A topology approach towards modeling activities and properties on a biomolecular surface

Aakriti Upadhyay
Department of Computer Science
University at Albany, SUNY
NY, USA
aupadhyay@albany.edu

Tuan Tran
Department of Computer Science
University at Albany, SUNY
NY, USA
ttran3@albany.edu

Chinwe Ekenna
Department of Computer Science
University at Albany, SUNY
NY, USA
cekenna@albany.edu

***Abstract*—Geometric features of protein surfaces play an important role in the identification of biomolecular structures, functions, and interactions. These features have been crucial in predicting binding sites for protein-ligand or protein-protein interactions. This paper introduces simplicial complexes and discrete Morse theory to extract important geometric information on the protein surface. Using the extracted geometric information, we provide possible intermediate conformations around the protein surface as the ligand travels to the binding site. We compare the efficiency of our method with the state-of-art-method in terms of computation time and total complexes needed to generate the topological structure of the protein surface. We also show comparable relevance of the binding affinity of our method in relation to the known native protein binding site.**

***Index Terms*—Protein surface, Simplicial complex, Discrete Morse theory.**

I. Introduction

Protein-ligand interactions are crucial to a wide range of biological activities and functions in any organism, including cell metabolism, signal transduction, muscle contraction, and immune systems. Many essential cellular processes such as controlling the function of enzymes, transport, and most regulatory mechanisms rely on physical interactions between proteins. Therefore, protein-ligand interactions network analysis is essential to gain comprehensive knowledge on the control mechanism and organization of a living cell.

Analyzing and extracting useful information from the molecular surface of proteins is a fundamental problem in structural biology. These surfaces contain essential biological information that helps us understand properties such as the geometrical organization of interacting residues, precise identification of the borders of each interaction site, energy potential at interaction sites that allow for strong versus weak binding, and the locations where artificial molecules (e.g., drugs) can best bind. Understanding these properties has life-saving biological implications, including aiding in the development of therapeutic drugs, vaccines, and point of care technologies. However, predicting protein-ligand interactions purely from structure remains an important challenge in structural biology [1]. The recent study proposed multiple approaches to capture molecular surface patterns with functional relevance, such as three-dimensional Zernike descriptors [2], [3] and geometric invariant fingerprint descriptors [4]. Nevertheless, the scope of these approaches is limited since they proposed hand-crafted descriptors and manually optimized the protein surface features, making it difficult to determine the right set of features for a given task.

This research is supported in part by NSF awards CRII-IIS-1850319.

Seminal work on the application of motion planning algorithms to the study of proteins was published, in [5]. The motion-planning-inspired methods are widely applied in molecular simulations for the computation of conformational transitions of proteins [6], the study of the protein folding process [7]–[9], and the analysis of protein-ligand interactions [10], [11]. One of the widely used motion planning methods is a sampling-based motion planning (SBMP) approach [12]. In this work, we apply an SBMP method to study the dynamics of ligand binding to an identified binding site.

Contribution: we present a framework that extracts the geometric features of the protein molecular surfaces via simplicial complexes and discrete Morse theory. We construct the topological structure, i.e., simplicial complex, around the protein surface using the Vietoris-Rips complex method. On the constructed simplicial complex, we apply the variant of the discrete Morse function defined in our work [13] to identify critical points on the surface of the protein molecule. The critical points information used in our algorithm provides different conformations of the ligand molecule around the protein surface, which help in planning the feasible trajectory of the ligand to a protein's active binding site. We perform experiments for ten different proteins with two ligand molecules, and our result shows a correlation between the binding affinity of our results to the known binding site of these ligands. We also show an improved performance of our method with the baseline method in [14].

II. Related Work

A. Protein-ligand interaction

Proteins do not function in isolation, and interactions help reveal important functions and properties. However, the efficiency of these interactions depends on the dynamical and kinetic features of any pair. Protein-ligand interactions are most often analyzed using a theoretical (physics-based) or statistical (knowledge-based) approach. The complicated interactions of biological molecules as captured by physics-based models carry a large overhead of needed processing resources [15]. Geometric features, such as coordinates, distances, angles,

surface areas [16], and curvatures [17] are important descriptors of biological molecules. However, geometric features often involve too much structural detail and are frequently computationally intractable for large biological molecules data sets [18]. Therefore, efficient extraction and processing of these geometric features still needs to be developed.

