



# LZerD Protein-Protein Docking Webserver Enhanced With *de novo* Structure Prediction

Charles Christoffer<sup>1</sup>, Vijay Bharadwaj<sup>1</sup>, Ryan Luu<sup>1</sup> and Daisuke Kihara<sup>1,2\*</sup>

<sup>1</sup>Department of Computer Science, Purdue University, West Lafayette, IN, United States, <sup>2</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN, United States

Protein-protein docking is a useful tool for modeling the structures of protein complexes that have yet to be experimentally determined. Understanding the structures of protein complexes is a key component for formulating hypotheses in biophysics regarding the functional mechanisms of complexes. Protein-protein docking is an established technique for cases where the structures of the subunits have been determined. While the number of known structures deposited in the Protein Data Bank is increasing, there are still many cases where the structures of individual proteins that users want to dock are not determined yet. Here, we have integrated the AttentiveDist method for protein structure prediction into our LZerD webserver for protein-protein docking, which enables users to simply submit protein sequences and obtain full-complex atomic models, without having to supply any structure themselves. We have further extended the LZerD docking interface with a symmetrical homodimer mode. The LZerD server is available at <https://lzerd.kiharalab.org/>.

**Keywords:** web server, LZerD, structure modeling, protein bioinformatics, protein-protein docking, protein structure prediction, symmetrical docking

## INTRODUCTION

Protein-protein interactions are key components of many biological processes, and the three-dimensional (3D) structures of the protein-protein complexes thus formed are a crucial resource for reasoning about their molecular functions. Ideally, structures of these complexes would be determined experimentally, through techniques such as X-ray crystallography or cryo-electron microscopy. However, it is possible to instead use computational methods to construct atomic structure models of protein complexes (Aderinwale et al., 2020). This class of methods is called protein-protein docking, and suitably constructed models from docking can be used to reason about how molecular functions are carried out in the living cell, even in the absence of an experimentally determined complex structure (Sanyal et al., 2021). Many protein-protein docking methods exist, such as LZerD (Venkatraman et al., 2009), Multi-LZerD (Esquivel-Rodríguez et al., 2012), ZDOCK (Mintseris et al., 2007), HADDOCK (Dominguez et al., 2003), ClusPro (Kozakov et al., 2017), RosettaDock (Lyskov and Gray, 2008), HEX (Ritchie and Venkatraman, 2010), SwarmDock (Torchala et al., 2013), ATTRACT (de Vries and Zacharias, 2013), and SymmDock (Schneidman-Duhovny et al., 2005). Previously, we released a web-based tool which allows free, easy, installation-free access to LZerD (Christoffer et al., 2021). Users can perform pairwise and multiple chain docking in the LZerD server. Users can also provide additional information in the form of distances of interacting or non-interacting residues to guide docking. The LZerD suite of methods has been ranked at or near the top of all server groups in recent rounds of CAPRI (Lensink et al., 2018; Lensink et al., 2019; Lensink et al., 2020), the blind communitywide assessment of protein docking methods.

## OPEN ACCESS

### Edited by:

Brian Jiménez-García,  
Utrecht University, Netherlands

### Reviewed by:

Sebastian Kmiecik,  
University of Warsaw, Poland  
Simone Orioli,  
University of Copenhagen, Denmark

### \*Correspondence:

Daisuke Kihara  
dkihara@purdue.edu

### Specialty section:

This article was submitted to  
Biological Modeling and Simulation,  
a section of the journal  
Frontiers in Molecular Biosciences

Received: 14 June 2021

Accepted: 21 July 2021

Published: 12 August 2021

### Citation:

Christoffer C, Bharadwaj V, Luu R and  
Kihara D (2021) LZerD Protein-Protein  
Docking Webserver Enhanced With  
*de novo* Structure Prediction.  
Front. Mol. Biosci. 8:724947.  
doi: 10.3389/fmolb.2021.724947

Biologists seeking to model a complex computationally may already have structures of individual subunits determined by experiment. However, where no such structures are available, a structure model can be constructed. Single-chain protein structure prediction methods have recently matured, and can often generate models in the absence of clear global template structures (Kryshtafovych et al., 2019). Such methods include our AttentiveDist (Jain et al., 2021), trRosetta (Yang et al., 2020), RaptorX (Xu and Wang, 2019), and QUARK (Zheng et al., 2019). Relative to the top existing servers participating in CASP13 (Kryshtafovych et al., 2019), AttentiveDist showed competitive performance when evaluated on the CASP13 (Kryshtafovych et al., 2019) dataset (Jain et al., 2021).

The underlying methods implemented in the LZerD webserver, LZerD (Christoffer et al., 2021), Multi-LZerD (Esquivel-Rodríguez et al., 2012), and AttentiveDist (Jain et al., 2021), and the ranksum model scoring function (Peterson et al., 2017a; Peterson et al., 2018a; Christoffer et al., 2020) have been rigorously examined in their original papers. In this article, we present the current version of the LZerD webserver, with new functionality of *de novo* prediction of subunit structures by AttentiveDist and applying symmetry constraints for homodimer modeling. We provide step-by-step instruction with examples of modeling in three different scenarios.

## MATERIALS

To perform protein docking, it is best if users have experimentally determined 3D structures. In case the structure is not available, structures can be modelled from the amino acid sequences of the proteins in question. In practice, it is recommended that as much information about the protein structures and their interactions as possible be gathered in advance. For example, there should be evidence that the proteins in question do in fact form a complex, e.g., from a biochemical assay or a biophysical experiment. It is even more desirable that information be known about specific residue interactions or non-interactions. As discussed in later sections, such information can even be provided directly to the server.

## METHODS

### AttentiveDist Protein Structure Prediction

For structure modeling of individual proteins, the LZerD server uses AttentiveDist (Jain et al., 2021). If users have 3D structures of individual proteins to dock, they can skip the AttentiveDist step. Here, we give a brief overview the algorithm of AttentiveDist.

In the first stage of AttentiveDist, four multiple sequence alignments (MSAs) with E-value cutoffs of 0.001, 0.1, 1, and 10 are generated using the DeepMSA (Zhang et al., 2020) method, which uses HH-suite (Steinegger et al., 2019) and HMMER (Johnson et al., 2010) to generate MSAs from the UniClust30 (Mirdita et al., 2017), UniRef90 (Suzek et al., 2015), and Metaclust (Steinegger and Soding, 2018) protein sequence databases. A trained neural network is then fed the amino acid types, the PSI-BLAST (Altschul et al., 1997) position-specific scoring

matrix, the HMM profile, the secondary structure and solvent-accessible surface area predicted by SPOT-1D (Hanson et al., 2019), rough contacts predicted by CCMPRED (Seemayer et al., 2014), mutual information, and a statistical pairwise potential. The output of this neural network is a prediction of the distribution of residue-residue distances.

