SHREC 2024: Non-rigid complementary shapes retrieval in protein-protein interactions

Taher Yacoub, Nika Zarubina, Camille Depenveiller, H-P Nguyen, V-T Vong, M-T Tran, Y Kagaya, T Nakamura, D Kihara, Florent Langenfeld, et al.

Eurographics Workshop on 3D Object Retrieval (2024)
S. Biasotti, B. Bustos, T. Schreck, I. Sipiran, and R. C. Veltkamp (Editors)

SHREC 2024: Non-rigid complementary shapes retrieval in protein-protein interactions

T. Yacoub[†1], N. Zarubina[1], C. Depenveiller[1], H.-P. Nguyen[2], V.-T. Vong[2], M.-T. Tran[2], Y. Kagaya[3], T. Nakamura[3], D. Kihara[3,4], F. Langenfeld[†1], M. Montes[†1,5],

[1]Laboratoire GBCM, EA 7528, Conservatoire National des Arts et Métiers, Hésam Université, Paris, France
[2]University of Science, VNU-HCM, Vietnam National University, Vietnam
[3]Department of Biological Sciences, Purdue University, West Lafayette, IN, 47906, USA
[4]Department of Computer Science, Purdue University, West Lafayette, IN, 47906, USA
[5]Institut Universitaire de France (IUF)

Abstract

The aim of this SHREC 2024 track is to compare different algorithms for retrieving non-rigid complementary shape pairs, applied in the context of 3D objects being more complex (e.g. with many folds and roughness) such as proteins. The dataset used for this benchmark is based on 52 selected protein-protein complexes for which an experimental structure is publicly available. One of the main difficulties of this challenge is the non-inclusion of the shapes derived from the ground truth conformations in the dataset. Different metrics were used to evaluate the retrieval performance (nearest-neighbor, first-tier, second-tier, and true positives) and to evaluate the quality of the predicted poses (TM-score, lDDT, ICS, IPS and DockQ — those metrics are classically used in the Critical Assessment of PRediction of Interactions challenges). Two teams took part in this challenge and were able to return the expected results. This paper discusses these results and prospects of retrieval methods based only on the protein shape information in the absence of atomic data, in a large context of protein-protein docking.

CCS Concepts
• ***Applied computing*** → ***Molecular structural biology;*** *Bioinformatics;* • ***Computing methodologies*** → ***Shape analysis;***

1. Introduction

Proteins interactions play a vital role within cells. Proteins are usually represented as a graph where each vertex is the center of an atom and each edge is a chemical bond between two atoms; but they can also be represented by their molecular surfaces (*i.e.* by their shapes). In the past, this representation was used to derive SHREC tracks focusing on protein classification [RFB*21; LAC*21] or protein-ligand binding site recognition [GRF*22]. In the present work, we focus on the protein-protein interactions (PPIs), that usually involves shape complementarity (Fig. 1, bottom panel). The characterization of protein interactions can be separated into two questions: i) the identification of the two interacting partners that form a complex, and ii) the prediction of the optimal molecular complex geometry.

While (ii) is a classic molecular docking problem that can be tackled by shape complementarity methods [ZSO09]; (i) can be viewed as a 3D object retrieval challenge that is classically explored among the SHREC (SHape REtrieval Challenges) community benchmarks [DGA*09; PSA*16]. They consist in evaluating the ability of computer vision methods to identify a list of objects from partial query information. The difficulty can be worsened by the intra-class variability, which consists of a class of different deformations of the same object due to a non-rigid behavior. Different fields of interest are concerned by this issue, like robotics for instance. In our context, the challenge is to retrieve two proteins considered as compatible to form a real-life complex, and where each of them taken separately constitutes partial information. As for the variability, it corresponds to the flexibility of a chain due to the internal degree of freedom of rotational bonds.

In this SHRE 2024 track, we address the 3D object retrieval challenge applied to protein-protein complexes with two assessments: i) the ability to retrieve the two real chains of each of the 52 experimentally-resolved complexes in the benchmark (called "ground truth"), and ii) the ability to find the real area of interaction between them. To make this challenge more difficult, the dataset used to retrieve the complexes includes different conformations associated with each chain, and excluding the chains from the ground

† Corresponding authors: T. Yacoub (taher.yacoub@lecnam.net), F. Langenfeld (florent.langenfeld@lecnam.net) & M. Montes (matthieu.montes@cnam.fr).

DOI: 10.2312/3dor.20241173

truth. Thus, we can evaluate the tolerance of applied algorithms to the non-rigid deformations.

Due to a constraint of timeline to adapt their algorithm for this challenge and despite the interest of several groups, only two participants were able to return the results in the given timeline. They downloaded the query and target datasets, and submitted their results including the prediction of query-target pairs, and their ranking based on a score indicating the quality of the prediction.

The paper is organized as follows: Section 2 describes the dataset and methods used to evaluate the retrieval performances, Section 3 describes the major obstacles of this challenge and participating methods, and Sections 4 and 5 provide a discussion and conclusion respectively.

