Using Guided Motion Planning to Study Binding Site Accessibility

Diane Uwacu
duwacu@tamu.edu
Texas A&M University
College Station, Texas

Abigail Ren
abigailren@vassar.edu
Vassar College
Poughkeepsie, New York

Shawna Thomas
sthomas@cse.tamu.edu
Texas A&M University
College Station, Texas

Nancy M. Amato*
namato@illinois.edu
University of Illinois Urbana-Champaign
Champaign, Illinois

ABSTRACT

Computational methods are commonly used to predict protein-ligand interactions. These methods typically search for regions with favorable energy that geometrically fit the ligand, and then rank them as potential binding sites. While this general strategy can provide good predictions in some cases, it does not do well when the binding site is not accessible to the ligand. In addition, recent research has shown that in some cases protein access tunnels play a major role in the activity and stability of the protein's binding interactions. Hence, to fully understand the binding behavior of such proteins, it is imperative to identify and study their access tunnels. In this work, we present a motion planning algorithm that scores protein binding site accessibility for a particular ligand. This method can be used to screen ligand candidates for a protein by eliminating those that cannot access the binding site. This method was tested on two case studies to analyze effects of modifying a protein's access tunnels to increase activity and/or stability as well as study how a ligand inhibitor blocks access to the protein binding site.

CCS CONCEPTS

• **Applied computing** → *Life and medical sciences*; *Computational biology*;

KEYWORDS

binding site accessibility, ligand binding, motion planning

*To whom correspondence should be addressed: namato@illinois.edu.

BCB '20, September 21–24, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7964-9/20/09…$15.00
https://doi.org/10.1145/3388440.3414707

ACM Reference Format:
Diane Uwacu, Abigail Ren, Shawna Thomas, and Nancy M. Amato. 2020. Using Guided Motion Planning to Study Binding Site Accessibility . In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '20), September 21–24, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3388440.3414707

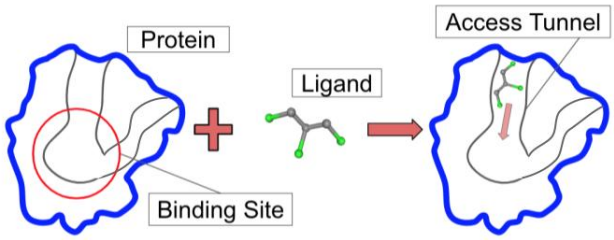

Figure 1: Binding Site Accessibility: The ligand can reach the binding site through an access tunnel

1 INTRODUCTION

Protein activity is triggered or inhibited by small molecules called ligands. Understanding protein-ligand interactions is essential in drug design where ligands can be made more specific to their target protein or in Bio-Engineering where proteins can be modified to increase the velocity or stability of their reactions with target ligands. A large number of proteins have buried binding sites that can only be reached when the ligand travels through the protein's tunnels [19]. The binding process of such structures thus depends on the accessibility of the binding site. Figure 1 illustrates this concept. In these types of proteins, residues along the tunnel can affect the binding process if they change in ways that alter the volume and energy of the tunnel. This can be explored in protein engineering research. Modifying binding site residues most of time results in changing the binding selectivity of the protein[11]. To increase a protein activity for a ligand, modifications can be applied to tunnel instead of binding site residues

Existing methods that describe protein tunnels for binding focus on the volume of the tunnels. These methods use a probe radius and find tunnels that would fit a sphere of such radius. By not factoring in the flexibility of the ligand, these methods are too restrictive because some rejected tunnels could be traversed by the ligand rotating its atoms. Furthermore, to use these methods correctly, the user needs to have an idea of a reasonable radius to use. The choice is not always intuitive since the radius of the probe must be smaller than that of the ligand in order to accomodate the margin of error created by the flexibility of the ligand.

We propose sampling-based motion planning to discover and analyze the accessibility of a protein as it relates to a specific ligand. Motion planning consists of finding a valid path for a movable object from a start to a goal position. This can be used to study the accessibility of a protein binding site by a ligand. Using motion planning, a ligand is considered a moving object and its valid configurations are sampled and connected through the protein tunnels until a path is found leading the ligand to the binding site.

Since a protein environment has narrow passages, we build a topological skeleton[28] of the protein free space to guide our planning algorithm. The idea of skeleton-guided planning was first explored by Denny et al [4] for robotics applications using the topological skeleton to guide a rapidly-exploring random tree (RRT). We implemented a topological skeleton-guided rapidly-exploring random graph (RRG) to increase our chances of finding paths with optimal energy. The algorithm builds a roadmap of all the free regions that a ligand can reach and finds low-energy paths connecting the binding site to the protein's outside surface.

We tested this method on case studies where protein binding activity depended on binding site accessibility. In all the cases considered, this method accurately described the accessibility of the binding site by a ligand. Comparison against a state-of-the-art protein tunnel identification algorithms shows the value of considering the geometric and energetic flexibility of the ligand when studying its binding potential with a given protein.

2 RELATED WORK

2.1 Predicting Protein-Ligand Binding

Protein-ligand binding research focusses on one of the two challenges: binding site identification or binding site accessibility. In this section we look at some of the existing solutions for each.

There are many computational methods that identify the binding site of a protein. *Geometry-based methods* [6, 22, 29, 34] focus on fitting a ligand inside the protein binding pocket. *Energy-based methods* [24] compute different forces between the ligand and the protein and return protein residues that minimize the energy cost upon interacting with the ligand. Based on one or both of these two features, *template-based methods* predict the protein's binding activity based on similarities between the protein and other known protein structures [12, 17]. *Statistical methods*[3, 33] apply statistical analysis to the features reported by geometry and/or energy based tools and return a percentage confidence.

Sampling-based methods have been used to rank protein binding sites before. Singh et al[26] used gradient descent methods to uniformly sample a set number of nodes in the protein environment and create dense sample clusters around those configurations that exhibit the lowest energy. 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. Bayazit et al [1] used Obstacle-based probabilistic roadmap strategy (PRM) to sample on the surface of the protein to detect a binding site based on geometric fit and optimal energy. This approach leverages human input to determine the optimal minima based on haptic feedback. This method relies on the expertise of the the user to give reliable feedback.