Once generated, the predicted distance distribution is converted into full-atom structure models by L-BFGS minimization of predicted short-, medium-, and long-range distance restraints in sequence using PyRosetta (Chaudhury et al., 2010). This minimization results in a pool of models which are then scored by a ranksum method (Christoffer et al., 2020; Peterson et al., 2017a; Peterson et al., 2018a) which aggregates the rankings of the pool by the knowledge-based scoring functions GOAP (Zhou and Skolnick, 2011), DFIRE (Zhou and Zhou, 2002), and ITScorePro (Huang and Zou, 2014), and additionally Rosetta's REF2015 score (Park et al., 2016), into a single ranking. This ranked pool of models is the end output of AttentiveDist.

Heuristically, a model output by AttentiveDist with a ranksum score  $\leq 20$  can be considered particularly confident, and a ranksum gap between models of  $\geq 2R$  can be considered significant, where  $R$  is the ranksum score of the top-ranked model. A ranksum of 20 is the score given when a model is ranked fifth by all component scores and guarantees that at least one component score has ranked the model within the top 5.

Although AttentiveDist was shown to have competitive performance at the time of the development (Jain et al., 2021), there are more recent methods that showed promising performance. Users are also encouraged to try such servers, perhaps those which performed well in recent CASP (Kryshtafovych et al., 2019). Single chain models built by an outside method can be uploaded to the LZerD server.

### Using the AttentiveDist Web Interface

The AttentiveDist web interface ([https://lzerd.kiharalab.org/upload/upload\\_sequence](https://lzerd.kiharalab.org/upload/upload_sequence)) allows users to submit up to six separate protein sequences for structure prediction at a time. Jobs are limited to six proteins because six is the maximum number of proteins the LZerD server can dock using Multi-LZerD. To submit a sequence, users can simply paste their sequence in FASTA format into the large text box, as shown in **Figure 1A**. Due to resource constraints, users are limited to 1,000 amino acids per sequence. To submit additional sequences, simply click the "+" button to create a new submission field. Finally, clicking Submit will submit the job. Users can further configure email notification settings, but this is not necessary.

After the prediction job has been submitted and has finished running, users are presented with a results summary page, shown in **Figure 1B**. This page contains rows of 3D visualizers showing the top five models for each submitted sequence. Below each row of visualizers is a table containing the scoring and ranking data for the output model set. Models can be downloaded in bulk as compressed archives or individually by clicking the appropriately labeled buttons.

From the five models presented for each chain, users need to choose one to perform docking by checking "Forward to



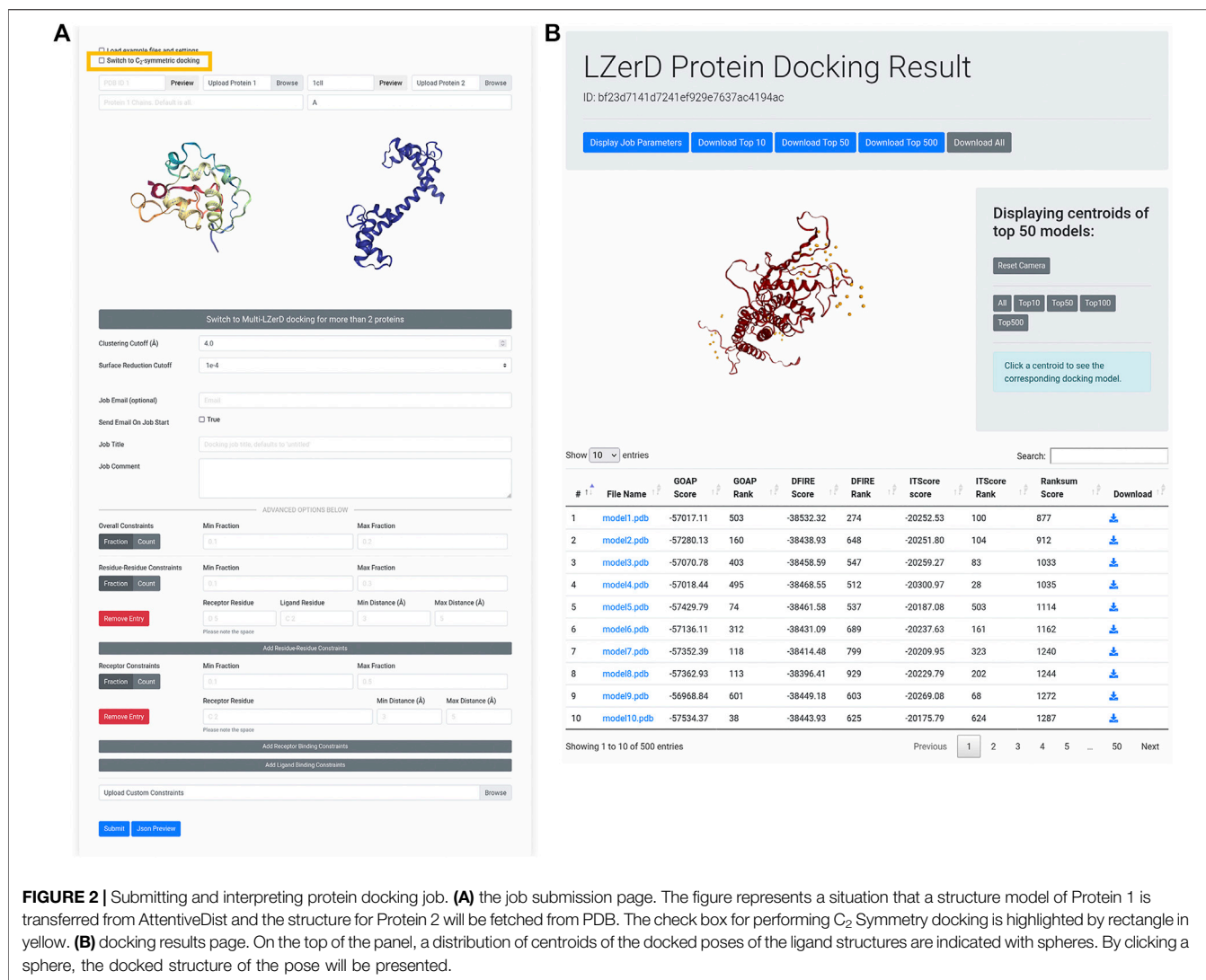
**FIGURE 1** | AttentiveDist panels. **(A)** The input page of AttentiveDist where users can input amino acid sequences of subunits to model. By clicking the “+” button, additional sequence submission field will appear. The maximum number of sequences users can submit is six. **(B)** the result panel of AttentiveDist. For each submitted sequence, top five scoring models are visualized. Scores of the five models are shown in a table below the visualization panel. The models are ranked by the ranksum score.

docking”. After choosing a model to dock, click either “Dock Using Pairwise LZerD” or “Dock Using Multi-LZerD” to send the models to the docking step. Then, the structure model will be sent to the subsequent step, the protein docking by LZerD. The panel only shows up to the top five models, but up to 20 models can be downloaded locally. If users want, they can examine models within top 20 locally, by using a structure viewer, such as PyMOL (The PyMOL Molecular Graphics System, 2019). Then upload the selected model directly in the input page of the LZerD server.