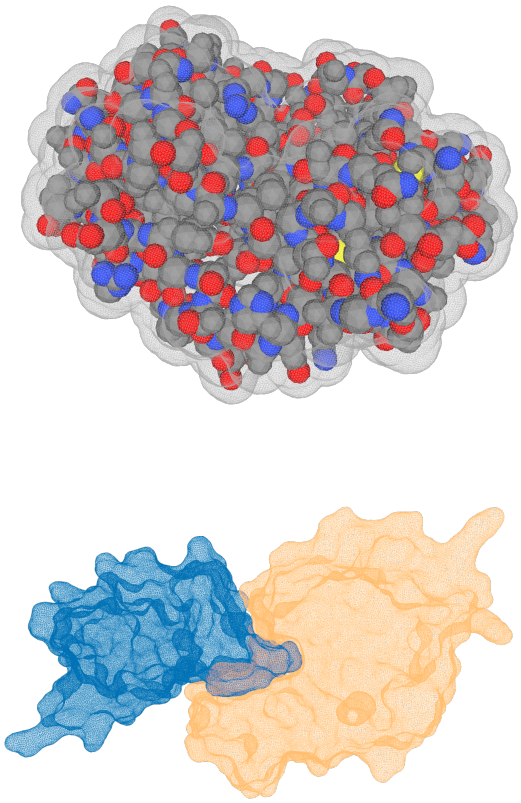


Figure 1: *Top panel: Example of query #1 (chain B of PDB 3n4i – Beta-lactamase inhibitory protein) represented in spheres (each sphere represents an atom from the experimentally resolved protein structure) and its corresponding SES mesh as provided to the participants. Bottom panel: The shape complementarity of the New antigen Receptor PBLA8 variable domain SES (blue mesh) and the lysozyme C SES (orange mesh). This example shows one of the 52 ground truth complexes used to generate the dataset (PDB structure: 2I25).*

2. Benchmark

2.1. Dataset & Ground Truth

All structural data are derived from the Protein Data Bank [BHN03], the Protein-Protein Docking Benchmark 5.5 [VMV*15; GVZ*21] and the PDBFlex database [ZSG07; HLS*15; ILJ*20]. The Protein-Protein Docking Benchmark 5.5 lists protein structures describing multi-chain complexes and their non-bonded counterparts. The PDBFlex database clusters PDB structures of individual chains from the same protein, highlighting the structural variations and the intrinsic flexibility of each protein.

Given a complex resolved in a single structure, the Protein-Protein Docking Benchmark 5.5 lists the non-bonded counterparts of each individual chain of the complex in a non-bonded state. We only retained complexes made of 2 protein chains to limit the complexity of the docking problem as the track timeline was limited. Thus, for each complex, we had a list of 3 PDB structures: one for a complex made of two chains, and one for each chain in the unbound state. The PDB structure of the complex was used to derive the ground truth; the two other chains were used to derive the queries and the targets of our dataset.

For the queries and the targets, we queried the PDBFlex [ZSG07; HLS*15; ILJ*20] whether the chain was part of a cluster; if so, we retrieved all the members of this cluster; if not, the complex was removed from the dataset. We also removed complexes for which the target cluster was smaller than 10 structures, resulting in a dataset (ground truth) of 52 complexes. For the queries, we randomly selected at most 10 members of the cluster, excluding the chain from the complex PDB structure. For the targets, we randomly selected 10 members, excluding the chain from the complex PDB structure, in order to have a balanced dataset with a consistent number of 10 targets for each query.

It is important to note that the queries and targets datasets do not necessarily contain complementary shapes. We queried the PDBFlex using the *unbound* structure of a protein, but the PDBFlex cluster of a query may contain various conformations of the same protein chain: *unbound*, *bound* to the target, or *bound* to another protein. The structure of a protein may undergo small to large structural rearrangements between each structure, as it is experimentally solved under different conditions (bound/unbound, physico-chemical conditions of the experiments, *etc.*). Therefore, the structures (and therefore the shapes derived from the structures) of a given protein may differ, reflecting the intrinsic flexibility of the protein structure. As an illustrative purpose, two examples of proteins in their bound and unbound states are provided in Fig. 2: while the Elongation factor 2 undergoes a large conformational rearrangement upon binding (Fig. 2A, left), the Complement C3 protein remains largely identical (Fig. 2B, left).

For the ground truth, each chain of the complexes was retrieved from the PDBe [ABC*19], individualized and EDTSurf [XZ09; XLZ13] was used to compute their Solvent-Excluded Surfaces in the PLY format. All the 387 queries and 520 targets were retrieved from the PDBe [ABC*19]. We cleaned these structures using the pdb_tools toolkit [RTTB18]: we only kept the relevant chain, removed the hetero-atoms and the hydrogen atoms. Then, EDTSurf [XZ09; XLZ13] was used to generate the queries and tar-

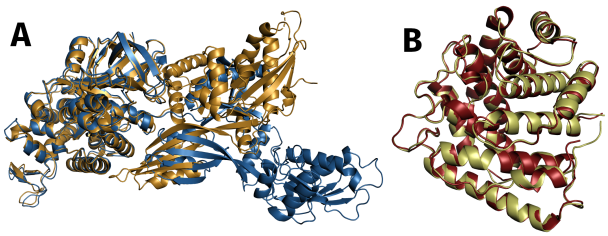


Figure 2: *A (left): Superposition of the unbound Elongation factor 2 protein (blue) and of the Elongation factor 2 protein (orange) bound to the exotoxin A (not shown on the picture, for clarity). B (right): Superposition of the unbound Complement C3 (light yellow) and of the Complement C3 (dark red) bound to the Integrin alpha-M (not shown on the picture, for clarity).*

gets surfaces in the PLY format, using the same parameters as for the ground truth surfaces. These were the shapes provided to the participants (Fig. 1). The dataset is freely available on the track webpage (`https://shrec2024.drugdesign.fr/`).

2.2. Evaluation

2.2.1. Retrieval performances

The retrieval performances of the participants' algorithms were evaluated using the standard nearest-neighbor, first-tier and second-tier retrieval metrics. These metrics compute the fraction of true positives, for each query, within the top-1, top-N and top-$2N$ results, with N being the number of true positives for the query considered. Finally, the number of true positives among the complexes returned by the participants was evaluated.