B. Geometric curvatures of protein surface

The exploration of the geometric features of a protein molecular surface enhances the understanding of molecular morphology and molecular mechanism and allows significant applications to drug design and protein-ligand interactions. The work in [14] presented a variational multiscale strategy for the unified geometric and physical modeling of an aqueous biomolecular system in the Lagrangian representation.

Another work [19] used the cartesian representation to evaluate second-order computational algorithms for six different 3D structure curvature descriptors leading to a good prediction of protein-ligand binding sites. However, the evaluation metric of the geometric properties turned out to be expensive and limited the robustness for use in the analysis of the interaction of macromolecules, e.g., proteins, membranes, DNAs, and RNAs.

C. Sampling based motion planners

Motion planning is, generally, formulated using the notion of configuration space [20]. The configuration space (or conformation space) is the set of all possible conformations an object (a robot or a biomolecule) can take, and the number of dimensions of this space equals the number of degrees of freedom of the object (i.e., the number of parameters needed to describe the pose of the robot or biomolecule). The regions in the configuration space free of all internal and external constraints are called $\mathcal{C}_{free}$. Sampling-based planners are often classified into two categories: graph-based methods such as the Probabilistic Roadmap Method (PRM) [21] and tree-based methods such as Rapidly-exploring Random Tree (RRT) [22]. These methods have shown several applications in the study of proteins [9], [23], [24]. Recent work in [11] generated a topology skeleton tunnel to plan a path for the ligand to the accessible binding site of protein using a skeleton guided rapidly-exploring random graph. However, the construction and analysis of the tunnel become complex as the complexity of protein increases.

In this work, we present a method to extract the geometric features of the protein surfaces using the discrete Morse function. Our algorithm extracts the non-degenerate critical points of the protein surfaces and provides possible ligand conformations around them. These features provide maximum and minimum curvatures information of the protein surface beneficial in the prediction of geometrically favorable protein-ligand interaction binding sites.

III. Methodology

A. Background definitions

Discrete Morse theory, originally defined by Forman [25], is a discrete analog of the classical smooth Morse theory. We first define the principal mathematical concepts, i.e., abstract simplicial complex and Vietoris-Rips complex.

Definition 1: (Abstract Simplicial complex) An abstract simplicial complex K, i.e., a collection of sets closed under the subset operation, is a generalization of a graph useful in representing higher-than-pairwise connectivity relationships.

The elements of the set are called vertices, and the set itself is a simplex. The vertices refer to ligand conformation in the conformation space.

Definition 2: (Vietoris-Rips complex) Given a set S of points in Euclidean space E, the Vietoris-Rips complex $R(S)$ is the abstract simplicial complex whose k-simplices are the subsets of $k+1$ points in X with diameter that is at most ε.

We perform steps from [26] to generate simplicial complex $R(S)$ to capture the topological structure of the protein surface, i.e., vertices, edges, triangles. We apply the discrete Morse function on the same simplicial complex to extract the critical points information of the surface. The discrete setting of Morse theory avoids the overhead of differential geometry, thus, reducing the computation complexity for high dimensional structures.

In this work, the protein surface is modeled as a rigid object. S is the set of all ligand conformations present in the simplicial complex $R(S)$. These conformations are generated at a radial distance 2ϱ away from the surface to avoid collisions, such that $S \subseteq \mathcal{C}_{free}$. We take ϱ as the diameter of the circumscribed circle of the ligand molecule. Considering the above parameters, we define the discrete Morse function as follows.

Definition 3: Let D be the Euclidean distance function that measures the distance between the point $x \in \mathcal{C}_{free}$ and the nearest point y on the protein surface P, that is, $D(x) = \min_{y \in P} \|x - y\|$.

Definition 4: Let $\Gamma(y, \varrho)$ be a density function where $\varrho > 0$ and y is the point on the protein surface. The function Γ counts all neighbors close to y in S within distance ϱ.

