A geometric and topological analysis of the binding behavior of Intrinsically Disordered Proteins

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

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

***Abstract*—Intrinsically disordered proteins (IDPs) play vital regulatory roles in biology, emphasizing the significance of understanding their conformational behavior and interaction mechanisms during protein-ligand or protein-protein interactions. However, IDP analysis becomes difficult due to the lack of a stable structure. In this work, we investigate the binding behavior of an IDP using the surface information of the interacting protein complex. Our algorithm extracts the protein surface model's topological and geometric features and predicts a geometrically favorable binding pose for an IDP around it. A transition path is planned to the predicted bound position to help evaluate the RMSD deviation in the IDP conformation structure. In our results, we use the Zernike descriptor metric to examine the structural homology of the binding pose and analyze the molar Gibbs free energy (binding affinity) of experimental conformation.**

***Index Terms*—Intrinsically Disordered Proteins, Geometric features, Binding behavior.**

I. Introduction

Intrinsically Disordered Proteins (IDPs) are involved in many biological processes, such as cell regulation and signaling, and their malfunction gets linked to severe pathologies [1]–[3]. Understanding the functional roles of IDPs requires studying their interactions with other proteins, which is very challenging and needs a tight coupling of experimental and computational methods. In contrast to structured/globular proteins, IDPs cannot be represented by a single conformation, and their models must be based on ensembles of conformations representing a distribution of states that the protein adopts in solution. The investigation of IDP interaction with structured/globular proteins is indispensable for understanding many biological mechanisms [4]. In terms of applications, understanding such molecular interactions is essential for drug design in pharmacology or protein engineering in biotechnology.

IDPs do not have distinct, well-defined secondary and tertiary structures because of their remarkable backbone flexibility [5]. During the binding of an IDP to a macromolecule (usually another protein), large interfaces get involved, resulting in specific but comparatively weak interactions. IDPs found in bacteria to higher living organisms have many occurrences in eukaryote groups. They are prevalent in various human diseases and enriched in cardiovascular disease, diabetes, cancer, and neuro-degenerative disease-related proteins [6]. Pathogens like *Plasmodium Falciparum* (PF) are among the list of bacteria that inflict damage to the human immune system and are responsible for most malaria-related deaths [7]. We consider the human cardiovascular, immune system, and PF proteins to study and analyze the binding behavior of IDPs in structure-based molecular interactions.

Based on the interaction of bio-molecules with each other, Molecular Dynamics (MD) simulations assist in measuring how every atom in a bio-molecule transforms over time, showing differences or similarities between conformations. The most widely used distance measure is the root-mean-squared-deviation (RMSD). Given two bio-molecular conformations represented as vectors of the Cartesian coordinates of their atoms, an RMSD is the square root of the average squared distance between the corresponding atoms. During protein-protein interactions (PPI), the IDP transforms into different possible conformations to bind with globular protein underlining the deviation in its structure at the time of binding. Research has shown that RMSD plots quantify the structural inspection of experimental binding conformation and help capture the structural change that initial conformation has undergone [8], [9]. However, insufficient data on an IDP can lead to under-determined structural possibilities with large RMSD values. Hence, developing a strategy to understand the structural properties of IDPs at the time of binding can provide new ways for biological probes to examine the functions of IDPs.

The intrinsic disorder poses a challenge for both experimental analyses of the conformation and computational modeling due to the lack of stable structure. Several structural analysis approaches have emerged as a new tool to capture molecular surface patterns between proteins from a given database [10]–[12]. Work in [13], [14] showed the application of a 3D Zernike descriptor that captures the 121 scalar invariant features defining the protein structure. Using this information, one can determine the changes in the shape of the bio-molecule if the position of atoms or residues changes. We apply this tool to validate the molecular structure homology between model conformation and the predicted experimental conformation of the IDPs.

Contribution: We propose an approach that extracts the

topological and geometric properties of the globular protein surface to predict the possible IDP conformation ensembles around it. To the best of the authors' knowledge, this is the first work that uses geometric features of protein surfaces to sample IDP conformations in the conformation space. Our algorithm computes RMSD values of the predicted conformations to analyze the transformation in the structure of IDP from its initial molecular pose. Using the geometric information, it then identifies the suitable binding pose on the globular protein surface to plan a feasible trajectory from the start conformation to the binding conformation using our incremental path planner [15]. Figure 1 shows an overview of our workflow.

We perform experiments for six globular proteins interacting with five IDP molecules in the conformation space. We consider the tertiary structure of the globular proteins, ranging between 173-1544 residues, as a static body and the rigid global body of IDPs as a moving object. Our results show structural homology between the predicted geometrically favorable binding pose and the initial conformation of IDP and report good binding affinity for the experimental conformation.