2.2.2. Alignment and generation of complexes

At first for the evaluation of the participants' results, we selected only the true positives among each query-target pair predicted by the participants. Each original query and target (before docking) in the PLY format have been aligned onto the participants' query-target pair using Open3D's RANSAC algorithm [ZPK18] to get a transformation matrix. Thus, for each query and each target, we obtain one transformation matrix respectively. Each matrix is then applied to the original PDBs (corresponding to a structural representation) with the MMTSB perl-based toolset [FKB04], in order to obtain participants' predictions from PLY format to PDB format, being essential for computing scores and evaluating predictions.

2.2.3. Comparison of query-target poses to reference complexes

To compare the query-target poses submitted by the participants to the reference complexes, several metrics from the structural bioinformatics community are used: TM-score (Template Modeling score) [ZS04], lDDT (local Distance Difference Test) [MBBS13], ICS (Interface Contact Similarity) [LBK*18a], IPS (Interface Patch Similarity) [LBK*18a] and DockQ [BW16]. To calculate ICS, IPS and DockQ, the distance threshold is set by default according to CAPRI definition (*i.e.* 5.0Å) [LBK*18b]. They allow evaluating the quality of predicted complexes compared to reference complex structures. These scores all have been computed using the OpenStructure framework [BSB*13].

In our analysis, we also computed the vectors between the centers of mass of the query and the target proteins both in the predicted poses and the reference structures, after applying a query-based superposition to the predicted poses. By calculating the angle between these vectors, we gained insights into the degree of alignment or divergence in the position of the target relative to the query in both the predicted poses and the reference structures.

2.2.4. Hex: a Shape-based Protein Docking Method

To highlight the difficulty of the challenge, we have applied *Hex*, a protein docking algorithm, based on spherical polar Fourier correlations of protein surface shape representation [RK00]. This algorithm is available for about 20 years, and we used the version 8.1.1 (`https://hex.loria.fr/`).

To run the docking, the following settings were used: the correlation type was set to "Shape Only", the samping method was set to "Range Angles" with Euler rotation angles for receptor, ligand and twist defined respectively to 180°, 180°and 360°with a "Step Size" equal to 7.5, and using 3D FFT (fast Fourier transform). The "Steric Scan" and "Final search" were defined at the order *N=16* and *N=25* respectively, with a "Distance Range" equal to 60. Finally, the "Grid dimension" and the "Box size" were set to 0.6Å and 10Å respectively.

2.3. Participants

Participants were afforded a span of 7 weeks from the dataset's release to submit their results. Multiple teams demonstrated their interest for the track, among which two were able to return the required results in the time limit:

1. Kihara *et al.* use LZerD, a structure-based protein docking approach, augmented by the 3D Zernike descriptor (3DZD) for shape representation 3.2. This approach has demonstrated success in the CAPRI challenge [CTS*19; LBM*21; LBR*23].
2. Tran *et al.* reduce the number of vertices and faces by a factor of 10 while preserving the object details. Each shape is then described by multiple tables, each one representing a different view of the shape, using a Fibonacci Sphere (FS) to generate unit vectors representing different views 3.1. The query and target tables are finally aligned to find complementary shapes.

3. Methods

In this section, we describe the two methods evaluated within the track timeline. Several other groups manifested their interest in the proposed problem, but were not able to produce the results in time. The major obstacles were 1) the constraints in the timeline, as most methods from the computer vision field are not directly applicable to protein surface meshes, and as such, need to be partially modified and tested before use; and 2) the lack of publicly available reference dataset of protein complexes (in the form of 3D meshes and / or cloud points) to train machine-learning algorithms against. Regarding the latest point, this track dataset may be a first step towards the publication of curated datasets for the training of

new, surface-based methods devoted to the prediction of molecular complexes. The participants were provided with two sets of protein shapes (queries and targets), and were asked 1) to evaluate the likelihood of each query-target complex, 2) to provide the 10 most probable predicted complexes.

3.1. FS-proj: Fibonacci-Sphere projected onto a plane

By: Y. Kagaya, T. Nakamura and D. Kihara.

3.1.1. Projection with multiple rotations

The 3D objects are simplified by reducing the number of vertices and faces. This process is done by using the Decimate Modifier of Blender with a ratio of 0.1. In this way, the number of vertices and faces is reduced by 10 times but still maintains the detail of the object. After simplifying, each 3D object is centered and rotated in different directions and angles. This is done by creating a Fibonacci Sphere, produced points on the sphere can be seen as the unit vectors representing different views of the object in different directions. For each unit vector, we create a plane going through the origin and have that unit vector as a normal vector. All the points of the object are translated by the direction of the normal vector so that they all lie within the positive region separated by the plane. All the points are projected from a three-dimensional space onto a two-dimensional plane with different rotations about the normal vector to produce multiple 100×100 height tables. We want to focus on the upper surface of the object, so the 2D projected coordinates are rounded to fit into the table and only the greatest heights in each position are stored in the table. In this stage, suppose we project the points with a directions and b rotations, each 3D protein mesh will produce an $a \times b \times 100 \times 100$ array.

3.1.2. Evaluating method

For each resulting alignment of the two projection tables of a query and a target, we calculate the score of the pair by taking the sum of some desired metrics over the distance between two corresponding cells.

As the resulting alignment is represented as the intersection of the target projection table and the query projection table by some **offset**, associating with a **distance value** between the tables, the distance between two corresponding cells is defined as the difference between associating distance value and the sum of the heights of two corresponding cells.

The applied metric should have the domain of the non-negative real number and should be non-increasing over the domain.

The chosen metric was $1/(1+x)^2$ with the motivation of imitating the gravitational/electric force. We applied the $(1+x)$ part to avoid the divided by zero problems.