## LZerD Protein-Protein Docking

The main protein docking engine of the webserver is LZerD for pairwise docking and Multi-LZerD for multiple subunit docking of up to six subunits. Here we briefly explain the algorithms of LZerD and Multi-LZerD.

LZerD takes two structures provided by users (conventionally referred to as receptor and ligand in descending order of size) as input and samples of all possible interaction interface regions and interaction angles exhaustively. If a putative complex structure contains excessive steric violations at the interface, has too small an



**FIGURE 2 |** Submitting and interpreting protein docking job. **(A)** the job submission page. The figure represents a situation that a structure model of Protein 1 is transferred from AttentiveDist and the structure for Protein 2 will be fetched from PDB. The check box for performing C<sub>2</sub> Symmetry docking is highlighted by rectangle in yellow. **(B)** docking results page. On the top of the panel, a distribution of centroids of the docked poses of the ligand structures are indicated with spheres. By clicking a sphere, the docked structure of the pose will be presented.

interaction area, or has low shape complementarity at the interface region, that model is rejected. In LZerD, a protein structure is represented by a molecular surface, which is segmented into overlapping local surface regions. Each local surface region is represented both by a mathematical moment-based shape descriptor called a 3D Zernike descriptor (3DZD) (Kihara et al., 2011). 3DZDs are rotation-invariant, which allows fast, alignment-free computation of shape complementarity, and also allows a soft representation of surface that is robust to a certain degree of conformational change induced by the interaction. This advantage extends to tolerance of small modeling inaccuracies. The conformational space is searched by the geometric hashing algorithm. If constraints of residue-residue distances, interface residues, or symmetry tolerance have been provided, models violating the constraints are rejected.

LZerD generates tens of thousands of docking models. After clustering (by default at an RMSD cutoff of 4.0 Å), which generally culls the model pool to a few thousand to a few tens of thousands, the complex models are then scored by a ranksum method (Peterson et al., 2017a; Peterson et al., 2018a; Christoffer

et al., 2020) which aggregates the rankings of the pool by the knowledge-based scoring functions GOAP (Zhou and Skolnick, 2011), DFIRE (Zhou and Zhou, 2002), and ITSPro (Huang and Zou, 2014). These three scoring functions essentially check if atom interactions in a model have distances and angles that agree with those observed in experimentally determined protein structures overall. If a model is consistently ranked as the top by all the component scores, then the ranksum of the model will be low, e.g. 3 if a model is ranked 1 by all component scores. Ranksum has been shown to perform very well in docking model ranking in CAPRI protein docking assessments (Lensink et al., 2019; Lensink et al., 2020).

Multi-LZerD takes three to six protein structures and builds them into a complex. It first runs pairwise LZerD for each pair of subunits to generate a pool of pairwise docking models. Then, it subsequently selects pairwise models and assembles into full subunit models. Combinations of pairwise models are iteratively optimized by a genetic algorithm. In Multi-LZerD, models are selected with a molecular mechanics force field with terms reweighted specifically for protein docking.

## Homodimer Docking ( $C_2$ Symmetry)

Symmetrical protein complexes are often observed in nature (Levy et al., 2006). Cyclic symmetry of general order  $n$  is usually referred to as  $C_n$  symmetry. The LZerD server supports docking with  $C_2$  symmetry, i.e., homodimers, in the current version of the server because homodimers are quite common. A measure of the  $C_2$  symmetry of a model can be constructed by transforming the atomic coordinates of the subunit by applying the rigid body transformation from docking twice and calculating the root-mean-square deviation (RMSD) to the original coordinates. If the model is perfectly symmetrical, this RMSD will be 0 Å. For the  $C_2$  symmetry constraint functionality provided in this webserver, this RMSD is cut off at 5.0 Å, and models exceeding it in symmetry mode are discarded.

## Using the Docking Web Interface

Figure 2 shows the interface for submitting individual protein structures for docking. Once users have forwarded subunit structures from the AttentiveDist web interface, the structures appear in the input panel as shown in Figure 2A. Alternatively, users can upload structures from their local disk or fetch them directly from the Protein Data Bank (PDB) (Berman et al., 2000) via the upload widget. It is also possible to model one of the subunit structures by AttentiveDist and dock it with a structure from PDB. Clicking the Submit button will start the docking computation.

This will run LZerD with the recommended default settings and without constraints. If users prefer, they can directly change parameters: the clustering RMSD cutoff, which controls the redundancy of the output model pool, as well as the surface reduction cutoff, which in part controls how finely the conformational space is sampled. If users are modeling a symmetrical homodimer, they can select the checkbox to model with a  $C_2$  symmetry constraint, which excludes parts of the conformational space greater than 5 Å RMSD from perfect symmetry. Users can further optionally supply an email for receiving job notifications, whether to receive a notification when a job begins running, a job title, and a job comment describing what the job is modeling.

Below, in the advanced options section, users can specify distance constraints for specific residue-residue interactions or residue-subunit interactions. Distances used for these constraints should ideally come from experiment, but can also come from computational predictions (La et al., 2013, La and Kihara, 2012). All distance constraints specify an allowed range for the closest heavy (non-hydrogen) atoms between two selections of atoms. For residue-residue constraints, the selections are the atoms belonging to the two specified residues. For receptor binding site constraints, the selections are the specified receptor residue and the entire ligand subunit. For ligand binding site constraints, the selections are the specified ligand residue and the entire receptor subunit. If a particular pair of residues should be in contact, a user could for example specify a distance range of 0–5 Å, or perhaps 0–8 Å or even broader depending on what available data from experiment might indicate. To specify that two residues should not interact, a user could for example specify an exclusionary minimum distance such as 15 Å. The same logic can be applied to receptor binding site and ligand binding site constraints. To control the

number of constraints that must be satisfied, users can set the min/max fraction fields, which directly specify what proportion of the distance constraints should be satisfied. A toggle is available for users to switch from specifying the proportion to specifying the actual numbers of constraints that are allowed to be satisfied.

The submission process is essentially the same for Multi-LZerD. Only the difference is, naturally, to specify three or more (up to six) subunit structures to assemble. To switch to Multi-LZerD, click the large gray button of “Switch to Multi-LZerD docking for more than 2 proteins”.