II. Related Work

A. Protein-protein interactions

An important area of study includes understanding how a protein binds to another protein's active site and what conformational changes both molecules undergo during docking to the active site or its exit from it. Such information allows for predicting the possibility of an association between protein-protein pairs, the strength of this association, and the protein activity level. In protein-protein interactions, geometric and topological features of the protein manifold play important roles. Especially for finding a suitable interaction site, [16], where the bio-molecules bind to the protein regions with potential coherence of matching (concave) curvatures. Several works in [17]–[19] studied and investigated the biological significance of targeting PPI for chemical biology and therapeutic intervention.

Recently, the topology of protein bio-molecule has shown to be surprisingly effective in simplifying biomolecular structural complexity attracting attention to a better understanding of biomolecular behavior during protein-protein interactions. Work in [20] proposed a set of topological methods to examine possible biases introduced in protein-protein interaction network data. Menglun et al. in [21] presented a topology-based network tree to predict PPI using convolutional neural networks (CNN). They characterized PPIs using an element- and site-specific persistent homology. Similarly, the authors in [22] introduced an ensemble learning approach for PPI prediction that integrated multiple learning algorithms and different protein-pair representations. Unlike the discussed strategies, we utilize the topological information of the protein surface to extract the geometric features that help predict the IDP conformation ensembles.

B. Studied biological mechanisms of IDPs

Studying the conformation of highly dynamic IDPs is a challenge in structural biology [23]. Nuclear Magnetic Resonance (NMR), often used in the study of IDPs [24], is a versatile spectroscopy method for studying proteins that, importantly, do not require crystallization. However, NMR spectral data from IDP ensembles have provided conformational constraints. The NMR-constrained molecular dynamics (MD) [25] simulations need multiple copies of the protein (known as replicate exchange MD) to generate possible structural models which fail to ensure the validity of the result regardless of the method used to sample the conformations using NMR data. Work in [26] used NMR to characterize the structure and dynamics of IDPs in various functional states and environments. It describes the NMR parameters of the structural ensemble to quantify the conformational propensities of IDPs and discusses the challenges associated with obtaining structural models of dynamic protein-protein complexes involving IDPs. Researchers have used the combination of molecular dynamics simulations and circuit topology (CT) to analyze the biological behavior of a human androgen receptor with a large N-terminal domain (AR-NTD) [8]. The method constructed the circuit topology of a potentially charged bio-molecule and analyzed the fluctuations of the chain using the root-mean-square-fluctuations (RMSF) and RMSD metrics. With a similar idea in this work, we use the surface topology to extract the geometric features and analyze the structural arrangements of non-charged bio-molecule using RMSD.

C. Sampling Based Motion Planners (SBMP)

A particular domain of molecular modeling relates to the prediction of the bound structure of protein-protein complexes; this problem is usually addressed with computational methods. The method is required to accurately predict the 3D conformation of the bio-molecule upon binding to the target receptor. A new research area has tried applying robotics-based motion planning techniques to this problem [27]–[30], where it randomly samples alternative conformations, in consideration to the position and orientation of the bio-molecule inside the receptor's binding cleft and plans a feasible path to the binding conformation. The space under which the *degrees of freedom* (i.e., the number of parameters, like residues or C-α atoms, needed to describe the pose) of a bio-molecule explored is called conformation space and the regions free of all internal and external constraints are called $\mathcal{C}_{free}$ space in the conformation space.

In this work, we apply our algorithms [15], [31], [32] to perform a random exploration of the IDP's rotational and translational *degrees of freedom*, without exploring its conformational flexibility, i.e., rigid docking. The approach utilizes the topological and geometric properties of the protein surface to examine the geometrically suitable structure arrangement of an IDP around a protein receptor and plans a feasible path to the predicted binding pose.

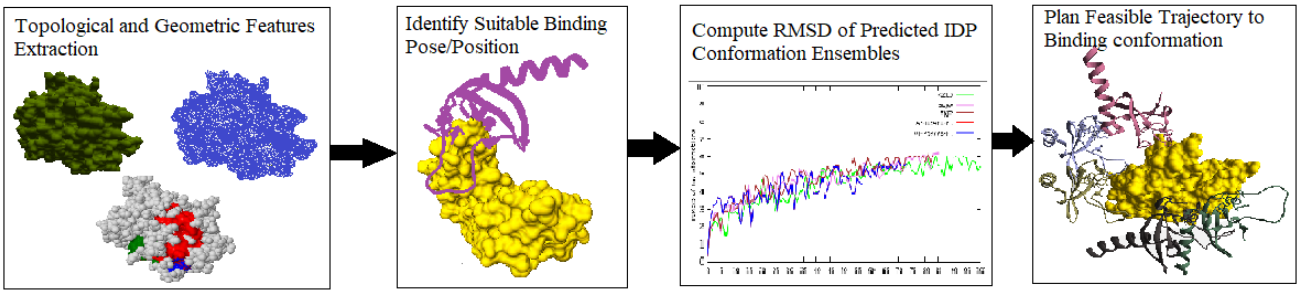


Fig. 1: Work flow of our approach.

III. Methodology

A. Background Definitions

We discuss some of the mathematical concepts used in our algorithm to extract the topological and geometric features of the protein surface.

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 IDP 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 S with diameter that is at most ε.