Table 1: Protein-Ligand interaction Prediction Methods(Guided RRG is shown in bold)

Method	**Properties**			
	Ligand-specific	***Geometry***	***Energy***	***Accessibility***
Probis-CHARMMing [12]	✓		✓	
CAVER [19]		✓		✓
MoMA-LigPath [5]	✓	✓		✓
Guided RRG	✓	✓	✓	✓

There are recent computational methods that focus on binding site accessibility. Prokop et al [19] presented *CAVER*, a software that predicts protein tunnels given a radius of a spherical probe that should fit the protein tunnels. Although the method is only geometry-based, it shed light on the presence of tunnels in protein, which has inspired a lot of research to explore the implications. Kaushik et al. [8] recently used CAVER to study the effects of access tunnel engineering on the protein catalytic properties. Vonasek et al [31] proposed a motion planning method that explores protein tunnels guided by the geometry of the environment. They run a Molecular Dynamics simulation of the protein and took snapshots of its different configurations. They used those snapshots to construct an approximate free space of the protein and used it for guidance to rapidly explore the environment. This method is not ligand-specific and uses a user-defined radius for a spherical probe. Simeon et al [5] used *MoMA-LigPath*, a motion planning method originally designed for disassembly problems. They defined protein-ligand binding as a disassembly problem in which a ligand is disassembled from a protein avoiding collisions between

non-bonded atoms. This method is also geometry-based and does not do an energy analysis, which may result in paths that are not energetically feasible.

Table 1 summarizes some of the methods described above and their specific properties. Note that Guided RRG is the only method with the potential to accurately find geometrically and energetically accessible tunnels for a specific ligand to a given protein's binding site.

2.2 Sampling-based planning

As the environment gets more complex or as the degrees of freedom of the movable object increase, exact motion planning solutions become intractable [23], hence the popularity of sampling-based algorithms that use randomness and heuristics to solve high-dimensional complex problems. Sampling-based algorithms can be subdivided into two categories, namely, the tree-based rapidly exploring random trees (RRT) [16] and the graph-based probabilistic roadmaps (PRM) [9]. Typically, RRT-based algorithms optimize speed and are easily applied to robotics problems with nonholonomic constraints, while PRM-based algorithms optimize connectivity and can be applied to multi-query problems. Rapidly exploring random graphs (RRG) [7] are a hybrid of both algorithms that combines their strengths. There have been variants of heuristics applied to these algorithms to improve planning in specific conditions.

Topological guidance involves using the workspace structure to direct how sampling-based planners explore the environment. Some motion planning approaches use the workspace decomposition for planning, especially to target narrow passages. For example, in [31] voronoi diagrams are used to guide an RRT through a protein environment and discover possible protein tunnels that could connect to a binding site.

Several topology-guided planners use workspace skeletons, which are graphs that capture the topological features of the environment. One such planner is Dynamic Region-biased RRT (DR-RRT) [4]. This strategy guides the planner to sample only from particular regions based on the workspace topology. A skeleton, or graph that maps the essence of the free workspace topology, directs region selection. These regions are created, sampled based on their previous exploration success rate, and then destroyed as the planner explores new regions.

3 OVERALL APPROACH

Our method involves three main steps: modeling, planning and analysis.

3.1 Modeling

We obtain protein data from the protein data bank (PDB)[21] and construct their geometric structure using CHIMERA [20]. Before constructing the geometry, we make sure that no ligand molecule is already occupying the binding site. figures 2 shows a cartoon representation of a protein geometry in ribbon and solid spherical format.

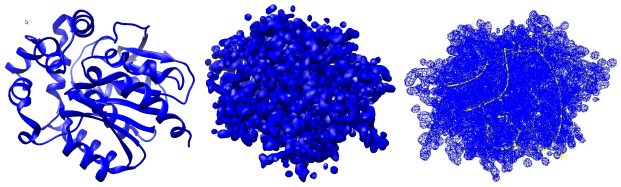

Figure 2: Protein model (DhaA): (Left)The cartoon format converted to a spherical model (middle). The free space decomposition and resulting mean curvature skeleton (right).

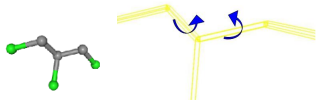

Figure 3: Ligand TCP (left) modeled as a flexible multi-linkage with two degrees of freedom(right).

Given the complexity of the protein geometry, it is hard to plan inside its free space. We start by computing the free space of the protein by subtracting the protein geometry from its convex hull. A skeleton representing the free space of the protein can be constructed by using CGAL's built-in mean curvature skeleton algorithm. The last image of Figure 2 shows resulting skeleton.

We model the ligand as a flexible robot. Given the coordinates of the ligand's atoms, we construct a multi-linkage. Every two adjacent atoms are represented by a link with the same length as the distance between them. Three successive atoms result into two links. The distance between the atoms is kept constant by conserving the plannar and dihedral angles between the links. Figure 3 shows a model of ligand 1,2,3-trichloropropane. The ligand has eight degrees of freedom: 6 degrees of freedom for moving and rotating in three dimensions and 2 degrees of freedom to model the flexibility of two links.

We consider the impact of two different properties on protein-ligand relationship: geometry and energy. As common in robotics applications, we make sure that the geometry of the ligand does not intersect with the geometry of the protein. The energy between a ligand and a protein is computed following equation[18] shown in Equation 1. For every pair of ligand atom i and protein atom j, $d_{i,j}$ is the distance between both atoms, A and B are Van der waals constants and q_i and q_j are charges:

$$E = \Sigma i, j \frac{332 * q_i q_j}{d_{i,j}} + \frac{A}{d_{i,j}^{12}} - \frac{B}{d_{i,j}^{6}} \quad (1)$$

3.2 Planning

We create a start configuration by sampling a valid configuration of the ligand that has maximum contact with the protein binding site residues. As shown in lines 3 and 4 of Algorithm 1, once we find that root configuration, we start building a roadmap of valid ligand configurations through

Algorithm 1 Guided-RRG

Input: Protein p, Site s, Ligand l, Clustering threshold ϵ

1: $WS \leftarrow$ `GetWorkspaceSkeleton`(p)
2: $start \leftarrow l$.`SampleRoot`(s)
3: $R \leftarrow \phi$
4: $r \leftarrow$ `GetInitialRegion`$(WS, start)$
5: **while** $\neg done$ **do**
6: $\quad C_r \leftarrow$ `GetChildren`(r)
7: $\quad r \leftarrow$ `SelectRegion`(C_r)
8: $\quad R \leftarrow$ `GrowRoadmap`(r)
9: $paths \leftarrow$ `Query`(R)
10: $tunnels \leftarrow$ `Cluster`$(paths, \epsilon)$ **return** $tunnels$

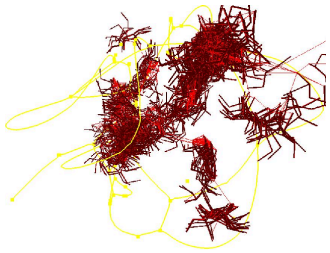

Figure 4: RRG roadmap displayed in red, biased along skeleton (yellow), guiding planning through protein narrow passages

the protein free space starting with the root configuration. We create a region that contains the root configuration and the closest skeleton node to it.