3.1.3. Results and runtimes

After evaluation, each query has a list of scores and information for creating the complexes. From the scores, we produce a 387×520 score matrix reflecting the likelihood of each query-target complex and get the top 10 highest-score query-target complexes. By using the information on the direction and angle projection of the query and target protein, we can restore the state of the two meshes and merge them into a protein-protein complex.

In the Projection with multiple rotations stage, we project the object's points with $a = 24$ directions and $b = 16$ rotations, a $24 \times 16 \times 100 \times 100$ array is produced in 30 seconds on average for each 3D mesh after simplification. In the Evaluation stage, because of the limitation of time, we reduce the number of directions and rotations to $a = 12$ directions and $b = 8$ rotations for each 3D mesh, each query-target is evaluated in 5 seconds on average. The source code can be accessed at: `https://github.com/nhphucqt/shrec2024-protein`.

3.2. LZerD: Local 3D Zernike descriptor-based Docking

By: H.-P. Nguyen, V.-T. Vong and M.-T. Tran.

3.2.1. Method description

We participated in this challenge using a method based on LZerD (Local 3D Zernike descriptor-based Docking program) [VYSK09], a surface-based protein docking method developed by our group that has been successful in the Critical Assessment of PRediction of Interactions (CAPRI) challenge for many years [CTS*19; LBM*21; LBR*23]. LZerD exhaustively samples the possible interaction interface areas and interaction angles of two structures. If the interfaces collide, are too small, or the shape complementarity of the interfaces is too low, the model is rejected. LZerD treats protein structures as surface structures, which are first divided into local surface regions. This local surface area is represented by a 3D Zernike descriptor (3DZD) [KSCE11]. Since 3DZD is rotationally invariant, shape complementarity can be calculated quickly and alignment-free.

Since the LZerD program was designed to dock proteins, it required protein atomic information in Protein Data Bank (PDB) file format as input. LZerD uses the atom information to calculate surfaces, the number of clashes of the docked conformation, and the final scores of docked conformations. For this reason, we needed to generate a reasonable PDB file and modify LZerD to disable the components related to atoms. To generate a PDB file, we first placed assumptive atom spheres in a given triangle mesh under the condition that the spheres have contact with the mesh from the inside. This process was intended to roughly recover surface atoms from a given mesh by reversely taking into consideration the process of calculating solvent-excluded surfaces from surface atoms. Atom spheres were placed per triangle of the mesh in the direction of the normal of the triangle with a certain distance from the center of the triangle. The distance was randomly chosen from van der Waals radii of nitrogen, carbon, alpha carbon, and oxygen in a peptide. If an atom sphere attempting to place overlapped with already placed spheres, placing a sphere for that triangle was skipped. We used Open3D Library [ZPK18] for calculating normals and checking whether a point is inside the mesh.

The scoring of LZerD first considers the directions of their surface normals and the correlation between their 3DZD at the interface. If the docking conformation had normals opposing each other and shapes were complementary, the conformation got a higher score. The final score was calculated by taking the weighted sum of

this score, the area of interface (larger is better), and crashed volumes (lower is better). This score was used to rank the conformations inside of one query-target pair as well as rank the targets for each query to report top 10 query-target pairs. The original LZerD had a further scoring method to re-rank top conformations using the ranksum method to combine several scoring functions. However, this was disabled because those scoring functions relied on the atomic structure written in the PDB file.

With those modifications, we ran LZerD with the original default parameters except for the threshold for pre-filtering by 3DZD correlation. Due to the time constraints of this challenge, we raised the threshold from the original 0.7 to 0.9. If the surface's 3DZD correlation was less than this threshold, subsequent calculations were skipped and contributed to reduced calculation time. However, this may lead to a risk of missing correct conformations which have weak surface shape correlation. The LZerD protein docking suite is free available at `https://kiharalab.org/proteindocking/lzerd.php`.

3.2.2. Results and runtimes

The calculations of LZerD were performed on Negishi and Bell, Purdue's community clusters run and maintained by the Rosen Center for Advanced Computing [HYM14]. The nodes of those clusters consist of 2-way AMD EPYC 7763 and 2-way AMD EPYC 7662, respectively, and each of them has 256GB RAM. Docking calculations for one query-target pair by LZerD took 7.75 CPU hours on average. Thus, for all the 387×520 pairs of processes, it took a total of approximately 1.56 million CPU hours. The preprocessing of LZerD such as creating a PDB file and the process of creating PLY files for submission from the rotation matrix obtained were negligible compared to the calculation time of LZerD itself.

4. Results & Discussion

4.1. Experimental Results

4.1.1. Retrieval performance

Metric	FS-proj	LZerD
Nearest-Neighbor	0.0	0.0
First-tier	0.014	0.023
Second-tier	0.035	0.046
TP (TP — %)	49 (1.3%)	99 (2.6%)
$TP_{complex}$ ($TP_{complex}$ — %)	22 (42.3%)	12 (23.1%)

Table 1: *Retrieval performances of the evaluated algorithms. TP stands for the number of true positives returned by the participants. In parenthesis, this number is expressed as a percentage (%) compared to the total number of complexes ($10 \times 387 queries$). $TP_{complex}$ is the number of complexes for which at least one query gave a true positive. In parenthesis, this number is expressed as a percentage (%) compared to the 52 complexes.*

Nearest-Neighbor, First-tier, Second-tier and true positives rates at the query and protein levels are calculated. True positives rate at the query level (TP) corresponds to the percentage of true positives returned by the participants, while true positives at protein level ($TP_{complex}$) corresponds to the percentage of complexes for which at least one query gave a true positive. The retrieval results are shown in Tab 1.