In this work, the protein surface is modeled as a static object. S is the set of all IDP conformations 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 IDP bio-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 be 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. We formally define f at any point in conformation space by

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

Please refer to [32] for our expanded definitions 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 IDP 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.

We perform steps from [31] to generate a simplicial complex $R(S)$ that captures the topological structure of the protein surface, i.e., vertices, edges, and triangles. We apply the discrete Morse function from [32] 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.

B. Root-Mean-Square Deviation (RMSD) metric

Root-Mean-Square deviation (RMSD) is the average distance measured between the two superimposed protein structures. This method measures the structural similarity/deviation of one protein with another or between two conformations of the same protein. The easiest way to compute RMSD is using the dihedral angle metric of the protein bio-molecule. Although the gradient of the dihedral RMSD is easy to compute, Cartesian RMSD gives a better measure of structural difference. The reason is that the dihedral RMSD does not capture the effect that perturbations of middle dihedral angles in a chain structure entail much larger structural changes than those of terminal angles. Hence, we use Cartesian RMSD to measure the structural deviation in this work. Given a N atoms protein bio-molecule, the RMSD between two conformations x and y is

$$RMSD(x, y) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - y_i\|^2, \tag{2}$$

where y is the predicted conformation and x is the model conformation of the same protein. Each conformation consists of both translation and rotation coordinates of the bio-molecule.

C. IDP conformation transformation during PPI

Algorithm 1 constructs a simplicial complex around the protein surface by sampling and connecting IDP conformations in method *ConstructComplex*. On satisfying the sampling condition from [31], 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 from [32] to this simplicial complex to identify the local maxima (protrusions) and minima (cavity) curvatures of the protein surface in line 4. The identified critical points are the highest and the lowest peak points on the surface at which function f reaches its extremum.

Algorithm 1 Path planning to suitable binding pose

Input: P: Protein surface model, R: A planned pathway to the binding site, s: initial IDP conformation, H: set of closest IDP conformations around the protein surface, g: suitable binding pose.

1: Let $R \leftarrow \{\phi\}$.
2: $S \leftarrow ConstructComplex(P)$; ◁ Refer Def. 2
3: $TopologicalCollapse(S)$; ◁ Refer [31]
4: $C \leftarrow IdentifyCriticalPoints(S)$; ◁ Refer Def. 5, 6
5: $F \leftarrow GetFeasiblePoints(S, C)$; ◁ Refer Def. 7
6: **for all** x ∈ F **do**
7: Compute RMSD(s, x)
8: **for all** $c \in C$ **do**
9: **if** x closest to c **then**
10: H[x] = distance(x,c)
11: **end if**
12: **end for**
13: **end for**
14: $g = \forall_{x \in H} min(H)$
15: $R = PlanPath(s, g)$ ◁ Refer [15]
16: **return** $\{S \bigcap F, R\}$

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 in close proximity to the protein surface and are part of a simplicial complex $R(S)$, refer to Def.7. The method computes RMSD from Eq.2 for the predicted IDP conformations of extracted geometric information map, line 7. Of the predicted conformations, a geometrically favorable binding position of the IDP gets selected with the conformation closest to protein surface curvature in lines 9-14. Finally, a path is planned for the IDP from the start conformation to the binding pose conformation using our method from [15] on taking the predicted IDP conformations as waypoints, in line 15. As a result, our algorithm outputs an extracted geometric information map consisting of critical points, feasible critical points (predicted IDP conformations), and a pathway from the start conformation to the binding pose conformation.

IV. Experimental Data

We obtain protein data from the protein data bank (PDB) [33], [34] and construct their tertiary structure using CHIMERA [35]. We obtain IDP data from PDB and AlphaFold Protein Structure Database (AlphaFold DB) [36]. We consider six proteins and five IDP bio-molecules to study and understand the biological binding mechanism of IDPs using protein surface geometries. The high-dimensional surface models of proteins represent a stationary rigid body in the conformation space. Figure 2 shows the graphical representation of 7A7H protein, its high-dimensional surface model, and the IDP conformation ensembles around it.

The proteins selected include three *Plasmodium Falciparum (PF)* pathogen proteins, i.e., 1SQ6, 1TQX, and 3NTJ, and two human cardiovascular proteins, i.e., 7A7H and 4JKQ, and one human immune protein (2FCB). *PF* is responsible for most malaria-related deaths and forms part of our ongoing research into identifying feasible protein drug targets. The high mutational capacity, coupled with the changing metabolism of the pathogen, makes the development of malaria drug treatments an evolving problem. In this work, we are interested in studying and analyzing the behavior of PF pathogens in the PPI network. Hence, these proteins were selected as they are the potential targets for malaria inflicts.