We incrementally grow the roadmap with steps shown in lines 5 to 9 by extending the sampling region along the edges of the skeleton. The skeleton shows the connected free space inside the protein (also known as *Workspace*), and it is through planning that we try to place a valid ligand configuration in the free space. The set of all possible valid ligand configurations (C_{free}) is a subset of the free workspace that the skeleton represents.

Once a region reaches the end of a skeleton edge, it sparks new regions along the adjacent edges adjacent to it. The function `SelectRegion`(C_r) in line 7 chooses the region to advance. This region can be chosen based on the previous success history along the edge that it's on. The steps described so far are modelled after the Dynamic Regions RRT algorithm [4]. In order to increase the chance of finding optimal paths, we construct a graph (RRG [7]) instead of a tree (RRT [15]). Another change is in how the next region to explore is chosen in line 7. One option is to annotate the skeleton with clearance and energy information around the skeleton edges, so that regions with high clearance and low energy are explored first.

Figure 4 shows the roadmap guided by workspace skeleton to discover tunnels of the protein. Note that not all edges of the skeleton are followed by the roadmap. This is because some areas represented by the skeleton are not accessible to the ligand due to the ligand size and energy constraints.

3.3 Analysis

In the post-processing steps described in the last two lines of Algorithm 1 we find through the roadmap all paths connecting the root configuration to the configurations outside the protein surface. Those paths are clustered into tunnels based on their topological distance [25], which is similar to the path edit distance between them. When two paths are in the same connected free space, their topological distance is close to their euclidean distance, while the distance between two paths separated by obstacle space is much longer.

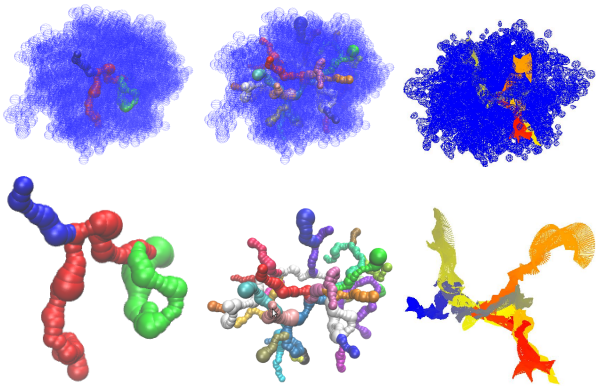

Figure 5: Tunnels found in DhaA. Left: 3 tunnels found with CAVER using probe radius of 0.85. Middle: 33 tunnels found with CAVER using probe radius of 0.7. Right: 5 tunnels found with Guided_RRG using ligand TCP.

Finally, we implemented a scoring function for the accessibility of tunnels. The tunnel score takes into consideration the volume of the tunnel and the energy minima and bottlenecks along the tunnel. To represent the volume of a tunnel, we compute the path spread along the tunnel. This is the distance between paths that make up a tunnel. This value is indicative of the ligand's accessible volume of the tunnel. For energy feasibility, minimum energy directly contributes to the score while maximum energy inversely affects the score. Lastly, we reward tunnels that are traversed by more than one path. We weight each of the three component equally as shown in Equation 2.

$$score = \frac{1}{3} * (min(E) + \frac{1}{max(E)}) + \frac{1}{3} * min(R) + \frac{1}{3} * |P| \quad (2)$$

where E is the interaction energy, R the path spread along the tunnel, and $|P|$ is the number of paths found in the tunnel. We consider minimum R because the smallest radius along the tunnel is limiting factor of accessibility.

Figure 5 shows tunnels found in protein DhaA for ligand TCP after clustering in comparison to what is found with CAVER. Our method identified five tunnels while CAVER identified three tunnels when we reduced the probe radius from the default of 0.9$\AA$ to 0.85$\AA$. When we reduced the radius further to 0.7$\AA$, CAVER returned 33 tunnels containing all the tunnels that we had identified. It is possible that

the ligand would not fit inside all the tunnels returned by CAVER.

4 EXPERIMENTS

4.1 Case Studies

We run experiments to analyze tunnels in different configurations of a protein that have undergone tunnel modifications. We intend to show that accessible proteins are easier to plan for than closed-off ones. We were able to study two problems related to protein-ligand binding. Figure 6 shows flexibility and size details of the studied ligands.

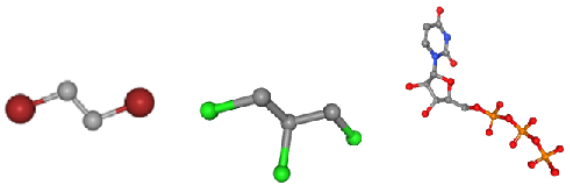

Figure 6: Tested ligands: (left) EDB has 6 DOFs and a radius of 2.16Å, (middle) TCP has 8 DOFs and a radius of 2.77Å, (right) rNTP has 13 DOFs and a radius of 8.36Å.

4.1.1 Effects of modifying access tunnel residues. Haloalkane dehalogenase enzymes (HLD) degrade halogenated hydrocarbons, which makes them essential in the biodegradation of soil pollutants like 1,2,3-trichloropropane (TCP, PDB ID: 3KP), an environmental pollutant with no known natural biodegradation pathway and 1,2-Dibromoethane (EDB), a synthetic chemical that was used as a fumigant until it was banned in the 1980s [28]. A variant of the HLD, DhaA degrades TCP and EDB into an alcohol, a proton and a halide. However, DhaA's activity is limited by its low stability and vulnerability to the harsh conditions of its reactive environment. Mutants have been engineered in laboratories in an attempt to enhance DhaA's stability and/or activity towards TCP and EDB specifically. In previous experiments, researchers showed how modifications made to residues lining the tunnels connecting the binding site of the protein affected the reaction between DhaA and the two haloalkane compounds.