After the docking job has been submitted and has finished running, the user is presented with a results summary page (Figure 2B). This page contains a 3D visualizer showing the distribution of ligand centroids of the top-scoring docked models, with a user selectable top-k threshold. Below the visualizer is a table containing the scoring and ranking data for the output model set. As default, models are ranked by the ranksum score. Users can choose another score to sort the models by clicking an arrow of the preferred score. Users can click on a model to display its 3D structure. Models belonging to particular centroids can be displayed by clicking on those centroids. Docked models can be downloaded in bulk as compressed archives or individually by clicking the appropriately labeled buttons.

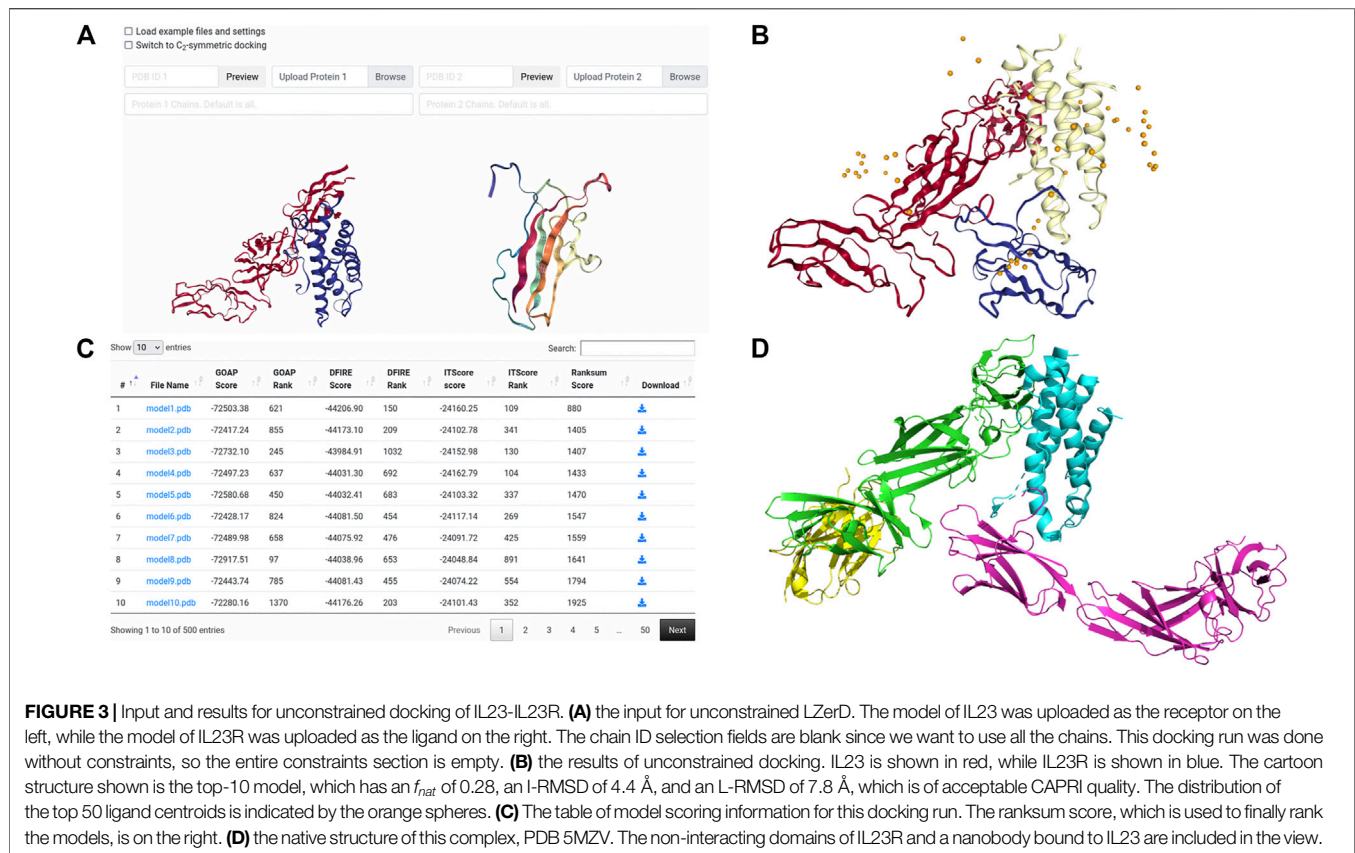
For more information about the job submission steps and interpretation of results page, users are encouraged to refer to the instructions on the LZerD web server. From the top bar, the information is available from the “About” pull-down menu.

## RESULTS

We discuss three case studies of docking modeling in different scenarios. The first case is the regular pairwise protein docking. The second case is homodimer docking, while the last case is docking with structure models built through the AttentiveDist pipeline.

### Case Study: Regular LZerD Docking of Human IL23-IL23R

The first case study provides an example of basic pairwise protein docking modeling. During CAPRI Round 39, a complex of a human cytokine heterodimer IL23 with a human IL23R monomer was given for prediction as target T122. CAPRI is a blind experiment where the structure of the complex to be modeled is not known to the predictors until well after all predictions have been submitted. In this example we show a similar result to the LZerD server group’s modeling performance on this target (Christoffer et al., 2020). As the LZerD server group did during CAPRI, here we used an unbound structure from X-ray crystallography provided by the organizers for IL23, which has a root-mean-square deviation (RMSD) of 1.7 Å to the native T122 structure. For the ligand, we used a model of IL23R generated by template-based modeling with MODELLER (Webb and Sali, 2021) which has been truncated to the actual interacting domain, which has an RMSD of 3.2 Å to the native T122 structure. As suggested by the low RMSD, the template-based modeling was reasonably successful at predicting the structure of IL23R. For the docking, we set no constraints. In Figure 3, we show the results of



this LZerD job. The top-10 model shown is acceptable under the CAPRI evaluation criteria, with a ligand RMSD (L-RMSD) of 7.8 Å and a fraction of native contacts ( $f_{nat}$ ) of 0.28 relative to the native structure PDB 5MZV shown in **Figure 3D**. Although this result is good, the richness of the result set can be increased by adding constraints, as discussed in the corresponding case study section in a previous paper (Christoffer et al., 2021). As seen by the centroid distribution in **Figure 3B**, the ligand models are not concentrated at the native interaction site.