We selected 4ZLX (106 residues), 5EJW (91 residues), and 7KPI (142 residues) proteins as IDP based on their high disorder behavior shown in the protein feature view plot available on the PDB database. The other 2 IDP bio-molecules, AF-I1E4Y1-F1 (117 residues) and AF-P59773-F1 (190 residues), from AlphaFold DB, are of mus-musculus and homo sapiens species, respectively. The mean per-residue confidence score (pLDDT) for AF-I1E4Y1-F1 is 48 and for AF-P59773-F1 is 59. The pLDDT measure estimates whether the predicted residue has similar distances to neighboring C-α atoms (within 15 $\mathring{A}$) in agreement with distances in the true structure and is scored between 0 and 100. The score assesses the local model quality of the structure, i.e., a lower score refers to the existence of larger disordered regions in a bio-molecule. Figure 3 shows a random combination of IDPs interacting with studied protein surface models.

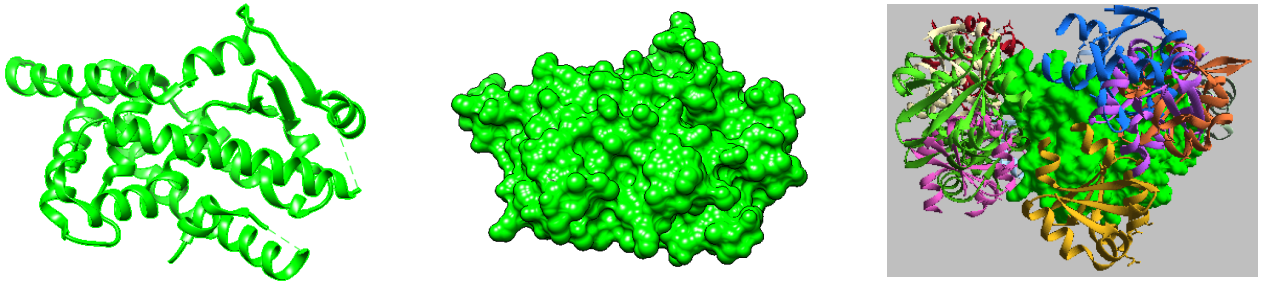
V. Result Analysis

We performed experiments on a Dell Optiplex 7040 desktop machine running OpenSUSE operating system and developed algorithms in C++ language. The results were evaluated for all IDPs with each globular protein for geometric feature extraction, RMSD computation, path planning to binding position, and binding affinity measure, i.e., a total of 150 trials. In this work, we do not change the atoms/molecules or dihedral angles of the IDPs but take the 3D structure of the global rigid body frames to calculate Cartesian RMSD in the conformation space, as discussed next.

A. Computing structural binding transformations

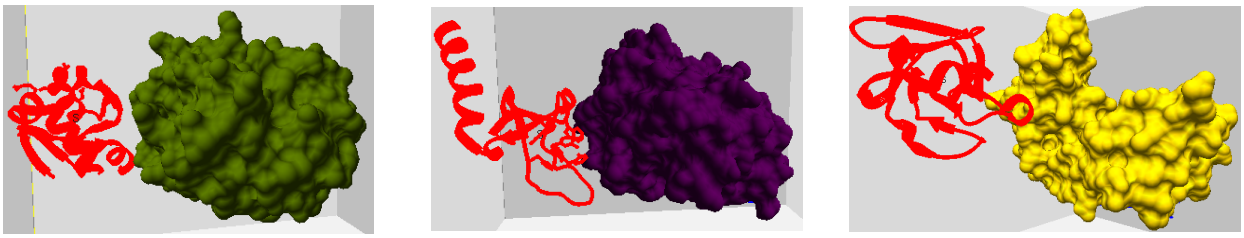
We evaluate the predicted conformation ensembles of IDPs using the RMSD formula from Eq.2 and report the global rigid body transformation (rotation-translation) measured at each IDP conformation in Angstrom (Å), as shown in Figure 4. We align the IDP conformation structures to their original start conformation structure to calculate the translation and rotational deviation in the IDP model that help minimize the Cartesian RMSD values. Our method predicts varying results of the conformation ensembles for IDPs in the conformation space of six proteins. Since the structure of these IDPs is large, the global rigid body frames of structural re-arrangements score between 1 and 10.

We observed that the IDP conformation transforms into different ensembles (i.e., translation and rotation) as it navigates

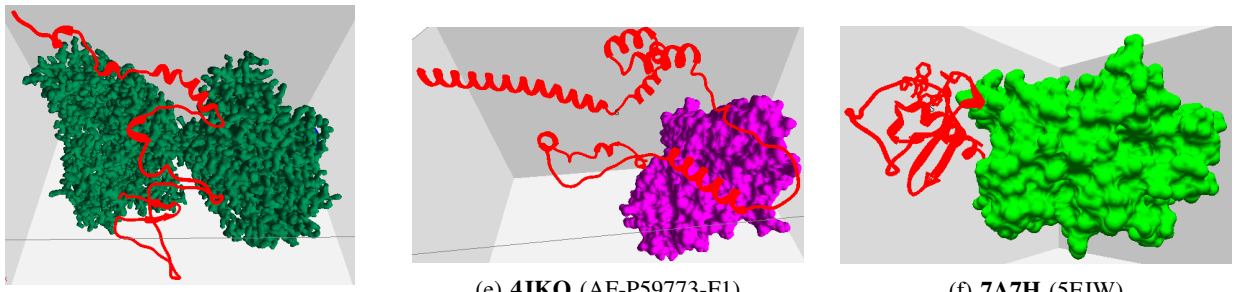


(a) **7A7H protein** (b) **Surface model** (c) **IDP conformation ensembles**

Fig. 2: The figure shows the multiscale surface model of the 7A7H protein and the predicted IDP conformations around detected geometric features (critical points) of the protein surface. The geometric information map provides the view of the 4ZLX bio-molecule conformation around the surface model.