The active site of DhaA is connected to the protein surface by two major access tunnels. Klavana et al[10] identified three more tunnels through molecular modeling. We used our method to study the changes in those tunnels.

We studied these five specific DhaA variants:

- **DhaA** (PDB ID: 4HZG) is the wild type structure described above.
- **DhaA15** (PDB ID: 3FBW) was proposed by Stiapanava et al[27] to increase the stability of the complex after it reacts with the ligand. To achieve this, residues close to the binding site were modified in order to reduce the size of the main and side tunnels and decrease the accessibility of the binding site to the solvent. As a result, DhaA15's activity towards ligand TCP was reported to be more stable compared to the wild type.
- **DhaA31** (PDB ID: 3RK4) was proposed by Lahoda et al [14] to increase its activity towards TCP, certain residues at the mouth of the tunnel were modified, which reduced the size of the binding pocket to fit the ligand and leave little space to the solvent. This was reported to increase DhaA's activity towards TCP.
- **DhaA80** (PDB ID: 4F60)) was proposed by Koudelakova et al[13] to increase the stability of the DhaA-EDB complex after reaction. Four residues were added at the opening ot the access tunnel, which reduced its size and decreased effects that the solvent dimethyl sulfoxide (DMSO) has on the binding process. Although higher stability was achieved, the modifications were reported to slow down the reaction between the protein and the ligand because tunnels were too narrow for the ligand.
- **DhaA106** (PDB ID: 4WCV) was proposed by Liskova et al[30] as a balance between stability and activity for ligand EDB. The size of the tunnel was increased by modifying two of the four residues added in DhaA80. This change increased activity and maintained the stability of the reaction between DhaA106 and ligand EDB.

4.1.2 Effects of inhibitors on accessibility. Nonstructural protein 5B (NS5B) is a protein in the Hepatitis C Virus genome and it plays the key role of replicating HCV's viral RNA. Inhibitors targetting NS5B are currently in clinical trials, which maked it an interesting protein to study. Yan et al [32] proposed an isothiazole compound that inhibits NS5B's binding site residues by making contact with the binding site of NS5B and disrupting the entrance of rNTP (PDB ID: UTP) substrate for de novo initiation.

We studied the two configurations of NS5B:

- **NS5B in unbound state** (PDB ID: 1C2P)[2]: This configuration is expected to have high affinity for rNTP
- **NS5B in bound state** (PDB ID: 2IJN)[32]: NS5B is bound to an isothiazole compound that is expected to block the accessibility of the binding site to rNTP

4.2 Experimental Setup

All methods were implemented in a C++ motion planning library developed in the Parasol Lab at Texas A&M University. All experiments were executed on a desktop computer running CentOS 7 with an Intel® Core™ i7-3770 CPU at 3.4 GHz, 16 GB of RAM, and the GNU g++ compiler version 4.8.5.

For each protein-ligand pair, the planning phase was capped at 30 minutes or until there were no new regions to discover. We performed 10 random trials and report the average roadmap construction time.

We run the systems with Probis-CHARMMing and CAVER for comparison. Finding a reasonable number of tunnels using CAVER is highly dependent on the radius of the probe. We used a radius of 0.9$\mathring{A}$ because it was suggested by the authors

to be the ideal radius, a lower radius to see if we would find more tunnels, and the actual radius of the ligand represented by R^* in Table 2. R^* ranges from 2.16for ligand EDB to 8.36 for nNTP. Probis-CHARMMing returns a list of ligands for each binding site with a score between 0 and 5 indicating the confidence in the predition. Scores shown in Table 3 are normalized.

The narrower the tunnels, the longer it took our planning algorithm to finish. Run times for the different steps of our algorithm are reported in Figure 10. The resulting roadmap was searched for optimal paths using Dijkstra's searching algorithm, and tunnels found after path clustering were scored with Equation 2 and the score ranges are reported in Table 2. In addition, we showed energy profile of the best path in each identified tunnel.

4.3 Results and Discussion

Run time can be indicative of accessibility. Figure 10 reports the run time for each tested ligand through the main three steps of our algorithm. We focus on planning time since skeleton time is only indicative of the complexity of the protein geometry. Time spent during analysis depends on the number of paths originally found after planning. Path profiles are shown in Figures 7, 8, and 9. The different colors signify paths representing each tunnel.

4.3.1 Analysis for TCP. We observe that DhaA15 took longer in planning. This is because there were energy bottlenecks close to the binding site which made it hard to grow the roadmap out of the binding pocket. DhaA31 on the other hand had the lowest planning time showing the balance between stability and activity. DhaA31's energy profile in Figure 7 does not have any energy bottleneck. This makes planning easier because there is a higher chance of creating valid configurations and connections compared to DhaA15.

Table 2 shows that Guided-RRG finds tunnels that indicate the accessibility observed in the shown DhaA mutants. The number of tunnels found using CAVER fluctuated a lot with the radius.

Table 3 shows that our protein score in relation to the ligand is more consistent to experimental observations than Probis. Probis returns lower scores for the mutants that are known to be more active and stable than the wild type structure.

Figure 7 shows energy changes along paths. Values are shown from left to right going from the binding site to the protein surface. Profiles show that wild type DhaA has a bottleneck at the mouth of the access tunnel that DhaA15 and DhaA31 do not have. DhaA15 has bottlenecks closer to the binding site, suggesting that a ligand would not easily leave the pockect after binding. DhaA31 is a balance between the two. As expected, paths found in DhaA15 reached very low optimal energies around the binding site.

4.3.2 Analysis for EDB. Planning time for EDB was lower than that of TCP across all variants, and this is because EDB is smaller in size and has fewer degrees of freedom than TCP. DhaA106 has the shortest planning time, because its tunnels are wider than the other two variants. The number of tunnels that we found is consistent with the documented modification made to each variant. All four tunnels found in DhaA80 were ranked high due to their low energy bottleneck. CAVER did not return any path for DhaA106 until we reduce radius to 0.5Å, which was surprising given that this variant is more open than the other two. Energy profiles in Figure 8 tell an interesting story. We observed that modifications made to widen DhaA106's access tunnels raised the energy close to the binding site while achieving a low optimal energy, which allows DhaA106 to achieve higher stability and accessibility. This advantage is indicated by the score in Table 3.