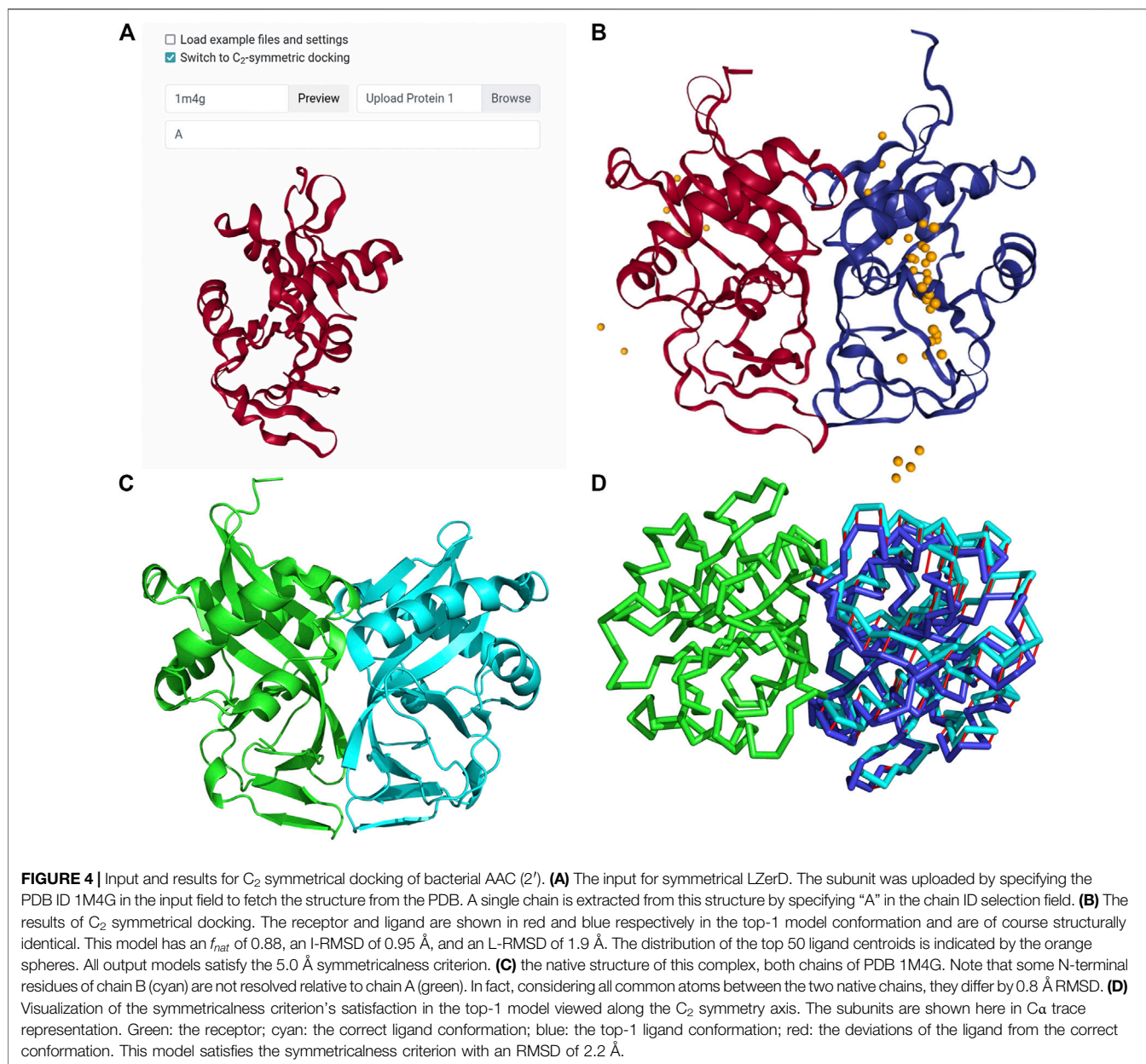
### Case Study: LZerD Docking of Bacterial Aminoglycoside 2'-N-Acetyltransferase Homodimer With $C_2$ Symmetry Constraint

This is an example of modeling a homodimer complex with the newly implemented homodimer constraint. The complex used is aminoglycoside 2'-N-acetyltransferase (AAC (2')) from *M. tuberculosis*, which appears to directly relate to the drug resistance of the organism (Vetting et al., 2002). This example was used in a work by Ritchie and Grudinin on symmetrical protein docking (Ritchie and Grudinin, 2016). To create the LZerD job, we used the PDB ID input method to specify that 1M4G should be fetched from the PDB. Then, we used the Protein 1 Chains field to specify that only chain A should be considered. Finally, we selected the checkbox to switch to symmetrical docking (**Figure 4A**) and clicked Submit. In the referenced work (Ritchie and Grudinin, 2016), a docked model was considered a hit if it had an RMSD within 10 Å of

the native structure. According to their paper (Ritchie and Grudinin, 2016), M-ZDOCK and SymmDock found no hits within the top 10, while SAM's top-1 model was a hit with RMSD 1.82 Å. LZerD in  $C_2$  symmetry mode's top-1 model, shown in **Figure 4B**, was a hit, with an even lower RMSD of 0.94 Å. By the CAPRI criteria, this model is of high quality, with an  $f_{nat}$  of 0.88, an I-RMSD of 0.95 Å, and an L-RMSD of 1.9 Å. **Figure 4C** shows the native structure. In **Figure 4D**, we examined the symmetry of our model. Our model (blue) has an RMSD of 2.2 Å to the correct pose (cyan) that is located at the perfect symmetrical position.

### Case Study: LZerD Docking of Designed colEdes3:Imdes3 With Site-Directed Mutagenesis Data

In this last example, we started from the sequence of individual proteins to model their tertiary structures by AttentiveDist, which were then docked to obtain complex models. In CAPRI round 43, a redesigned version of the *E. coli* colicin-E2:DNase-Im2 complex was presented as the target T133. The structure of this complex has since been released, and is available as an entry in PDB, 6ERE. The designed sequences of the first assembly, i.e. chains B and C, can be taken from the PDB entry page (<https://www.rcsb.org/structure/6ERE>). We modeled this complex from the sequence information without using any template structures. We clicked Upload Protein Sequences to bring up the AttentiveDist submission form and pasted the colEdes3 sequence. Then, we clicked the green plus button to add