The Nearest-Neighbor, First-tier and Second-tier values tend toward 0, indicating that the fraction of true positives in the top-scores is low. Although the percentage of TP is about 1.3% and 2.6% for FS-proj and LZerD respectively, the percentage of $TP_{complex}$ is about 42.3% and 23.1% respectively.

4.1.2. Comparison of query-target pairs to reference

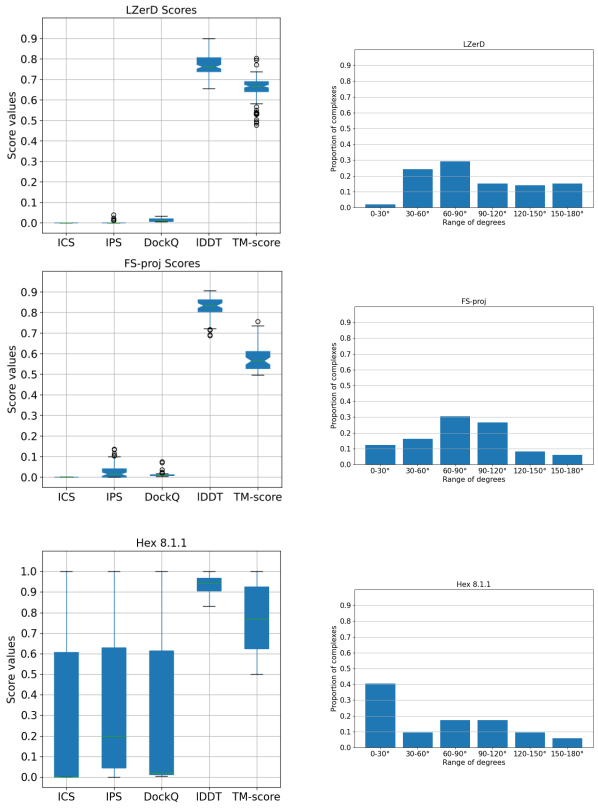


Figure 3: *Left panel: Evaluation of the quality of results in comparison with the ground truth. 5 parameters were assessed such as ICS, IPS DockQ, lDDT and TM-score, for each participant: LZerD (upper) and FS-proj (lower). Scores range between 0 and 1. Overall, ICS, IPS and DockQ show a low score (between 0 and 0.2) whereas TM-score and lDDT show a high score (between 0.5 and 0.9). Right panel: Proportion of complexes as a function of the angle between the center-of-mass vectors. The angle is calculated between two vectors, the reference's vector and the predicted model's vector. These vectors are computed between the center of mass of the query protein and the center of mass of the target protein. The values are calculated for each range of 30° angle.*

In Fig. 3, we present in the left panel the distribution of key scoring metrics including ICS, IPS, DockQ, TM-score and lDDT for

each true positive identified by participating teams in our study. Notably, we observed similar trends in the results obtained by both teams and their respective collaborators. For both methods, ICS, IPS and DockQ scores are below 0.1, indicating a limited agreement in the predicted interaction interfaces when compared to the ground truth. In contrast, lDDT and TM-score exhibit significantly higher values, ranging between 0.5 and 0.9, indicating a closer structural alignment between the predicted and reference complexes. The divergence in performance between these metrics can be attributed to their distinct evaluation criteria. ICS, IPS and DockQ assess the retrieval of the interaction interface between both partners of the complex, while lDDT and TM-score reflect the overall similarity between the ground truth complex and those submitted by the participants. Consequently, instances arise where low ICS, IPS and DockQ scores coexist with good lDDT and TM-score values, suggesting partial alignment of certain regions of the complex, typically observed in the query structures, despite incomplete retrieval of the interaction interface.

Although IPS and ICS are used to compare strictly two interfaces, a third vector-based criterion was used to assess whether the predicted query-target zone of interaction was correct. Unlike ICS and IPS which are based on a comparison of the contacts at the interface, the center-of-mass vector-based criterion evaluates the position of the target relative to the query. Thus, the vector between the query and target centers of mass from each model is compared to the vector between the query and target centers of mass from its corresponding reference. The results are shown in Fig. 3.

For both participants, about 20% of predicted complexes locate the target within the threshold of 60° around the position of the reference target. Thus, even if the participants' methods were unable to accurately predict the reference interface, they show an ability to predict an interaction zone in the vicinity of the actual interface.

4.2. Discussion

As shown in Tab. 1, LZerD method achieves a higher true positive rate, but FS-proj finds at least one true positive for more complexes. More specifically, LZerD founds an average of 8.25 true positives ($99/12$) by $\mathrm{TP}_{\mathrm{complex}}$ (complex for which it found at least one true positive), ranging from 1 and up to 39 true positives by $\mathrm{TP}_{\mathrm{complex}}$, respectively. In contrast, FS-proj founds an average of 2.23 true positives ($49/22$) by $\mathrm{TP}_{\mathrm{complex}}$, from 1 to 5 true positives by $\mathrm{TP}_{\mathrm{complex}}$. 22 $\mathrm{TP}_{\mathrm{complex}}$ were retrieved by the FS-proj, and 12 by the LZerD workflow. The FS-proj produced slightly better complex for the true positives, as evidenced by slightly higher IPS values (cf. Fig. 3) and more predicted poses with an angle inferior to 30° between the vectors of the center of mass of the query protein to the center of mass of the target protein extracted from the ground truth and the predicted poses (cf. Fig. 3, right panel).

LZerD method achieves a higher average recall level, when a true positive was found for the given complex. However, the ICS, IPS or DockQ scores remain low (cf. Fig. 3, upper left panel). This highlights the fact that when the correct query-target pair is found, the relative position and / or orientation of the target relative to the query remains perfectible, as evidenced by the distribution of the angles between the center-of-mass vectors (cf. Fig. 3, upper right panel).