4.3.3 Analysis for rNTP. This ligand has 13 degrees of freedom, which increased its planning time compared to the previous two ligands. Planning in the NS5B-isothiazole complex did not result in any paths through the main tunnel. The side tunnel that we had previously discovered in the unbound NS5B was still accessible. This suggests that although the isothiazole compound successfully blocks access through the main tunnel, NS5B might still be able to react with rNTP through the side tunnel. Compared to Probis-CHARMMing, our method's score agrees more with the change in activity between rNTP and the two versions of protein NS5B.

Table 2: Number of tunnels found

Ligand	**Protein**	**CAVER**			**Guided-RRG**		
		Probe Radius			**Tunnel score**		
		0.7	0.9	R*	> 0.85	> 0.65	All
TCP	DhaA	33	0	0	1	2	5
	DhaA15	19	1	0	3	3	7
	DhaA31	16	0	0	4	5	5
EDB	DhaA	33	0	0	1	3	5
	DhaA80	18	0	0	4	4	4
	DhaA106	0	0	0	5	6	8
rNTP	Unbound NS5B	75	12	0	1	1	2
	Bound NS5B	61	4	0	0	0	1

[a]R* = ligand's true radius.

5 CONCLUSION

We presented a fully automated framework for studying the dynamics of ligand binding by considering the accessibility of the binding site to the ligand. The guided planning method rapidly explores the free space of the protein, a problem that is made hard by the high dimensionality of the ligand's degrees of freedom and the complexity of the protein geometry. Tests run over haloalkane dehalogenase enzymes as well as Hepatitis C protein configurations show that motion planning algorithms can indicate the activity and stability

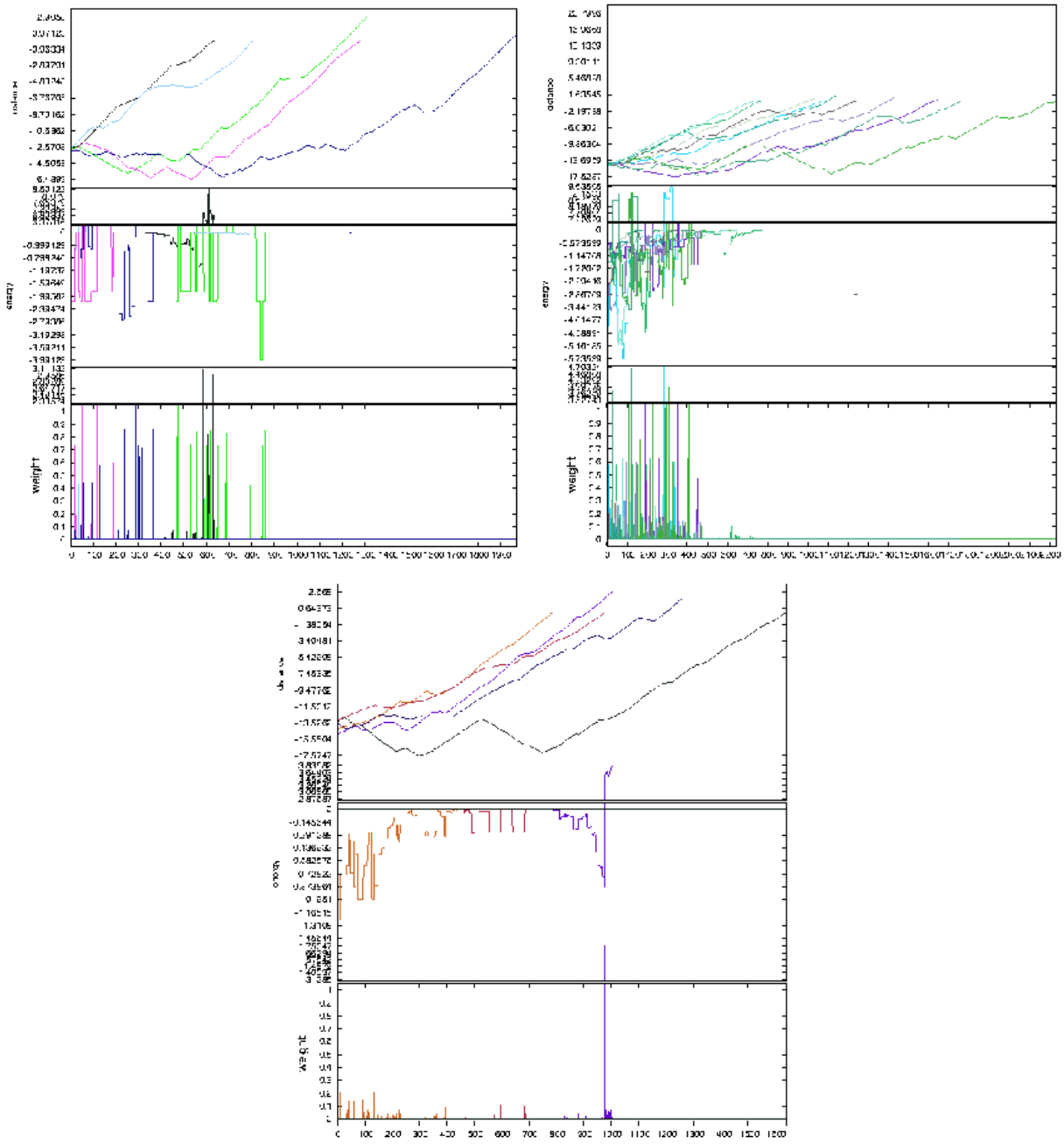

Figure 7: Energy profile for identified tunnels in DhaA variants with ligand TCP. Top: distance to the protein convex surface (negative when inside the protein tunnels). Middle: energy value between a ligand configuration and the protein atoms. Bottom: edge weights indicating the cost to transition from one configuration to the next. (Left) DhaA, (middle) DhaA15, (right) DhaA31.

of a protein-ligand reaction based on the accessibility of the protein's binding site. We showed that with skeleton guidance, identification of and trajectories through the protein's tunnels is facilitated. For faster and more effective planning, we can annotate the skeleton with environment information such as clearance and energy and bias planning towards regions with higher clearance and lower energy in order to find higher ranked tunnels first. In the future, we plan to improve the analysis step of our algorithm by using machine learning clustering methods to identify tunnels faster after planning.