(a) **1SQ6** (4ZLX) (b) **1TQX** (5EJW) (c) **2FCB** (7KPI)



(d) **3NTJ** (AF-I1E4Y1-F1) (e) **4JKQ** (AF-P59773-F1) (f) **7A7H** (5EJW)

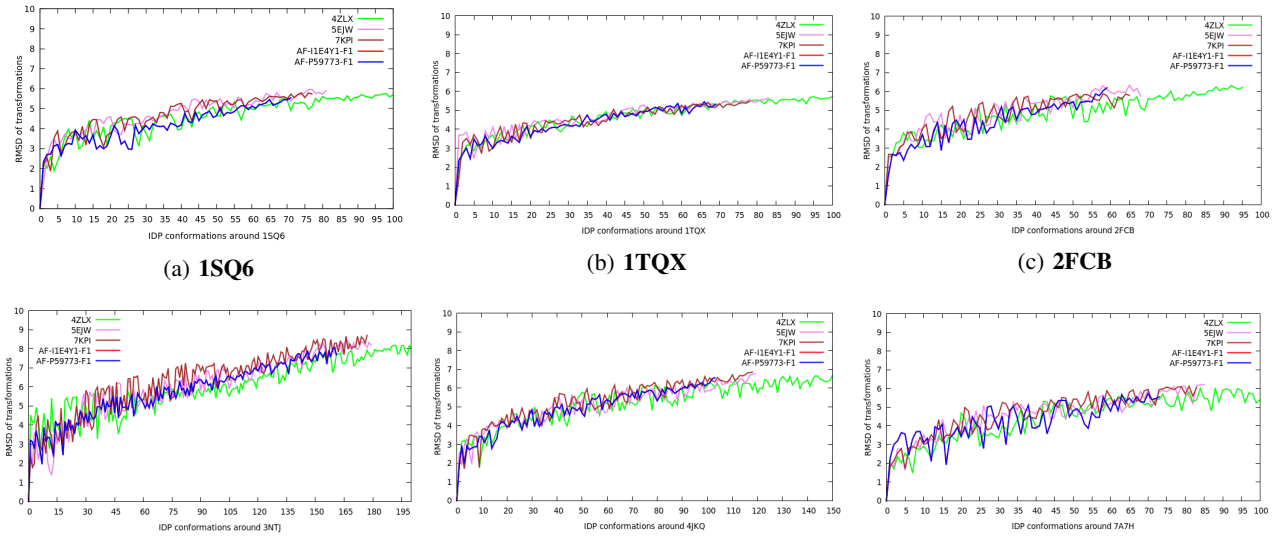
Fig. 3: A random combination of a globular protein surface model and an IDP was captured, from the experimental analysis, with IDP names mentioned in the brackets and IDPs shown in red.

to a geometrically favorable interaction pose in all six proteins. Hence, we analyzed that using the geometric information of protein surface, it is still possible to predict diverse structural arrangements of IDPs around the proteins to find the closest interacting binding pose between two bio-molecules.

B. Planning path to suitable binding pose

We study the total time taken (in seconds) for all IDPs in all six globular protein conformation spaces for geometric feature extraction and path planning to binding pose, as shown in Figure 5. Geometric feature extraction time evaluates the time taken to extract features and predict feasible conformations around the globular protein surface. The path planning time inspects the total time taken for an IDP to move feasibly from start conformation to binding conformation in the vicinity of the protein surface. The structure and size of globular proteins like 1SQ6, 1TQX, 2FCB, and 7A7H are smaller and more compact. So, planning a path around these proteins takes much lesser time than 3NTJ and 4JKQ proteins. However, the disordered regions of the IDPs affect the feature extraction time required to find the geometric structural alignments around globular proteins. The uncertain behavior of IDPs around the studied proteins helps us analyze the feasibility of their interaction with a particular protein, that is, how easily they align around a protein structure for an association. The path planning time helps us analyze the locomotion of the IDP around the protein to find the most suitable binding pose for their rigid body structure. We provide this information to help understand the time required to obtain the feasible binding pose and the surface information of the protein as relevant to future biological studies.