LZerD method was developed to operate on protein structures, and proved to be successful [CTS*19; LBM*21; LBR*23] at predicting correctly protein complexes based on the structures of the individual partners. To apply this structure-based algorithm to the shape-based problem proposed in this SHREC 2024 track, the input meshes had to be converted back to synthetic PDB structures as an initial step, which are subsequently subdivided into local surface areas. It is therefore likely, given the performance gap observed between this track and the CAPRI experiment [CTS*19; LBM*21; LBR*23], that the mesh-to-structure retro-conversion did not produce synthetic structures that are accurate enough to produce synthetic structures from which the LZerD docking algorithm may produce the expected high-quality results, as observed in pure structure-based benchmarks [CTS*19; LBM*21; LBR*23].

One key aspect of the LZerD method is the scoring function used to rank the query-target poses. It is composed of several elements, among which the area of interface (cf. Subsection 3.2): a larger area results in a greater score. For one specific complex, the Importin beta-1 subunit / GTP-binding nuclear protein RAN, the overall shape of the Importin beta-1 subunit is of particular interest. This protein forms an irregular crescent, a concave site that offers a large surface area (cf. Fig. 4 and Fig. 5, middle panel).

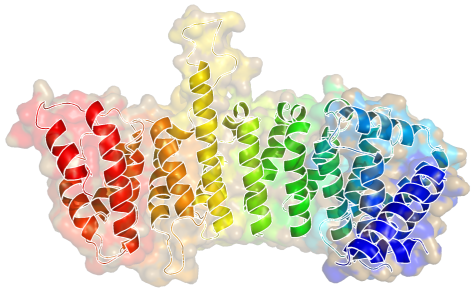


Figure 4: *Importin beta-1 subunit structure. The rainbow shows the different secondary structures formed mainly by around twenty alpha helices. The SASA (solvent accessible surface) is equal to 22 $\mathring{A}^2$). The average of SASA among all proteins (sub-units) of the ground truth is about 10 $\mathring{A}^2$ (standard deviation: 6 $\mathring{A}^2$).*

Despite the limited performance of the mesh-to-structure step of the LZerD workflow, the overall complementary shapes of the Importin beta-1 subunit and of the GTP-binding nuclear protein RAN seem to be retrieved, as 39 true positives were found for this complex only. This conserved relative shape complementarity combined with a scoring function that favors larger interface area may explain why we observed such an over-representation of this complex among the true positives retrieved from the dataset.

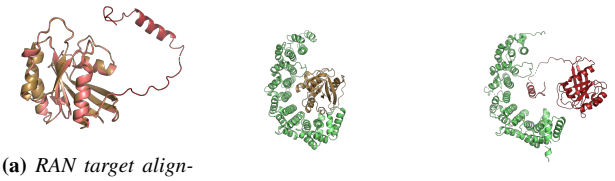
FS-proj tries to find query-target matches using a muti-view approach. Each mesh is reduced by 90%, and the resulting 3D shapes are displaced in multiple directions and rotated in order to produce multiple views of each object that are projected on a 2D plane encoding a 3D information into a 100×100 table. Then, each table

of the queries and of the targets are aligned, and a distance metric evaluates how distant those two tables are, *i.e.* how well the two interfaces (described by the tables) fit to each other. In this workflow, the interfaces are summarized in a 100×100 table reflecting the 3D geometry of a surface observed from a given point of view. Such an approach would depend on the conservation of the critical details about the surfaces throughout the successive steps of the workflow (mesh reduction, multi-view choice and 2D projection) to retrieve the relevant complementary interfaces.

The results presented in Tab. 1 and in Fig. 3 indicate a limited success, both at retrieving the true positives query-target pairs and at predicting the right interaction. Unlike the LZerD method, the ratio true positive / $TP_{complex}$ remains low (2.23 as compared to 8.25) with a maximum of 5 for the complexes formed by the NR1-NR2A subunits of the NMDA receptor and the Complement C3 protein bound to the integrin αM, but in contrast the number of $TP_{complex}$ is high (42.3%). Similarly to LZerD, the interface scores (ICS, IPS and DockQ) are low while the overall alignment scores (lDDT and TM-score) are significantly higher (cf. Fig. 3, middle left panel). Similarly to LZerD, the position and orientation of the target relative to the query were not necessarily well retrieved (cf. Fig. 3, middle right panel), indicating that the table alignment distance metrics is perfectible. Given the constrained timeline of the track, the main parameters of this method have been selected to be able to produce the results (cf. Subsection 3.1). Therefore, a better calibration of the parameters might improve the results.

As a reference point, the results obtained with *Hex* (*cf.* section 2.2.4) are provided as well (bottom panels of Fig. 3). To be noted, these results were obtained using the 52 pairs of protein shapes from the complexes of the ground-truth, and not the queries-targets shapes from the dataset, resulting in a slightly easier challenge. Nevertheless, the results obtained from *Hex* show that *Hex* is able to retrieve the native complex for a few examples (as exemplified by the maximum ICS, IPS and DockQ scores of nearly 1.0, Fig. 3, bottom left panel). Overall, the median scores remain low (< 0.2 for ICS, IPS and DockQ), however. Furthermore, the fraction of complexes for which *Hex* found a query-to-target center-of-mass vector within 30°compared to the ground truth is significantly higher than the participants' results (in about 40% compared to less than 15% for FSProj and less than 5% for LZerD). But in the rest of the cases (60%), *Hex* is unable to find the correct positioning of the query-target chains. Taken together, these results indicate that the correct positioning of the complexes' chains remains a challenge, even in a best case scenario where a method is given the ground truth shapes.