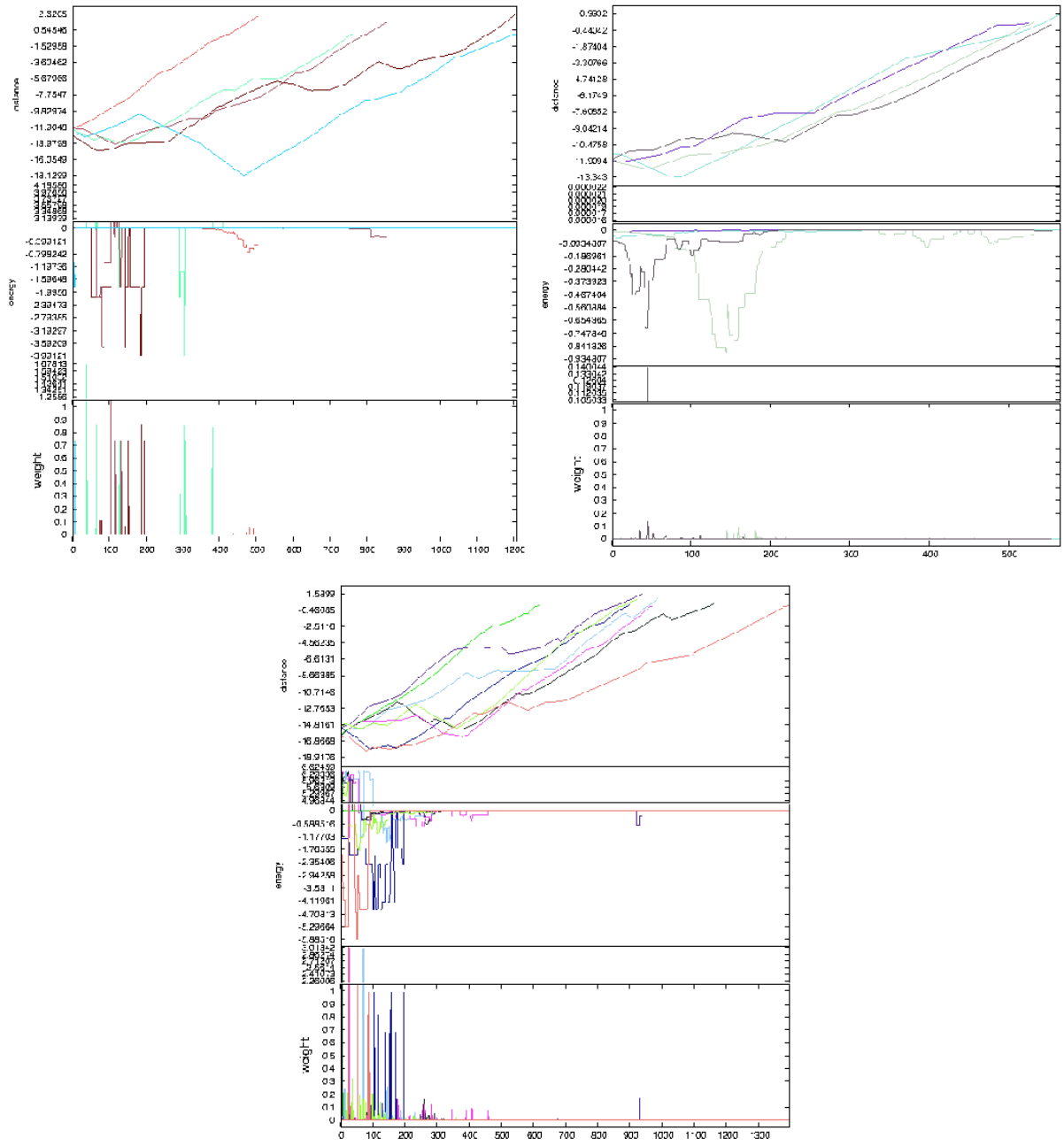

Figure 8: Energy profile for identified tunnels in DhaA variants with ligand EDB. Top: distance to the protein convex surface (negative when inside the protein tunnels). Middle: energy value between a ligand configuration and the protein atoms. Bottom: edge weights indicating the cost to transition from one configuration to the next. (Left) DhaA, (middle) DhaA80, (right) DhaA106.

In addition, we plan to expand the guidance of annotated skeletons to dynamic environments to explore the flexibility of the protein.

REFERENCES

[1] O. B. Bayazit, G. Song, and N. M. Amato. 2001. Ligand Binding with OBPRM and Haptic User Input: Enhancing Automatic Motion Planning with Virtual Touch. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*. 954–959. This work was also presented as a poster at *RECOMB 2001*.

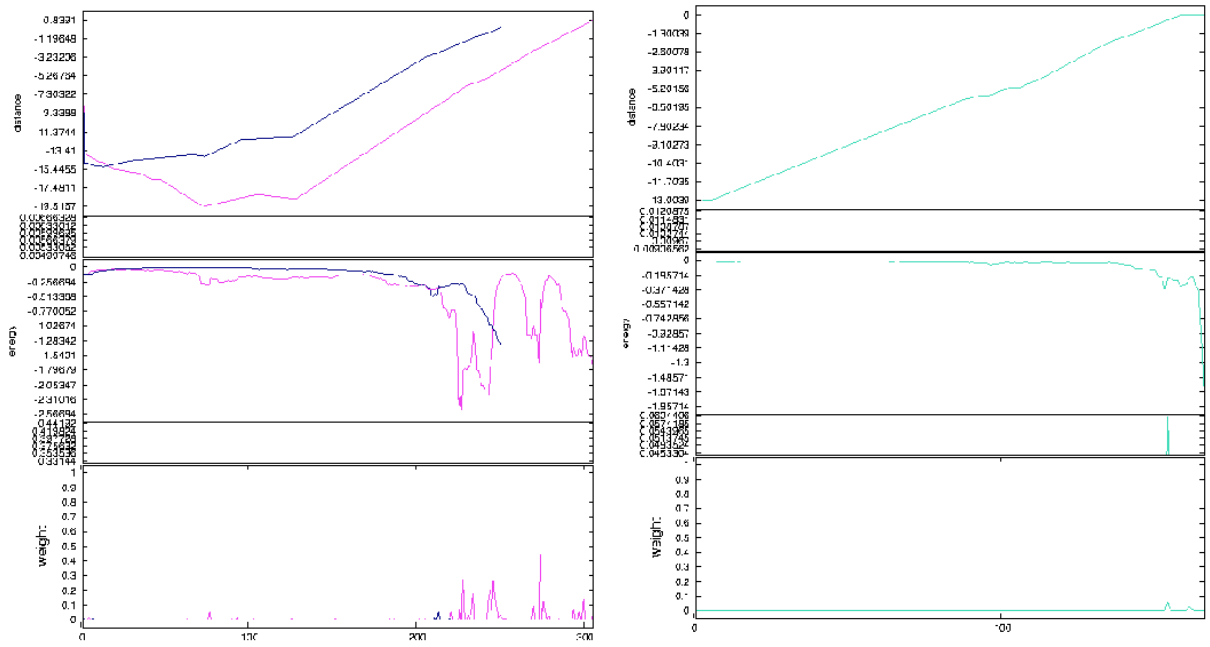

Figure 9: Energy profile for identified tunnels in NS5B with rNTP. Top: distance to the protein convex surface (negative when inside the protein tunnels). Middle: energy value between a ligand configuration and the protein atoms. Bottom: edge weights indicating the cost to transition from one configuration to the next. (Left) NS5B in bound state, (right) NS5B in open state.