Definition 5: Let f be a discrete Morse function on $R(S)$ restricted to the vertices of the Vietoris-Rips complex. One option was defined in [13]. This is applied in this paper and formally defined at any point in conformation space by

$$f(x) = D(x) \times \Gamma(y, \varrho). \tag{1}$$

Please refer to [13] for our expanded definition and theorems.

Definition 6: (Critical points) The set of critical points is defined as the set of non-degenerate points on the surface of protein when the given discrete Morse function f reaches its extreme values, i.e. local minima or maxima.

Definition 7: (Feasible critical points) This set is defined as all possible ligand conformations in S at a radial distance of ϱ from a critical point on the protein surface. In other words, it is the union of intersections of vertices in S within the metric balls of radius ϱ centered at some critical point.

B. Extraction of geometric features

Algorithm 1 describes how we construct a simplicial complex around the protein surface by sampling and connecting ligand conformations in method *ConstructComplex*. Using the sampling condition from [26], the algorithm performs topological collapse to remove redundant topological information, i.e.,

vertices and edges, and provides a skeleton of the simplicial complex around the protein surface in line 3, i.e., a surface mesh. It applies discrete Morse function f to this simplicial complex to identify the local maxima and minima curvatures of the protein surface, in line 4. The identified critical points of the surface are the highest and the lowest peak points on the surface at which function f reaches its extremum. For function f, the distance becomes an equalizer, and the density becomes an essential contributing factor affecting the morse value density at the surface curvatures.

Algorithm 1 Path planning to protein binding site

Input: P: Protein surface model, R: Planned path to the binding site, H: set of ligand conformations around the protein surface.

1: Let $R \leftarrow \{\phi\}$.
2: $S \leftarrow ConstructComplex(P)$; ◁ Refer Def. 2
3: $TopologicalCollapse(S)$; ◁ Refer [26]
4: $C \leftarrow IdentifyCriticalPoints(S)$; ◁ Refer Def. 5, 6
5: $F \leftarrow GetFeasiblePoints(S, C)$; ◁ Refer Def. 7
6: $H = S \bigcap F$
7: $R = Query(H)$
8: **return** $\{H, R\}$

The algorithm extracts the feasible critical points at radial distance ϱ from the identified critical points of the protein surface, in line 5. These conformations are identified at proximity to the protein surface and are part of a simplicial complex $R(S)$, refer Def.7. The extracted geometric information map is used to plan a path for the ligand conformation from start to the binding site conformation in lines 6-7. The output of our algorithm is an extracted geometric information map consisting of critical points, feasible critical points, and a pathway from the start conformation to the binding site conformation.

IV. Model Transformation

We obtain protein data from the protein data bank (PDB) [27], [28] and construct their geometric structure using CHIMERA [29]. Figure 1 shows the graphical representation of 4JNO protein, its high-dimensional surface model, and the extracted geometric map.

We consider 10 proteins and 2 ligand bio-molecules to study and understand the protein surfaces and their geometries. The high dimensional surface models of proteins are represented as a stationary rigid body in the conformation space. We construct a flexible linkage model of the ligand using the covalent bond length and angle measurement derived from CHIMERA. Figure 2 shows the backbone structure of the SIA ligand molecule and our corresponding model. Each covalent bond is simulated as one link of the robot of length 1.53 $\mathring{A}$, and the part of the C-N bond of the molecule is surrounded with a sphere of radius 2.7 $\mathring{A}$. Similarly, we perform the transformation for SO4 ligand where the covalent bond length is 1.47 $\mathring{A}$, and the angle between the pair of S-O bonds is 120° each, as shown in Figure 3.

The protein studied are listed in Table I and range from 36 to 1708 residues. The proteins selected include three *Plasmodium Falciparum (PF)* pathogen proteins, i.e., 1SQ6, 1TQX, and 3NTJ, and one DNA protein (7OXS). *PF* inflicts the most damage and is responsible for most malaria-related deaths. The high mutational capacity coupled with its changing metabolism makes the development of malaria drug treatments an evolving problem. These proteins were selected because they have been identified as potential drug targets from our previous work [30].