Finally, the workflow we used to generate the dataset produced a few difficult, if not intractable, cases where the query or the target protein conformations from the ground truth and the dataset largely differ. An illustrative purpose was given in Fig. 2 (left panel). As the ground truth and dataset structures were experimentally solved through different experiments and conditions, the accuracy and the portion of proteins actually solved in the structures may mislead the docking algorithms as the changes in the structures and the surficial shapes are substantial. The RAN protein is a good example of how two protein shapes can radically differ while representing the same protein. The C-terminal extremity of this protein is so flexible that it is absent in most PDB structures, as the experimental methods are not able to solve such dynamic portions of proteins. Due to the presence of some protein partners, however, this region may fold into a stable helix, and therefore be solved and appear in the PDB structure (cf. Fig. 5). The problem difficulty is then considerably increased, and the methods may not handle such difficult cases, thus lowering the results.



(a) *RAN target alignment from ground truth and dataset.* **(b)** *Ground truth: query-target complex.* **(c)** *Prediction: query-target complex.*

Figure 5: *Impact of the structural difference with the RAN protein example, mostly formed by alpha helices. In (a), we show the RAN protein difference between conformations from the ground truth (sand color) and dataset (lig red color). The dark red show a C-extremity of about 40 residues, with an alpha helix, present in the dataset's structure and absent in the ground truth's structure. Subfigures (b) and in (c) show the complex between RAN (target) and Importin beta-1 (query) from the ground truth and from a prediction respectively, to highlight the impact of this C-extremity.*

5. Conclusion

In this paper, we have proposed a challenge on the non-rigid shape complementarity retrieval in protein-protein interactions. This is a particularly difficult challenge for several reasons: i) the restricted time available for the challenge, ii) the complex and unusual shape of 3D objects such as proteins, iii) a dataset that does not include the same protein conformations as the ground truth and, iv) the presence of several rearrangements, sometimes significant, derived from the latter. We have described the approaches of two participants: LZerD by Kihara *et al.* and FS-proj by Tran *et al.* , and compared their results to those obtained with *Hex*. Apart from the difficulties of the challenge, the mixed results (particularly the retrieval) may be explained by the fact that all parts of the participants' workflows were developed *ad hoc*, probably with sub-optimal parameters. Hex produced slightly better results, but remain far from the ideal. Nevertheless, this challenge poses the bases for the development of protein shapes docking methods in the future, with a first curated dataset made publicly available which could be further expanded and prospects for optimizing existing algorithms.

Acknowledgments

Y. Kagaya, T. Nakamura and D. Kihara are thankful for technical help and discussion by Charles Christoffer. Y. Kagaya, T. Nakamura and D. Kihara are also grateful to the Rosen Center for Advanced Computing of Purdue University for providing computational resources. The work of Y. Kagaya, T. Nakamura and D. Kihara is partly supported by fundings from the National Institutes of Health (R01GM133840, R01GM123055, and 3R01GM133840-02S1) and the National Science Foundation (CMMI1825941,

MCB1925643, DBI2003635, and DBI2146026). T. Yacoub, F. Langenfeld and M. Montes are supported by the French Agence Nationale de la Recherche (ANR22-CE23-GRADIENT) and by the European Research Council Executive Agency under the research grant number 640283.

References

[ABC*19] ARMSTRONG, DAVID R et al. "PDBe: improved findability of macromolecular structure data in the PDB". *Nucleic Acids Research* (Nov. 2019). DOI: 10.1093/nar/gkz990.

[BHN03] BERMAN, HELEN, HENRICK, KIM, and NAKAMURA, HARUKI. "Announcing the worldwide Protein Data Bank". *Nature Structural & Molecular Biology* 10.12 (Dec. 2003), 980–980. DOI: 10.1038/nsb1203-980.

[BSB*13] BIASINI, M. et al. "OpenStructure: an integrated software framework for computational structural biology". *Acta Crystallographica Section D* 69.5 (May 2013), 701–709. DOI: 10.1107/S0907444913007051.

[BW16] BASU, SANKAR and WALLNER, BJÖRN. "DockQ: A Quality Measure for Protein-Protein Docking Models". *PLOS ONE* 11.8 (Aug. 2016), 1–9. DOI: 10.1371/journal.pone.0161879.

[CTS*19] CHRISTOFFER, CHARLES et al. "Performance and enhancement of the LZerD protein assembly pipeline in CAPRI 38-46". *Proteins: Structure, Function, and Bioinformatics* 88.8 (Nov. 2019), 948–961. DOI: 10.1002/prot.25850.

[DGA*09] DUTAGACI, H. et al. "SHREC'09 track: querying with partial models". *Proceedings of the 2nd Eurographics Conference on 3D Object Retrieval*. 3DOR '09. Munich, Germany: Eurographics Association, 2009, 69–76. ISBN: 9783905674163.

[FKB04] FEIG, MICHAEL, KARANICOLAS, JOHN, and BROOKS, CHARLES L. "MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology". *Journal of Molecular Graphics and Modelling* 22.5 (2004). Conformational Sampling, 377–395. DOI: 10.1016/j.jmgm.2003.12.005.

[GRF*22] GAGLIARDI, LUCA et al. "SHREC 2022: Protein–ligand binding site recognition". *Computers & Graphics* 107 (Oct. 2022), 20–31. DOI: 10.1016/j.cag.2022.07.005.

[GVZ*21] GUEST, JOHNATHAN D. et al. "An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants". *Structure* 29.6 (June 2021), 606–621.e5. DOI: 10.1016/j.str.2021.01.005.

[HLS*15] HRABE, THOMAS et al. "PDBFlex: exploring flexibility in protein structures". *Nucleic Acids Research* 44.D1 (Nov. 2015), D423–D428. DOI: 10.1093/nar/gkv1316.