Table 3: Protein binding score

Ligand	Protein	Probis	Guided-RRG
TCP	DhaA	0.96	0.67
	DhaA15	0.82	0.71
	DhaA31	0.59	0.93
EDB	DhaA	0.86	0.67
	DhaA80	0.93	0.77
	DhaA106	0.73	0.87
rNTP	Unbound NS5B	0.96	0.69
	Bound NS5B	0.96	0.49

[2] Lesburg Charles, Cable Michael, Ferrari Eric, Hong Zhi, Mannarino Anthony, and Weber Patricia. 1999. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Structural Biology* 6 (1999), 937–943.

[3] B. Y. Chen. 2014. 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] Jory Denny, Read Sandström, Andrew Bregger, and Nancy M. Amato. 2020. Dynamic Region-biased exploring Random Trees. In *Alg. Found. Robot. XII*. Springer. (WAFR '16).

[5] Didier Devaurs, Léa Bouard, Marc Vaisset, Christophe Zanon, Ibrahim Al-Bluwi, Romain Iehl, Thierry Siméon, and Juan Cortés. 2013. MoMA-LigPath: A web server to simulate protein-ligand unbinding. *Nucleic Acids Research* vol. 41 (May 2013), 297–302.

[6] Bingding Huang and Michael Schroeder. 2006. LIGSITE cs: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Structural Biology* 6 (2006), 1472–6807. http://dx.doi.org/10.1186/1472-6807-6-19

[7] Rahul Kala. 2013. Rapidly exploring random graphs: motion planning of multiple mobile robots. *Advanced Robotics* 27, 14 (2013), 1113–1122. https://doi.org/10.1080/01691864.2013.805472 arXiv:https://doi.org/10.1080/01691864.2013.805472

[8] Shubhangi Kaushik, Sérgio M. Marques, Prashant Khirsariya, Kamil Paruch, Lenka Libichova, Jan Brezovsky, Zbynek Prokop, Radka Chaloupkova, and Jiri Damborsky. 2018. Impact of the access tunnel engineering on catalysis is strictly ligand specific. *Federation of European Biochemical Societies Journal* 285, 8 (2018), 1456–1476.

[9] L. E. Kavraki, P. Švestka, J. C. Latombe, and M. H. Overmars. 1996. Probabilistic Roadmaps for Path Planning in High-Dimensional Configuration Spaces. *IEEE Trans. Robot. Automat.* 12, 4 (Aug. 1996), 566–580.

[10] Martin Klvana, Martina Pavlova, Tana Koudelakova, Radka Chaloupkova, Pavel Dvorak, Zbynek Prokop, Alena Stsiapanava, Michal Kuty, Ivana Kuta-Smatanova, Jan Dohnalek, Petr Kulhanek, Rebecca C. Wade, and Jiri Damborsky. 2009. Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored Using Classical and Random Acceleration Molecular Dynamics Simulations. *Journal of Molecular Biology* 392, 5 (2009), 1339 – 1356. https://doi.org/10.1016/j.jmb.2009.06.076

[11] Piia Kokkonen, David Bednar, Gaspar Pinto, Zbynek Prokop, and Jiri Damborsky. 2019. Engineering enzyme access tunnels. *Biotechnology Advances* (2019), 107386. https://doi.org/10.1016/j.biotechadv.2019.04.008

[12] Janez Konc, Benjamin T. Miller, Tanja Štular, Samo Lešnik, H. Lee Woodcock, Bernard R. Brooks, and Dušanka Janežič. 2015. ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *Journal of Chemical Information and Modeling* 55, 11 (2015), 2308–2314. https://doi.org/10.1021/acs.jcim.5b00534 arXiv:https://doi.org/10.1021/acs.jcim.5b00534 PMID: 26509288.

[13] Tana Koudelakova, Radka Chaloupkova, Jan Brezovsky, Zbynek Prokop, Eva Sebestova, Martin Hesseler, Morteza Khabiri, Maryia Plevaka, Daryna Kulik, Ivana Kuta Smatanova, Pavlina Rezacova,

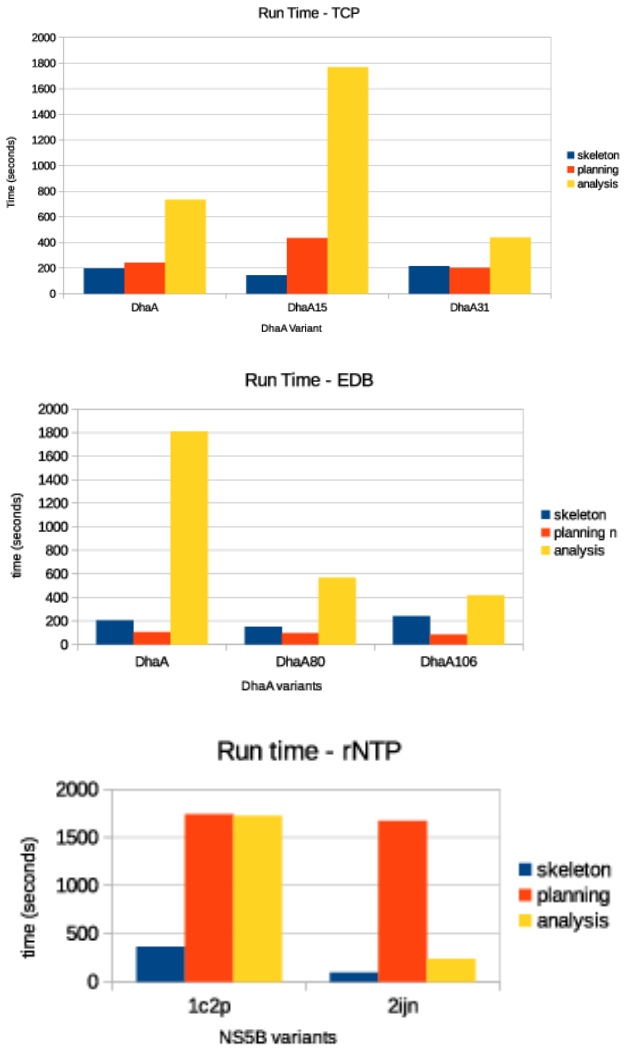

Figure 10: Run time results for the three cases studied.