We follow the same steps of transformation for all protein and ligand bio-molecules to avoid the loss of biological significance. We transform the coordinates of the known binding site provided in the PDB file of each protein into the goal conformation for our ligand model in the conformation space. The start and goal positions are highlighted in red and blue color, respectively, as shown in Figure 4. The caption of each protein surface model represents the ligand molecule used for the protein-ligand interaction experiment. Here, protein 4JNO binds with ligand SIA, and the rest binds with ligand SO4.

V. Result Analysis

The experiments were executed on a Dell Optiplex 7040 desktop machine running OpenSUSE operating system, and the algorithms were implemented in C++ language. All results were averaged over 5 random trials for geometric map generation and 5 random trials for roadmap planning time for each protein.

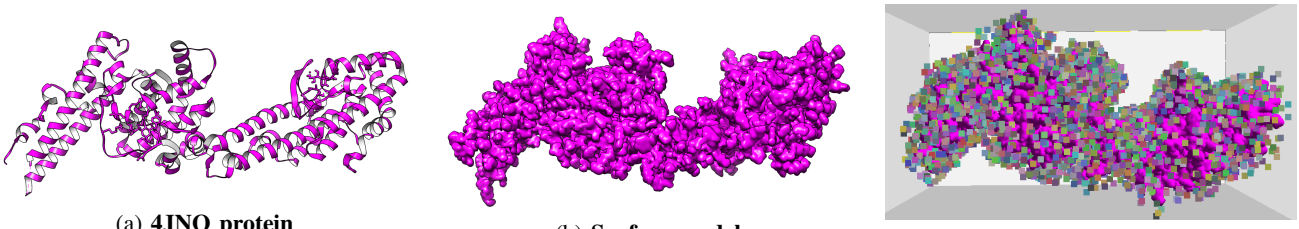
A. Computing surface complexes

We compare the performance of our method with the Delaunay-refinement-based method from the TetGen library due to its relevance to our approach, as seen in [14]. Computation time and total complexes generated to capture the topological structure of the protein surface were recorded. In Figure 5, we observe that the Delaunay method required a longer time to generate complexes for all proteins compared to our method. The numbers of complexes generated by the Delaunay method for proteins 3NTJ and 5JBE are equivalently highest among all proteins, but the computation time difference between them is significant and inconsistent. On the other hand, to maintain the skeleton structure of the simplicial complex our method shows a consistent relation between the computation time and the total generated complexes. The calculation performed to construct complexes around the surface does not affect the overall performance resulting in less memory overhead. Thus, our method is more reliable and faster in capturing the topological structure of the protein surface by generating fewer complexes in low computation time than the baseline method.

On the generated surface complex, the work [14] performed numerical calculations to identify the surface curvatures. Instead, our method provides an automated framework for identifying the minimum and maximum curvatures of the protein surfaces, i.e., critical points, and generates feasible critical points around them, as discussed next.

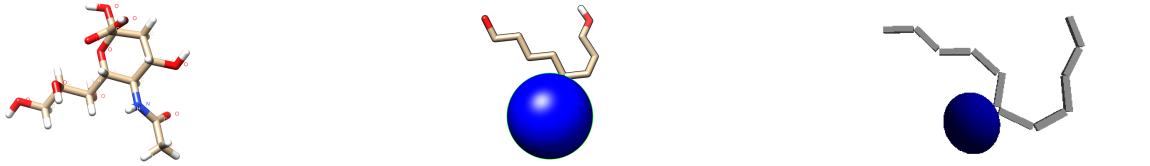
B. Motion planning towards a binding site

Our method generates a geometric information map for all 10 proteins, e.g., shown in Figure 1c. We input the map to PRM [21] method as an initial graph to solve query from start conformation to the binding site conformation. We show the



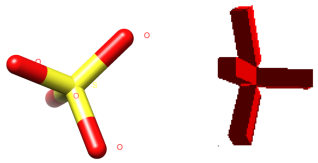
(a) **4JNO protein** (b) **Surface model** (c) **Geometric information map**

Fig. 1: The figure shows the multiscale surface model of the 4JNO protein and the geometric features (critical points) detected on the protein surface. The geometric information map provides the point size view of the ligand molecule conformation around the surface.