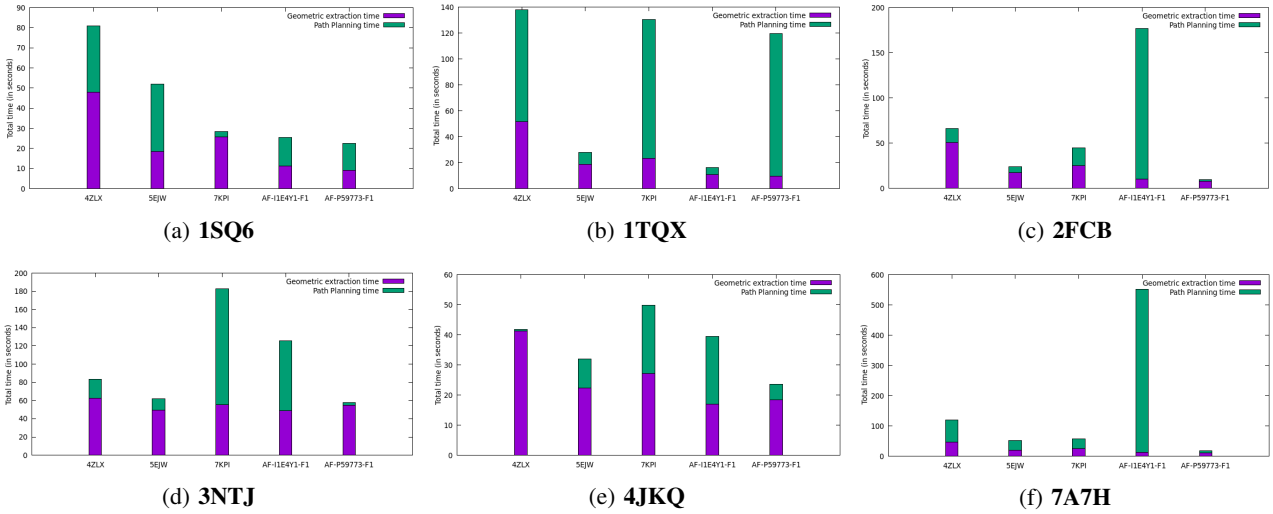
Figure 6 shows screenshots of the planned path for the 5EJW IDP around the 7A7H protein surface to the binding pose conformation. The different view angles reflect the mo-



(a) **1SQ6** (b) **1TQX** (c) **2FCB**

(d) **3NTJ** (e) **4JKQ** (f) **7A7H**

Fig. 4: The plots show measured RMSD for all predicted IDP conformation ensembles around the protein surface model.



(a) **1SQ6** (b) **1TQX** (c) **2FCB**

(d) **3NTJ** (e) **4JKQ** (f) **7A7H**

Fig. 5: The total time taken (in seconds) to extract geometric features and plan a path for all IDPs in each protein's conformation space.

tion of the IDP bio-molecule around the protein surface using the predicted IDP conformations generated by our method.

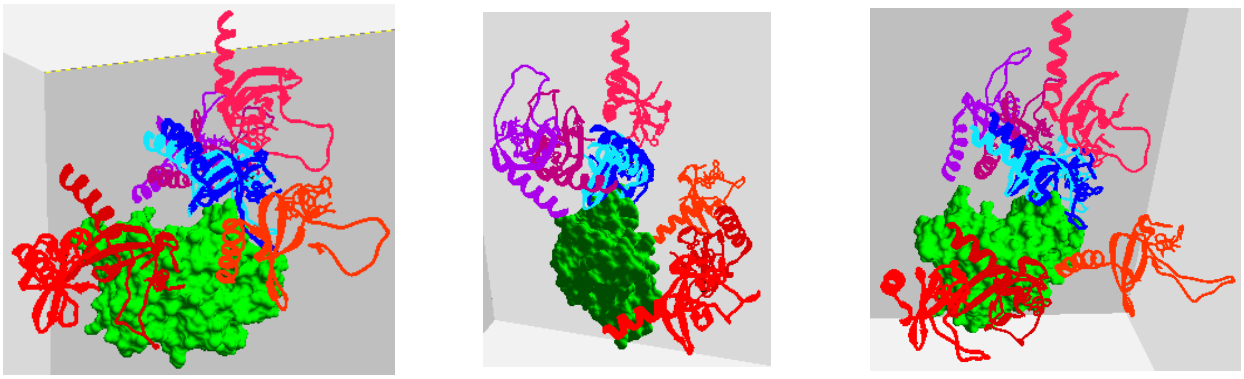
C. Structural analysis of binding conformation

We analyze the structural similarity of our binding goal conformation structure with the model structure of IDP at the initial conformation to understand the changes or movement of the atoms that happen internally within a bio-molecule or externally for a body model. We first examine the external transformation of the identified IDP binding conformations for each protein, aligned along the initial conformation structure, using the RMSD value, as shown in Table I. It helps us analyze the displacement or deviation of the IDP body model from its initial form. We observe that the binding pose conformation showed 60-90 % transformation in the structure from its initial model conformation, i.e., the translation and rotation displacement of the body model from its initial pose, for all IDPs in all proteins' conformation space.

TABLE I: RMSD values computed for the predicted binding pose structure.