Rudiger Ettrich, T. Bornscheuer Uwe, and Jiri Damborsky. [n. d.]. Engineering Enzyme Stability and Resistance to an Organic Cosolvent by Modification of Residues in the Access Tunnel. *Angewandte Chemie International Edition* 52, 7 ([n. d.]), 1959–1963. https://doi.org/10.1002/anie.201206708

[14] Maryna Lahoda, Jeroen R. Mesters, Alena Stsiapanava, Radka Chaloupkova, Michal Kuty, Jiri Damborsky, and Ivana Kuta Smatanova. 2014. Crystallographic analysis of 1,2,3-trichloropropane biodegradation by the haloalkane dehalogenase DhaA31. *Acta Crystallographica Section D* 70, 2 (Feb 2014), 209–217. https://doi.org/10.1107/S1399004713026254

[15] Steven M. Lavalle. 1998. *Rapidly-Exploring Random Trees: A New Tool for Path Planning.* Technical Report. Iowa State University.

[16] S. M. LaValle and J. J. Kuffner. 2001. Randomized Kinodynamic Planning. *Int. J. Robot. Res.* 20, 5 (May 2001), 378–400.

[17] Wolfson H Nussinov R Ma B, Elkayam T. 2003. Protein–protein 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. 5772–5777. https://doi.org/10.1073/pnas.1030237100

[18] N. Pattabiraman, M. Levitt, T. E. Ferrin, and R. Langridge. 1985. Computer Graphics in Real Time Docking with Energy Calculation and Minimization. *J. of Computational Chemistry* 6 (1985), 432–436.

[19] Antonin Pavelka, Eva Sebestova, Barbora Kozlikova, Jan Brezovsky, Jiri Sochor, and Jiri Damborsky. 2016. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 13, 3 (2016), 505–517.

[20] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. UCSF Chimera: A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* 25, 13 (2004), 1605–16012.

[21] Protein Data Bank [n. d.]. The Protein Data Bank. `http://www.rcsb.org/pdb/`.

[22] P. A. Ravindranath, S. Forli, D. S. Goodsell, A. J. Olson, and M. F. Sanner. 2015. 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

[23] J. H. Reif. 1979. Complexity of the Mover's Problem and Generalizations. In *Proc. IEEE Symp. Found. Comp. Sci. (FOCS)*. San Juan, Puerto Rico, 421–427.

[24] Boris Aguilar Saeed Izadi and Alexey V. Onufriev. 2015. Protein-Ligand Electrostatic Binding Free Energies from Explicit and Implicit Solvation. *Journal of ChemicalTheory and Computation* 11, 9 (2015), 4450–4459. https://doi.org/10.1021/acs.jctc.5b00483

[25] Read Sandström, Andrew Bregger, Ben Smith, Shawna Thomas, and Nancy M. Amato. 2018. Topological Nearest-Neighbor Filtering for Sampling-based Planners. In *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*. 3053–3060.

[26] Amit P. Singh, Jean-Claude Latombe, and Douglas L. Brutlag. 1999. A Motion Planning Approach to Flexible Ligand Binding. In *Int. Conf. on Intelligent Systems for Molecular Biology (ISMB)*. 252–261.

[27] A. Stsiapanava, J. Dohnalek, J. A. Gavira, M. Kuty, T. Koudelakova, J. Damborsky, and I. Kuta Smatanova. 2010. Atomic resolution studies of haloalkane dehalogenases DhaA04, DhaA14 and DhaA15 with engineered access tunnels. *Acta Crystallographica Section D* 66, 9 (Sep 2010), 962–969. https://doi.org/10.1107/S0907444910027101

[28] Andrea Tagliasacchi, Ibraheem Alhashim, Matt Olson, and Hao Zhang. 2012. Mean Curvature Skeletons. *Eurographics Symposium on Geometry Processing 2012* 27, 1 (2012).

[29] Kuan Pern Tan, Thanh Binh Nguyen, Siddharth Patel, Raghavan Varadarajan, and M. S. Madhusudhan. 2013. 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 (2013), W314. https://doi.org/10.1093/nar/gkt503

[30] Liskova Veronika, Bednar David, Prudnikova Tatyana, Rezacova Pavlina, Koudelakova Tana, Sebestova Eva, Smatanova Ivana Kuta, Brezovsky Jan, Chaloupkova Radka, and Damborsky Jiri. [n. d.]. Balancing the Stability–Activity Trade-Off by Fine-Tuning Dehalogenase Access Tunnels. *ChemCatChem* 7, 4 ([n. d.]), 648–659. https://doi.org/10.1002/cctc.201402792

[31] V Vonasek and B Kozlikova. 2018. Tunnel detection in protein structures using sampling-based motion planning. 1 (2018), 713–720.

[32] Shunqi Yan, Todd Appleby, Esmir Gunic, Jae Hoon Shim, Tania Tasu, Hongwoo Kim, Frank Rong, Huaming Chen, Robert Hamatake, Jim Z. Wu, Zhi Hong, and Nanhua Yao. 2007. Isothiazoles as active-site inhibitors of HCV NS5B polymerase. *Bioorganic & Medicinal Chemistry Letters* 17, 1 (2007), 28 – 33. https://doi.org/10.1016/j.bmcl.2006.10.002

[33] Zengming Zhang, Yu Li, Biaoyang Lin, Michael Schroeder, and Bingding Huang. 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* 27, 15 (2011), 2083. https://doi.org/10.1093/bioinformatics/btr331

[34] Hongbo Zhu and M. Teresa Pisabarro. 2011. MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics* 27, 3 (2011), 351. https://doi.org/10.1093/bioinformatics/btq672