[HYM14] HACKER, THOMAS, YANG, BAIJIAN, and MCCARTNEY, GERRY. "Empowering Faculty: A Campus Cyberinfrastructure Strategy for Research Communities". *Educause Review* (July 2014).

[ILJ*20] IYER, MALLIKA et al. "Difference contact maps: From what to why in the analysis of the conformational flexibility of proteins". *PLOS ONE* 15.3 (Mar. 2020). Ed. by SALSBURY, FREDDIE, e0226702. DOI: 10.1371/journal.pone.0226702.

[KSCE11] KIHARA, DAISUKE et al. "Molecular Surface Representation Using 3D Zernike Descriptors for Protein Shape Comparison and Docking". *Current Protein & Peptide Science* 12.6 (Sept. 2011), 520–530. DOI: 10.2174/138920311796957612.

[LAC*21] LANGENFELD, FLORENT et al. "SHREC 2021: Surface-based Protein Domains Retrieval". *Eurographics Workshop on 3D Object Retrieval* (2021). DOI: 10.2312/3DOR.20211308.

[LBK*18a] LAFITA, ALEIX et al. "Assessment of protein assembly prediction in CASP12". *Proteins: Structure, Function, and Bioinformatics* 86.S1 (2018), 247–256. DOI: 10.1002/prot.25408.

[LBK*18b] LAFITA, ALEIX et al. "Assessment of protein assembly prediction in CASP12". *Proteins: Structure, Function, and Bioinformatics* 86.S1 (2018), 247–256. DOI: 10.1002/prot.25408.

[LBM*21] LENSINK, MARC F. et al. "Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment". *Proteins: Structure, Function, and Bioinformatics* 89.12 (Sept. 2021), 1800–1823. DOI: 10.1002/prot.26222.

[LBR*23] LENSINK, MARC F. et al. "Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment". *Proteins: Structure, Function, and Bioinformatics* 91.12 (Oct. 2023), 1658–1683. DOI: 10.1002/prot.26609.

[MBBS13] MARIANI, VALERIO et al. "lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests". *Bioinformatics* 29.21 (Aug. 2013), 2722–2728. DOI: 10.1093/bioinformatics/btt473.

[PSA*16] PRATIKAKIS, I. et al. "Partial Shape Queries for 3D Object Retrieval". *Eurographics Workshop on 3D Object Retrieval*. Ed. by FERREIRA, A., GIACHETTI, A., and GIORGI, D. The Eurographics Association, 2016, 79–88. ISBN: 978-3-03868-004-8. DOI: 10.2312/3dor.20161091.

[RFB*21] RAFFO, ANDREA et al. "SHREC 2021: Retrieval and classification of protein surfaces equipped with physical and chemical properties". *Computers & Graphics* 99 (Oct. 2021), 1–21. DOI: 10.1016/j.cag.2021.06.010.

[RK00] RITCHIE, DAVID W. and KEMP, GRAHAM J.L. "Protein docking using spherical polar Fourier correlations". *Proteins: Structure, Function, and Bioinformatics* 39.2 (2000), 178–194. DOI: 10.1002/(SICI)1097-0134(20000501)39:2<178::AID-PROT8>3.0.CO;2-6.

[RTTB18] RODRIGUES, JPGLM et al. "pdb-tools: a swiss army knife for molecular structures [version 1; peer review: 2 approved]". *F1000Research* 7.1961 (2018). DOI: 10.12688/f1000research.17456.1.

[VMV*15] VREVEN, THOM et al. "Updates to the Integrated Protein–Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2". *Journal of Molecular Biology* 427.19 (Sept. 2015), 3031–3041. DOI: 10.1016/j.jmb.2015.07.016.

[VYSK09] VENKATRAMAN, VISHWESH et al. "Protein-protein docking using region-based 3D Zernike descriptors". *BMC Bioinformatics* 10.1 (Dec. 2009). DOI: 10.1186/1471-2105-10-407.

[XLZ13] XU, DONG, LI, HUA, and ZHANG, YANG. "Protein Depth Calculation and the Use for Improving Accuracy of Protein Fold Recognition". *Journal of Computational Biology* 20.10 (Oct. 2013), 805–816. DOI: 10.1089/cmb.2013.0071.

[XZ09] XU, DONG and ZHANG, YANG. "Generating Triangulated Macromolecular Surfaces by Euclidean Distance Transform". *PLoS ONE* 4.12 (Dec. 2009). Ed. by BUEHLER, MARKUS J., e8140. DOI: 10.1371/journal.pone.0008140.

[ZPK18] ZHOU, QIAN-YI, PARK, JAESIK, and KOLTUN, VLADLEN. "Open3D: A Modern Library for 3D Data Processing". *CoRR* abs/1801.09847 (2018). arXiv: 1801.09847.

[ZS04] ZHANG, YANG and SKOLNICK, JEFFREY. "Scoring function for automated assessment of protein structure template quality". *Proteins: Structure, Function, and Bioinformatics* 57.4 (2004), 702–710. DOI: 10.1002/prot.20264.

[ZSG07] ZHANG, YING, STEC, BOGUSLAW, and GODZIK, ADAM. "Between Order and Disorder in Protein Structures: Analysis of "Dual Personality" Fragments in Proteins". *Structure* 15.9 (Sept. 2007), 1141–1147. DOI: 10.1016/j.str.2007.07.012.

[ZSO09] ZHANG, QING, SANNER, MICHEL, and OLSON, ARTHUR. "Shape Complementarity of Protein-Protein Complexes at Multiple Resolutions". *Proteins* 75 (May 2009), 453–67. DOI: 10.1002/prot.22256.