(a) **Ligand SIA** (b) **SIA with backbone structure** (c) **Corresponding robot model**

Fig. 2: Robot transformation of SIA ligand. The C-N bond part of the ligand is enclosed by a blue sphere.



(a) **Ligand SO4** (b) **Robot model**

Fig. 3: SO4 ligand into robot transformation.

distribution of time taken to construct a simplicial complex, compute a geometric map, and plan a path to the binding site. Figure 6a shows the total computation time taken by our method to capture protein surface topological and geometrical information, and Figure 6b shows the time taken to plan a successful path to the goal conformation. We observe that for large protein structures like 1ZRL, 3NTJ, 4JNO, 5JBE, and 5ZT1, the computation of geometric information is higher than the complex construction time. On the other hand, the ligand conformations generated around the identified curvatures of protein surfaces contribute to the smooth navigation of the ligand to the goal conformation at a safe distance from the protein. The path planning time for all proteins is negligible as the highest planning time recorded was 150 seconds for the SIA ligand around the 4JNO protein surface.

Figure 7 shows screenshots of the planned path for the ligand SIA around the 4JNO protein surface to the goal conformation. The different view angles reflect the motion of the ligand biomolecule around the protein surface using the ligand conformation generated by our method. Our method was able to capture the biological aspect of the protein-ligand interaction using the geometric information map. We produce the same results for the remaining nine proteins.

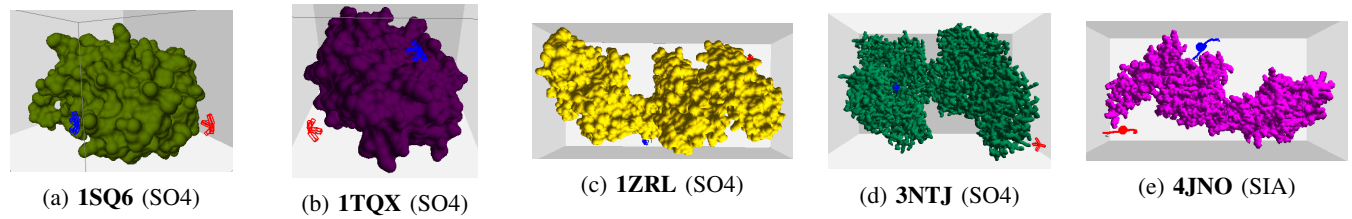
C. Binding affinity of goal conformation

We are interested in analyzing the relevance of our goal conformation with known binding sites for the ligand. We validate the identified ligand conformations for each protein with the native binding pose using the binding affinity measure. We use the molar Gibbs free energy ΔG (binding affinity) to determine the relevance for the binding pose. Gibbs free energy is a thermodynamic potential that measures the capacity of a thermodynamic system to do maximum or reversible work at a constant temperature and pressure (isothermal, isobaric) [31]. The protein-ligand binding occurs only when the change in Gibbs free energy ΔG of the system is negative, i.e., when the system reaches an equilibrium state at constant pressure and temperature. Table I show the ΔG value for the ligand at goal conformation for all 10 proteins.

Our method provides a closely relevant binding affinity for our ligand goal conformation compared with the native binding poses of the ligand for each protein. Our method successfully captures the geometric features of the protein surfaces and plans a path for ligand biomolecule to the binding site without losing its biological significance.