	1SQ6	**1TQX**	**2FCB**	**3NTJ**	**4JKQ**	**7A7H**
4ZLX	6.8	6.8	7	9	7.7	6.9
5EJW	6.7	6.5	6.7	8.8	7.4	6.6
7KPI	6.8	6.5	6.9	9.2	7.6	6.8
AF-I1E4Y1-F1	5.8	5.9	6.3	8.3	7	6.2
AF-P59773-F1	6.2	6.4	6.6	8.7	7.3	6.5

We also validate the structural resemblance of the binding pose conformation with the IDP model conformation structure using 3D-SURFER [12] to compare and analyze the difference



(a) **Front view** (b) **Side view** (c) **Top view**

Fig. 6: The path planned for 5EJW IDP around the 7A7H protein surface model using the geometrically favorable conformation ensembles. The start conformation is in red, and the binding goal position is in dark blue.

in protein structures internally. We input the PDB files of both the model structure and the predicted geometrically favorable binding pose structure to the 3D-SURFER tool. The tool provides the difference in protein structures using Zernike feature vectors. It helps us evaluate the displacement of atoms/molecules or residues of the protein structure and the dihedral angles between them. The computation of structural resemblance between the two protein models gets calculated as

$$P_{sim} = \frac{\sum_1^{121} Z_{binding}}{\sum_1^{121} Z_{initial}}, \tag{3}$$

where P_{sim} denotes the percentage of resemblance in structure, $Z_{binding}$ denotes the Zernike invariant (121 scalar values) of the predicted binding structure, and $Z_{initial}$ refers to the Zernike invariant of initial model structure. We observed that our predicted IDP binding goal conformations have a 100% structure resemblance with the initial model structure for all IDPs in each globular protein's conformation space. Thus proving no changes in the positions of atoms or residues within the IDP structure, i.e., our method preserves the internal backbone structures of the IDPs.

D. Binding Affinity of experimental conformation

We inspect the binding affinity of our IDP binding pose conformation with the known binding affinity for a ligand for all six proteins. We use the molar Gibbs free energy ΔG (binding affinity) to determine the relevance of 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) [37]. The protein 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 II shows the binding affinity of the predicted IDP binding pose for each protein compared to the protein's known binding affinity for a ligand.

We observe that our predicted IDP conformations show a higher binding affinity for all proteins than their known binding affinity for a ligand. We see that the geometrically feasible binding pose with good binding affinity becomes useful for interesting biological studies/findings to help investigate the interaction mechanism of IDPs with unknown or known protein bio-molecules. We conclude that our approach successfully captures the geometric features of the protein surfaces and plans a path for IDP biomolecule to the geometrically favorable binding pose showing a higher affinity than ligands. Thus, showing the significance of our approach for further biological studies.

VI. Discussion and Future Work

The paper presented a framework that uses the topological and geometric information of the structured protein surface to examine the possible IDP binding behavior. The work analyzed the deviation in the experimental conformations of IDP with its model structure conformation using the RMSD metric and reported the computation time taken to extract the surface features and plan a transition path. Our experiments show that our method predicts a geometrically favorable binding position for the IDPs around the protein surface models for rigid docking, showing good binding affinities. We also validate the structural homology of binding conformation with the IDP's initial conformation showing a 100% similarity. This work serves as the first step toward the further analysis of IDPs and their interaction with other bio-molecules, given the geometric and topological representation of these bio-molecules. We plan to use a dynamically flexible model of IDPs to study its interaction with other proteins in our future work.

References

[1] V. Csizmok, A. V. Follis, R. W. Kriwacki, and J. D. Forman-Kay, "Dynamic protein interaction networks and new structural paradigms in signaling," *Chemical reviews*, vol. 116, no. 11, pp. 6424–6462, 2016.

[2] M. M. Babu, R. van der Lee, N. S. de Groot, and J. Gsponer, "Intrinsically disordered proteins: regulation and disease," *Current opinion in structural biology*, vol. 21, no. 3, pp. 432–440, 2011.

[3] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Intrinsically disordered proteins in human diseases: introducing the d2 concept," *Annual review of biophysics*, vol. 37, no. 1, pp. 215–246, 2008.

TABLE II: Binding Affinity of the experimental binding conformations

Proteins	$\Delta G_{known\ for\ a\ ligand}$	$\Delta G_{Experimental\ Conformation}$				
		4ZLX	**5EJW**	**7KPI**	**AF-I1E4Y1-F1**	**AF-P59773-F1**
1SQ6	-5.4	-7.21	-8.26	-6.95	-9.72	-9.55
1TQX	-5.4	-10.03	-11.40	-11.37	-11.98	-12.97
2FCB	-5.5	-6.79	-8.39	-7.32	-10.83	-9.71
3NTJ	-8.7	-24.85	-26.47	-27.07	-25.52	-28.68
4JKQ	-6.5	-6.41	-8.87	-8.24	-9.77	-10.03
7A7H	-5.7	-7.56	-8.03	-6.60	-9.94	-9.82

[4] P. Tompa, E. Schad, A. Tantos, and L. Kalmar, "Intrinsically disordered proteins: emerging interaction specialists," *Current opinion in structural biology*, vol. 35, pp. 49–59, 2015.

