Probis-ligands: a web server for prediction of ligands by examination of protein binding sites. Nucleic Acids Research (2014)
11. Lahoda, M., Mesters, J.R., Stsiapanava, A., Chaloupkova, R., Kuty, M., Damborsky, J., Kuta Smatanova, I.: Crystallographic analysis of 1,2,3-trichloropropane biodegradation by the haloalkane dehalogenase DhaA31. Acta Crystallographica Section D 70(2), 209–217 (Feb 2014), https://doi.org/10.1107/S1399004713026254
12. Latombe, J.C.: Robot Motion Planning. Kluwer Academic Publishers, Boston, MA (1991)
13. Latombe, J.C.: Motion planning: A journey of robots, molecules, digital actors, and other artifacts. Int. Journal of Robotics Research 18(11), 1119–1128 (1999)
14. Lavalle, S.M.: Rapidly-exploring random trees: A new tool for path planning. Tech. rep. (1998)

15. Levitt, M.: Protein folding by restrained energy minimization and molecular dynamics. J. Mol. Biol. 170, 723–764 (1983)
16. Ma B, Elkayam T, W.H.N.R.: Proteinprotein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. In: Proceedings of the National Academy of Sciences of the United States of America. vol. 100, pp. 5772–5777 (2003)
17. Pavelka, A., Sebestova, E., Kozlikova, B., Brezovsky, J., Sochor, J., Damborsky, J.: Caver: Algorithms for analyzing dynamics of tunnels in macromolecules. IEEE/ACM Transactions on Computational Biology and Bioinformatics 13(3), 505–517 (2016)
18. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.: Ucsf chimera: A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25(13), 1605–16012 (2004)
19. Ravindranath, P.A., Forli, S., Goodsell, D.S., Olson, A.J., Sanner, M.F.: Autodockfr: Advances in protein-ligand docking with explicitly specified binding site flexibility. PLOS Computational Biology 11(12) (2015), https://doi.org/10.1371/journal.pcbi.1004586
20. S. MacKerell, J.A., Brooks, I.C.: Charmm fluctuating charge force field for proteins: Ii protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. Journal of Computational Chemistry 25, 1504–1514 (2004)
21. Saeed Izadi, B.A., Onufriev, A.V.: Protein-ligand electrostatic binding free energies from explicit and implicit solvation. Journal of ChemicalTheory and Computation 11(9), 44504459 (2015)
22. Singh, A.P., Latombe, J.C., Brutlag, D.L.: A motion planning approach to flexible ligand binding. In: Int. Conf. on Intelligent Systems for Molecular Biology (ISMB). pp. 252–261 (1999)
23. Stsiapanava, A., Dohnalek, J., Gavira, J.A., Kuty, M., Koudelakova, T., Damborsky, J., Kuta Smatanova, I.: Atomic resolution studies of haloalkane dehalogenases DhaA04, DhaA14 and DhaA15 with engineered access tunnels. Acta Crystallographica Section D 66(9), 962–969 (Sep 2010), https://doi.org/10.1107/S0907444910027101
24. Tagliasacchi, A., Alhashim, I., Olson, M., Zhang, H.: Mean curvature skeletons. Eurographics Symposium on Geometry Processing 2012 27(1) (2012)
25. Tan, K.P., Nguyen, T.B., Patel, S., Varadarajan, R., Madhusudhan, M.S.: Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pka of ionizable residues in proteins. Nucleic Acids Research 41(W1), W314 (2013), +http://dx.doi.org/10.1093/nar/gkt503
26. Tana, K., Radka, C., Jan, B., Zbynek, P., Eva, S., Martin, H., Morteza, K., Maryia, P., Daryna, K., Ivana, K.S., Pavlina, R., Rudiger, E., T., B.U., Jiri, D.: Engineering Enzyme Stability and Resistance to an Organic Cosolvent by Modification of Residues in the Access Tunnel. Angewandte Chemie International Edition 52(7), 1959–1963, https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201206708
27. Veronika, L., David, B., Tatyana, P., Pavlina, R., Tana, K., Eva, S., Kuta, S.I., Jan, B., Radka, C., Jiri, D.: Balancing the Stability–Activity Trade-Off by Fine-Tuning Dehalogenase Access Tunnels. ChemCatChem 7(4), 648–659, https://onlinelibrary.wiley.com/doi/abs/10.1002/cctc.201402792
28. Yu R, Peethambaram HS, F.R.e.a.: Kinetics of 1,2-dichloroethane and 1,2-dibromoethane biodegradation in anaerobic enrichment cultures. applied and environmental microbiology 79(4), 1359–1367 (2013)
29. Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B.: Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27(15), 2083 (2011), +http://dx.doi.org/10.1093/bioinformatics/btr331
30. Zhu, H., Pisabarro, M.T.: Mspocket: an orientation-independent algorithm for the detection of ligand binding pockets. Bioinformatics 27(3), 351 (2011), +http://dx.doi.org/10.1093/bioinformatics/btq672