VI. Discussion and Future Work

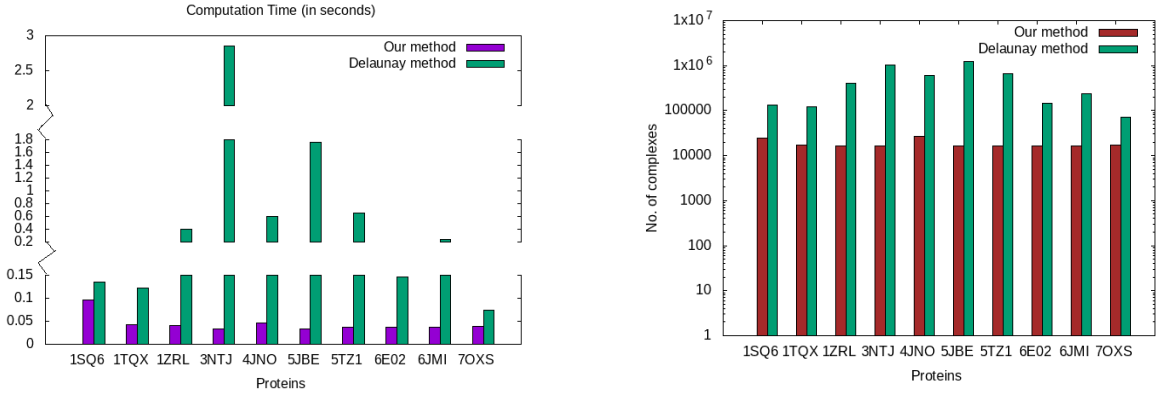
The paper presented an automated framework that extracts the geometric features, i.e., maximum and minimum curvature, of the protein surface and provides the possible ligand conformation around the identified curvatures. Our algorithm applies the discrete Morse function to the constructed simplicial complex around the protein surface and generates a geometric information map. The results show that our method captures the topological structure of the protein surface more efficiently compared to the baseline method in terms of computation time and the number of complexes generated. Our results show a strong correlation between the goal conformation of the ligand with known binding sites for the studied proteins via



(a) **1SQ6** (SO4) (b) **1TQX** (SO4) (c) **1ZRL** (SO4) (d) **3NTJ** (SO4) (e) **4JNO** (SIA)

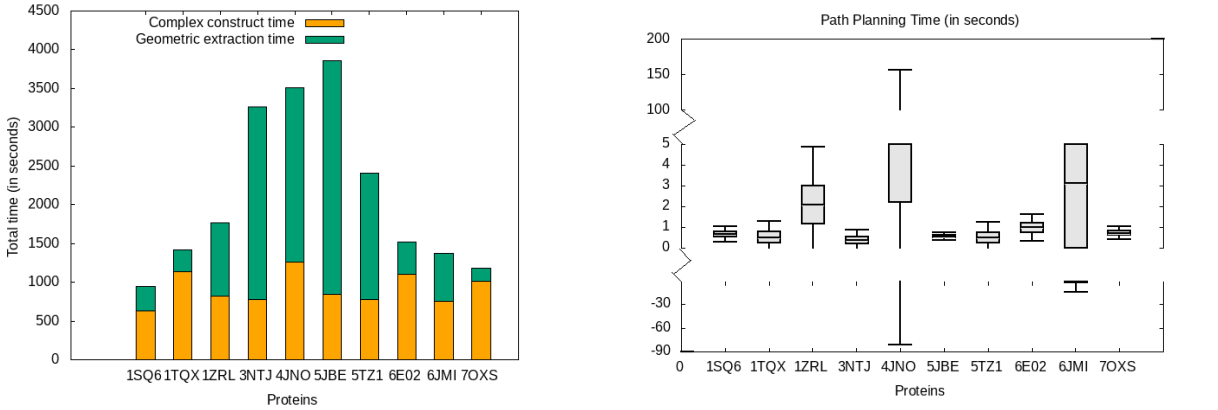
(f) **5JBE** (SO4) (g) **5ZT1** (SO4) (h) **6E02** (SO4) (i) **6JMI** (SO4) (j) **7OXS** (SO4)

Fig. 4: Protein surface models studied



(a) **Total computation time (in seconds)** (b) **Total complexes generated**

Fig. 5: Qualitative performance analysis of generated surface complex.