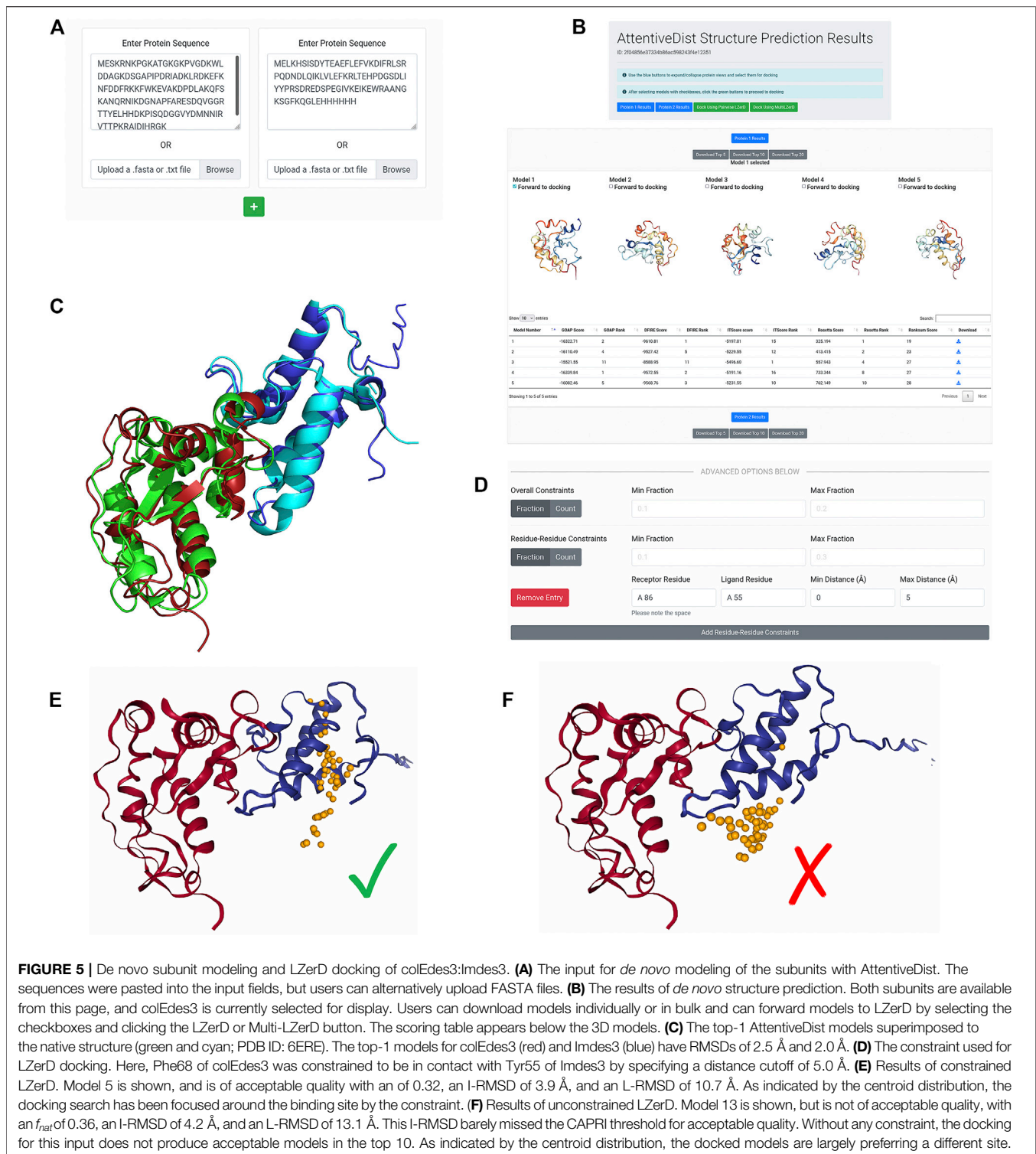


another sequence input field and pasted the Imdes3 sequence there. The filled submission form is shown in **Figure 5A**. **Figure 5B** shows the result of the single-chain modeling. From this panel, we selected the first model for each subunit as shown in **Figure 5B**, and then clicked Dock Using Pairwise LZerD to forward both models to the LZerD submission page. The top-1 models for colEdes3 and Imdes3 have RMSDs of 2.5 Å and 2.0 Å to the native structure, respectively, and can be seen superimposed in **Figure 5C**.

From the site-directed mutagenesis experiments, we knew that Tyr55 of DNase-Im2 is a hotspot residue (Netzer et al., 2018). Further, this residue is conserved by the designed sequence of Imdes3 (Wojdyla et al., 2012). The same is true for Phe86 of colicin-E2, a key specificity site. Thus, to focus the docking, we integrated this information into our docking job. To accomplish

this, we clicked Add Residue-Residue Constraints and create a constraint between receptor residue "A 86" indicating chain A residue sequence number 86, and ligand residue "A 55", with minimum distance 0 Å and maximum distance 5 Å (**Figure 5D**). This configuration tells the LZerD server that Phe86 of colEdes3 and Tyr55 of Imdes3 should be in direct contact with each other. We left the Min/Max Fraction fields blank since we were only specifying one single constraint.

On the results page, shown in **Figure 5E**, the distribution of ligand centroids about Phe86 of colEdes3 is clearly visible. The top-10 model pool contains two acceptable models: model 5 is acceptable with an  $f_{nat}$  of 0.32, an I-RMSD of 3.9 Å, and an L-RMSD of 10.7 Å; model 8 is acceptable with an  $f_{nat}$  of 0.40, an I-RMSD of 3.8 Å, and an L-RMSD of 11.6 Å. In this example, the residue constraints were



**FIGURE 5** | De novo subunit modeling and LZerD docking of colEdes3:Imdes3. **(A)** The input for *de novo* modeling of the subunits with AttentiveDist. The sequences were pasted into the input fields, but users can alternatively upload FASTA files. **(B)** The results of *de novo* structure prediction. Both subunits are available from this page, and colEdes3 is currently selected for display. Users can download models individually or in bulk and can forward models to LZerD by selecting the checkboxes and clicking the LZerD or Multi-LZerD button. The scoring table appears below the 3D models. **(C)** The top-1 AttentiveDist models superimposed to the native structure (green and cyan; PDB ID: 6ERE). The top-1 models for colEdes3 (red) and Imdes3 (blue) have RMSDs of 2.5 Å and 2.0 Å. **(D)** The constraint used for LZerD docking. Here, Phe68 of colEdes3 was constrained to be in contact with Tyr55 of Imdes3 by specifying a distance cutoff of 5.0 Å. **(E)** Results of constrained LZerD. Model 5 is shown, and is of acceptable quality with an of 0.32, an I-RMSD of 3.9 Å, and an L-RMSD of 10.7 Å. As indicated by the centroid distribution, the docking search has been focused around the binding site by the constraint. **(F)** Results of unconstrained LZerD. Model 13 is shown, but is not of acceptable quality, with an  $f_{\text{best}}$  of 0.36, an I-RMSD of 4.2 Å, and an L-RMSD of 13.1 Å. This I-RMSD barely missed the CAPRI threshold for acceptable quality. Without any constraint, the docking for this input does not produce acceptable models in the top 10. As indicated by the centroid distribution, the docked models are largely preferring a different site.

effective in guiding the docking. To compare, in **Figure 5F**, we show docking results when the constraints were not provided. As shown, without the constraints the ligand was attracted at an incorrect place, which made a larger interface between the two proteins.

This example demonstrates several qualities of the pipeline components and qualities of the pipeline as a whole. As shown

quantitatively by the metrics and visually in **Figure 5C**, AttentiveDist was capable of accurately modeling the structures of individual protein chains. As shown by the unconstrained result, the correct interface was sampled in docking with no template or extra interaction information, although a model of acceptable CAPRI quality was not ranked at the top. As shown quantitatively by the metrics and



visually in **Figure 5E**, the barest of residue-residue interaction information was sufficient to overcome this and produce acceptable models among the top ranks.