[5] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nature reviews Molecular cell biology*, vol. 16, no. 1, pp. 18–29, 2015.

[6] P. Kulkarni and V. N. Uversky, "Intrinsically disordered proteins in chronic diseases," p. 147, 2019.

[7] P. A. Holding and R. W. Snow, "Impact of plasmodium falciparum malaria on performance and learning: review of the evidence," *The Intolerable Burden of Malaria: A New Look at the Numbers: Supplement to Volume 64 (1) of the American Journal of Tropical Medicine and Hygiene*, 2001.

[8] V. Sheikhhassani, B. Scalvini, J. Ng, L. W. Heling, Y. Ayache, T. M. Evers, E. Estébanez-Perpiñá, I. J. McEwan, and A. Mashaghi, "Topological dynamics of an intrinsically disordered n-terminal domain of the human androgen receptor," *Protein Science*, vol. 31, no. 6, p. e4334, 2022.

[9] F. Ballante and G. R. Marshall, "An automated strategy for binding-pose selection and docking assessment in structure-based drug design," *Journal of Chemical Information and Modeling*, vol. 56, no. 1, pp. 54–72, 2016.

[10] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped blast and psi-blast: a new generation of protein database search programs," *Nucleic acids research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[11] 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.

[12] D. La, J. Esquivel-Rodríguez, V. Venkatraman, B. Li, L. Sael, S. Ueng, S. Ahrendt, and D. Kihara, "3d-surfer: software for high-throughput protein surface comparison and analysis," *Bioinformatics*, vol. 25, no. 21, pp. 2843–2844, 2009.

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

[14] 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.

[15] A. Upadhyay, B. Goldfarb, and C. Ekenna, "Incremental path planning algorithm via topological mapping with metric gluing," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[16] H. Lu, Q. Zhou, J. He, Z. Jiang, C. Peng, R. Tong, and J. Shi, "Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials," *Signal transduction and targeted therapy*, vol. 5, no. 1, pp. 1–23, 2020.

[17] C. Maniaci and A. Ciulli, "Bifunctional chemical probes inducing protein–protein interactions," *Current Opinion in Chemical Biology*, vol. 52, pp. 145–156, 2019.

[18] P. Bryant, G. Pozzati, and A. Elofsson, "Improved prediction of protein-protein interactions using alphafold2," *Nature communications*, vol. 13, no. 1, pp. 1–11, 2022.

[19] X. Meng, J. Xiang, R. Zheng, F.-X. Wu, and M. Li, "Dpcmne: detecting protein complexes from protein-protein interaction networks via multi-level network embedding," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 3, pp. 1592–1602, 2021.

[20] A. W. Nowakowska and M. Kotulska, "Topological analysis as a tool for detection of abnormalities in protein-protein interaction data," *Bioinformatics*, 2022.

[21] M. Wang, Z. Cang, and G.-W. Wei, "A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 116–123, 2020.

[22] K.-H. Chen, T.-F. Wang, and Y.-J. Hu, "Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–17, 2019.

[23] K. Pauwels, P. Lebrun, and P. Tompa, "To be disordered or not to be disordered: is that still a question for proteins in the cell?" *Cellular and Molecular Life Sciences*, vol. 74, no. 17, pp. 3185–3204, 2017.

[24] M. R. Jensen, R. W. Ruigrok, and M. Blackledge, "Describing intrinsically disordered proteins at atomic resolution by nmr," *Current opinion in structural biology*, vol. 23, no. 3, pp. 426–435, 2013.

[25] J. R. Allison, P. Varnai, C. M. Dobson, and M. Vendruscolo, "Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements," *Journal of the American Chemical Society*, vol. 131, no. 51, pp. 18 314–18 326, 2009.

[26] S. Milles, N. Salvi, M. Blackledge, and M. R. Jensen, "Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments," *Progress in nuclear magnetic resonance spectroscopy*, vol. 109, pp. 79–100, 2018.

[27] 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.

[28] T. Adamson, J. A. Camarena, L. Tapia, and B. Jacobson, "Optimizing low energy pathways in receptor-ligand binding with motion planning," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 2041–2048.

[29] A. Estana, "Algorithms and comptutational tools for the study of intrinsically disordered proteins," Ph.D. dissertation, Toulouse, INSA, 2020.

[30] A. Upadhyay, T. Tran, and C. Ekenna, "A topology approach towards modeling activities and properties on a biomolecular surface." in *BIBM: IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, 2021.

[31] 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.

[32] A. Upadhyay, B. Goldfarb, W. Wang, and C. Ekenna, "A new application of discrete morse theory to optimizing safe motion planning paths," in *15th International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2022.

[33] 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.

[34] 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.

[35] 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.

[36] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon *et al.*, "Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models," *Nucleic acids research*, vol. 50, no. D1, pp. D439–D444, 2022.

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