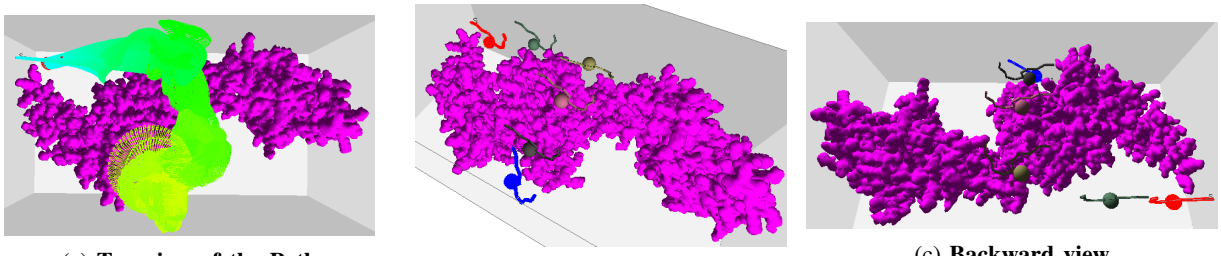


(a) **Total time taken to construct the simplicial complex and extract geometric features of the protein surface.** (b) **Path planning time range shown over mean and standard deviation values of all trials.**

Fig. 6: Quantitative time analysis of our algorithm.

TABLE I: Binding Affinity (Kcal/mol) for the ligand at goal conformation compared to ligand native pose for each protein.

	1SQ6	1TQX	1ZRL	3NTJ	4JNO	5JBE	5ZT1	6E02	6JMI	7OXS
Our method	-5.6	-5.2	-5.3	-5.4	-7.0	-5.2	-5.2	-5.2	-5.2	-5.2
Native poses	-5.4	-5.4	-5.2	-5.2	-7.0	-5.2	-5.3	-5.2	-5.2	-5.7



(a) **Top view of the Path** (b) **Side view** (c) **Backward view**

Fig. 7: Path planned using feasible critical points information (ligand conformations generated around 4JNO protein surface) to the binding site. The side and backward views show intermediate ligand conformations of the path.

a binding affinity measure. While our method considers the tertiary-level structure of the protein, i.e., protein surfaces, the geometric information map shows the potential to contribute to the prediction of potential drug target sites when combined with machine learning techniques. In future work, we plan to utilize the geometric information of protein surface as classified potential binding sites and inculcate machine learning techniques with a diverse combination of protein and ligand bio-molecules to analyze the prediction accuracy of the protein active binding sites based on the learned geometric features.

References

[1] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter *et al.*, "Structure-based prediction of protein–protein interactions on a genome-wide scale," *Nature*, vol. 490, no. 7421, pp. 556–560, 2012.

[2] S. Daberdaku and C. Ferrari, "Antibody interface prediction with 3d zernike descriptors and svm," *Bioinformatics*, vol. 35, no. 11, pp. 1870–1876, 2019.

[3] X. Zhu, Y. Xiong, and D. Kihara, "Large-scale binding ligand prediction by improved patch-based method patch-surfer2. 0," *Bioinformatics*, vol. 31, no. 5, pp. 707–713, 2015.

[4] S. Yin, E. A. Proctor, A. A. Lugovskoy, and N. V. Dokholyan, "Fast screening of protein surfaces using geometric invariant fingerprints," *Proceedings of the National Academy of Sciences*, vol. 106, no. 39, pp. 16 622–16 626, 2009.

[5] A. P. Singh, J.-C. Latombe, D. L. Brutlag *et al.*, "A motion planning approach to flexible ligand binding." in *ISMB*, 1999, pp. 252–261.

[6] A. Enosh, B. Raveh, O. Furman-Schueler, D. Halperin, and N. Ben-Tal, "Generation, comparison, and merging of pathways between protein conformations: gating in k-channels," *Biophysical journal*, vol. 95, no. 8, pp. 3850–3860, 2008.

[7] D. Becerra Romero, "Predicting protein folding pathways using ensemble modeling and sequence information," *Doctoral dissertation*, 2017.

[8] M. Moll, D. Schwarz, and L. E. Kavraki, "Roadmap methods for protein folding," *Protein Structure Prediction*, pp. 219–239, 2008.

[9] C. Ekenna, S. Thomas, and N. M. Amato, "Adaptive local learning in sampling based motion planning for protein folding," *BMC systems biology*, vol. 10, no. 2, p. 49, 2016.