## DISCUSSION

With the upgraded LZerD web interface, biologists can conveniently construct protein complex models that they can use to reason about the interactions in their system. Information about the system can be integrated in the form of geometric distance and symmetry constraints. Now, even without known or template-modeled subunit structures, users can generate *de novo* predictions of subunit structures and dock them with the click of a button. Future development of the LZerD web platform is expected to include modeling complexes with intrinsically disordered proteins using IDP-LZerD (Peterson et al., 2017b), as well as modeling the assembly order of multimeric complexes using Path-LZerD (Peterson et al., 2018b).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## REFERENCES

- Aderinwale, T., Christoffer, C. W., Sarkar, D., Alnabati, E., and Kihara, D. (2020). Computational Structure Modeling for Diverse Categories of Macromolecular Interactions. *Curr. Opin. Struct. Biol.* 64, 1–8. doi:10.1016/j.sbi.2020.05.017
- Altschul, S., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Chaudhury, S., Lyskov, S., and Gray, J. J. (2010). PyRosetta: a Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinformatics* 26, 689–691. doi:10.1093/bioinformatics/btq007
- Christoffer, C., Chen, S., Bharadwaj, V., Aderinwale, T., Kumar, V., Hormati, M., et al. (2021). LZerD Webserver for Pairwise and Multiple Protein-Protein Docking. *Nucleic Acids Res.* 49, W359, W365. doi:10.1093/nar/gkab336
- Christoffer, C., Terashi, G., Shin, W. H., Aderinwale, T., Maddhuri Venkata Subramaniya, S. R., Peterson, L., et al. (2020). Performance and Enhancement of the LZerD Protein Assembly Pipeline in CAPRI 38–46. *Proteins* 88, 948–961. doi:10.1002/prot.25850
- de Vries, S., and Zacharias, M. (2013). Flexible Docking and Refinement with a Coarse-Grained Protein Model Using ATTRACT. *Proteins* 81, 2167–2174. doi:10.1002/prot.24400
- Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* 125, 1731–1737. doi:10.1021/ja026939x10.3892/or\_00000289
- Esquivel-Rodríguez, J., Yang, Y. D., and Kihara, D. (2012). Multi-LZerD: Multiple Protein Docking for Asymmetric Complexes. *Proteins* 80, 1818–1833. doi:10.1002/prot.24079
- Hanson, J., Paliwal, K., Litfin, T., Yang, Y., and Zhou, Y. (2019). Improving Prediction of Protein Secondary Structure, Backbone Angles, Solvent Accessibility and Contact Numbers by Using Predicted Contact Maps and

## AUTHOR CONTRIBUTIONS

All the authors discussed and conceived the study. CC, VB, and RL implemented the described update of the LZerD web server. CC developed the pipeline of the AttentiveDist protein structure prediction. CC performed the computation of the docking examples. CC wrote the initial draft of the article, and DK and CC edited it.

## FUNDING

This work was partly supported by the National Institutes of Health (R01GM133840 and R01GM123055) and the National Science Foundation (DMS1614777, CMMI1825941, MCB1925643, and DBI2003635). CC is supported by National Institute of General Medical Sciences-funded predoctoral fellowship to CC (T32GM132024).

## ACKNOWLEDGMENTS

The authors are grateful to Information Technology at Purdue, West Lafayette, IN for providing computational resources. The tools covered in this work are available at <https://lzerd.kiharalab.org>.

- an Ensemble of Recurrent and Residual Convolutional Neural Networks. *Bioinformatics* 35, 2403–2410. doi:10.1093/bioinformatics/bty1006
- Huang, S.-Y., and Zou, X. (2014). ITScorePro: an Efficient Scoring Program for Evaluating the Energy Scores of Protein Structures for Structure Prediction. *Methods Mol. Biol.* 1137, 71–81. doi:10.1007/978-1-4939-0366-5\_6
- Jain, A., Terashi, G., Kagaya, Y., Maddhuri Venkata Subramaniya, S. R., Christoffer, C., and Kihara, D. (2021). Analyzing Effect of Quadruple Multiple Sequence Alignments on Deep Learning Based Protein Inter-residue Distance Prediction. *Sci. Rep.* 11, 7574. doi:10.1038/s41598-021-87204-z
- Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. *BMC Bioinformatics* 11, 431. doi:10.1186/1471-2105-11-431
- Kihara, D., Sael, L., Chikhi, R., and Esquivel-Rodríguez, J. (2011). Molecular Surface Representation Using 3D Zernike Descriptors for Protein Shape Comparison and Docking. *Cpps* 12, 520–530. doi:10.2174/138920311796957612
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., et al. (2017). The ClusPro Web Server for Protein-Protein Docking. *Nat. Protoc.* 12, 255–278. doi:10.1038/nprot.2016.169
- Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moutl, J. (2019). Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XIII. *Proteins* 87, 1011–1020. doi:10.1002/prot.25823
- La, D., and Kihara, D. (2012). A Novel Method for Protein-Protein Interaction Site Prediction Using Phylogenetic Substitution Models. *Proteins* 80, 126–141. doi:10.1002/prot.23169
- La, D., Kong, M., Hoffman, W., Choi, Y. I., and Kihara, D. (2013). Predicting Permanent and Transient Protein-Protein Interfaces. *Proteins* 81, 805–818. doi:10.1002/prot.24235
- Lensink, M. F., Brysbaert, G., Nadzirin, N., Velankar, S., Chaleil, R. A. G., Gerguri, T., et al. (2019). Blind Prediction of homo- and Hetero-protein Complexes: The CASP13-CAPRI experiment. *Proteins* 87, 1200–1221. doi:10.1002/prot.25838
- Lensink, M. F., Nadzirin, N., Velankar, S., and Wodak, S. J. (2020). Modeling Protein-protein, Protein-peptide, and Protein-oligosaccharide Complexes: CAPRI 7th Edition. *Proteins* 88, 916–938. doi:10.1002/prot.25870