[10] M. S. Apaydın, C. E. Guestrin, C. Varma, D. L. Brutlag, and J.-C. Latombe, "Stochastic roadmap simulation for the study of ligand-protein interactions," *Bioinformatics*, vol. 18, no. suppl_2, pp. S18–S26, 2002.

[11] D. Uwacu, A. Ren, S. Thomas, and N. M. Amato, "Using guided motion planning to study binding site accessibility," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2020, pp. 1–10.

[12] H. Choset, K. M. Lynch, S. Hutchinson, G. A. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge, MA: MIT Press, June 2005.

[13] A. Upadhyay, B. Goldfarb, W. Wang, and C. Ekenna, "Planning among obstacles: using discrete morse theory for topological extraction of geometric features," State University of New York, Albany, NY (United States), Tech. Rep., 2020. [Online]. Available: https://aakupadhyay.github.io/assets/images/DMT_in_cspace.pdf

[14] X. Feng, K. Xia, Z. Chen, Y. Tong, and G.-W. Wei, "Multiscale geometric modeling of macromolecules ii: Lagrangian representation," *Journal of computational chemistry*, vol. 34, no. 24, pp. 2100–2120, 2013.

[15] J. Graves, J. Byerly, E. Priego, N. Makkapati, S. V. Parish, B. Medellin, and M. Berrondo, "A review of deep learning methods for antibodies," *Antibodies*, vol. 9, no. 2, p. 12, 2020.

[16] Q. Zheng, S. Yang, and G. Wei, "Molecular surface generation using pde transform," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 28, pp. 291–316, 2012.

[17] D. D. Nguyen and G.-W. Wei, "The impact of surface area, volume, curvature, and lennard–jones potential to solvation modeling," *Journal of computational chemistry*, vol. 38, no. 1, pp. 24–36, 2017.

[18] Z. Cang, L. Mu, and G.-W. Wei, "Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening," *PLoS computational biology*, vol. 14, no. 1, p. e1005929, 2018.

[19] K. Xia, X. Feng, Z. Chen, Y. Tong, and G.-W. Wei, "Multiscale geometric modeling of macromolecules i: Cartesian representation," *Journal of Computational Physics*, vol. 257, pp. 912–936, 2014.

[20] T. Lozano-Perez, "Spatial planning: A configuration space approach," *Computers, IEEE Transactions on*, vol. 100, no. 2, pp. 108–120, 1983.

[21] L. E. Kavraki, P. Švestka, J. C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE Trans. Robot. Automat.*, vol. 12, no. 4, pp. 566–580, August 1996.

[22] S. M. LaValle and J. J. Kuffner, "Randomized kinodynamic planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 1999, pp. 473–479.

[23] S. Thomas, X. Tang, L. Tapia, and N. M. Amato, "Simulating protein motions with rigidity analysis," *J. Comput. Biol.*, vol. 14, no. 6, pp. 839–855, 2007, special issue of Int. Conf. Comput. Molecular Biology (RECOMB) 2006.

[24] L. Tapia, S. Thomas, and N. M. Amato, "A motion planning approach to studying molecular motions," *Communications in Information and Systems*, vol. 10, no. 1, pp. 53–68, 2010, special issue in honor of Michael Waterman.

[25] R. Forman, "A user's guide to discrete morse theory," *Sém. Lothar. Combin*, vol. 48, p. 35pp, 2002.

[26] A. Upadhyay, W. Wang, and C. Ekenna, "Approximating c free space topology by constructing vietoris-rips complex," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2517–2523.

[27] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.

[28] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, "The protein data bank: a computer-based archival file for macromolecular structures," *Journal of molecular biology*, vol. 112, no. 3, pp. 535–542, 1977.

[29] 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.

[30] T. Tran and C. Ekenna, "Protein binding pose prediction via conditional variational autoencoding for plasmodium falciparum," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 2448–2455.

[31] M. K. Gilson and H.-X. Zhou, "Calculation of protein-ligand binding affinities," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 36, pp. 21–42, 2007.