- Lensink, M. F., Velankar, S., Baek, M., Heo, L., Seok, C., and Wodak, S. J. (2018). The challenge of Modeling Protein Assemblies: the CASP12-CAPRI experiment. *Proteins* 86 (Suppl. 1), 257–273. doi:10.1002/prot.25419
- Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2006). 3D Complex: a Structural Classification of Protein Complexes. *Plos Comput. Biol.* 2, e155. doi:10.1371/journal.pcbi.0020155
- Lyskov, S., and Gray, J. J. (2008). The RosettaDock Server for Local Protein-Protein Docking. *Nucleic Acids Res.* 36, W233–W238. doi:10.1093/nar/gkn21610.1093/nar/gkn216
- Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R., and Weng, Z. (2007). Integrating Statistical Pair Potentials into Protein Complex Prediction. *Proteins* 69, 511–520. doi:10.1002/prot.2150210.1002/prot.21502
- Mirdita, M., von den Driesch, L., Galiez, C., Martin, M. J., Söding, J., and Steinegger, M. (2017). Uniclust Databases of Clustered and Deeply Annotated Protein Sequences and Alignments. *Nucleic Acids Res.* 45, D170–D176. doi:10.1093/nar/gkw1081
- Netzer, R., Listov, D., Lipsh, R., Dym, O., Albeck, S., Knop, O., et al. (2018). Ultrahigh Specificity in a Network of Computationally Designed Protein-Interaction Pairs. *Nat. Commun.* 9, 5286. doi:10.1038/s41467-018-07722-9
- Park, H., Bradley, P., Greisen, P., Jr., Liu, Y., Mulligan, V. K., Kim, D. E., et al. (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theor. Comput.* 12, 6201–6212. doi:10.1021/acs.jctc.6b00819
- Peterson, L. X., Roy, A., Christoffer, C., Terashi, G., and Kihara, D. (2017). Modeling Disordered Protein Interactions from Biophysical Principles. *Plos Comput. Biol.* 13, e1005485. doi:10.1371/journal.pcbi.1005485
- Peterson, L. X., Togawa, Y., Esquivel-Rodriguez, J., Terashi, G., Christoffer, C., Roy, A., et al. (2018). Modeling the Assembly Order of Multimeric Heteroprotein Complexes. *Plos Comput. Biol.* 14, e1005937. doi:10.1371/journal.pcbi.1005937
- Peterson, L. X., Kim, H., Esquivel-Rodriguez, J., Roy, A., Han, X., Shin, W. H., et al. (2017). Human and Server Docking Prediction for CAPRI Round 30-35 Using LZerD with Combined Scoring Functions. *Proteins* 85, 513–527. doi:10.1002/prot.25165
- Peterson, L. X., Shin, W. H., Kim, H., and Kihara, D. (2018). Improved Performance in CAPRI Round 37 Using LZerD Docking and Template-based Modeling with Combined Scoring Functions. *Proteins* 86 (Suppl. 1), 311–320. doi:10.1002/prot.25376
- Ritchie, D. W., and Grudinin, S. (2016). Spherical Polar Fourier Assembly of Protein Complexes with Arbitrary point Group Symmetry. *J. Appl. Cryst.* 49, 158–167. doi:10.1107/s1600576715022931
- Ritchie, D. W., and Venkatraman, V. (2010). Ultra-fast FFT Protein Docking on Graphics Processors. *Bioinformatics* 26, 2398–2405. doi:10.1093/bioinformatics/btq444
- Sanyal, A., Zbornik, E. A., Watson, B. G., Christoffer, C., Ma, J., Kihara, D., et al. (2021). Kinetic and Structural Parameters Governing Fic-Mediated adenylation/AMPylation of the Hsp70 Chaperone, BiP/GRP78, *Cell Stress Chaperones*. 26, 639–656. doi:10.1007/s12192-021-01208-2
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic Acids Res.* 33, W363–W367. doi:10.1093/nar/gki481
- Seemayer, S., Gruber, M., and Söding, J. (2014). CCMpred-Fast and Precise Prediction of Protein Residue-Residue Contacts from Correlated Mutations. *Bioinformatics* 30, 3128–3130. doi:10.1093/bioinformatics/btu500
- Steinegger, M., Meier, M., Mirdita, M., Vohringer, H., Haunsberger, S. J., and Söding, J. (2019). HH-suite3 for Fast Remote Homology Detection and Deep Protein Annotation. *BMC Bioinformatics* 20, 473. doi:10.1186/s12859-019-3019-7
- Steinegger, M., and Söding, J. (2018). Clustering Huge Protein Sequence Sets in Linear Time. *Nat. Commun.* 9, 2542. doi:10.1038/s41467-018-04964-5
- Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., and UniProt, C. (2015). UniRef Clusters: a Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches. *Bioinformatics* 31, 926–932. doi:10.1093/bioinformatics/btu739
- FIWARE (2019). Version 2.3.0. Schrödinger, LLC
- Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J., and Bates, P. A. (2013). SwarmDock: a Server for Flexible Protein-Protein Docking. *Bioinformatics* 29, 807–809. doi:10.1093/bioinformatics/btt038
- Venkatraman, V., Yang, Y. D., Sael, L., and Kihara, D. (2009). Protein-protein Docking Using Region-Based 3D Zernike Descriptors. *BMC Bioinformatics* 10, 407. doi:10.1186/1471-2105-10-407
- Vetting, M. W., Hegde, S. S., Javid-Majd, F., Blanchard, J. S., and Roderick, S. L. (2002). Aminoglycoside 2'-N-Acetyltransferase from Mycobacterium Tuberculosis in Complex with Coenzyme A and Aminoglycoside Substrates. *Nat. Struct. Biol.* 9, 653–658. doi:10.1038/nsb830
- Webb, B., and Sali, A. (2021). Protein Structure Modeling with MODELLER. *Methods Mol. Biol.* 2199, 239–255. doi:10.1007/978-1-0716-0892-0\_14
- Wojdyla, J. A., Fleishman, S. J., Baker, D., and Kleanthous, C. (2012). Structure of the Ultra-high-affinity Colicin E2 DNase-Im2 Complex. *J. Mol. Biol.* 417, 79–94. doi:10.1016/j.jmb.2012.01.019
- Xu, J., and Wang, S. (2019). Analysis of Distance-based Protein Structure Prediction by Deep Learning in CASP13. *Proteins* 87, 1069–1081. doi:10.1002/prot.25810
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc. Natl. Acad. Sci. USA.* 117, 1496–1503. doi:10.1073/pnas.1914677117
- Zhang, C., Zheng, W., Mortuza, S. M., Li, Y., and Zhang, Y. (2020). DeepMSA: Constructing Deep Multiple Sequence Alignment to Improve Contact Prediction and Fold-Recognition for Distant-Homology Proteins. *Bioinformatics* 36, 2105–2112. doi:10.1093/bioinformatics/btz863
- Zheng, W., Li, Y., Zhang, C., Pearce, R., Mortuza, S. M., and Zhang, Y. (2019). Deep-learning Contact-map Guided Protein Structure Prediction in CASP13. *Proteins* 87, 1149–1164. doi:10.1002/prot.25792
- Zhou, H., and Zhou, Y. (2002). Distance-scaled, Finite Ideal-Gas Reference State Improves Structure-Derived Potentials of Mean Force for Structure Selection and Stability Prediction. *Protein Sci.* 11, 2714–2726. doi:10.1110/ps.0217002
- Zhou, H., and Skolnick, J. (2011). GOAP: a Generalized Orientation-dependent, All-Atom Statistical Potential for Protein Structure Prediction. *Biophysical J.* 101, 2043–2052. doi:10.1016/j.bpj.2011.09.012

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Christoffer, Bharadwaj, Luu and Kihara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.