Published in final edited form as:

Methods Mol Biol. 2020 ; 2165: 231–244. doi:10.1007/978-1-0716-0708-4_13.

IDP-LZerD: software for modeling disordered protein interactions

Charles Christoffer¹, Daisuke Kihara^{1,2}

¹Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA

²Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA

Summary

Modeling the tertiary structure of protein-protein interaction complex has been well studied over many years, especially in the case where the structures of both binding partners are roughly the same before and after binding. However, the assembly of complexes with less-ordered partners is a much harder problem, and modeling even small amounts of flexibility can pose a challenge. In an extreme case, where one of the binding partners is intrinsically disordered before binding, we have previously shown that by initially disregarding the coupling between windows of these intrinsically disordered proteins (IDPs), we can reliably assemble complexes involving IDPs up to at least 69 residues long. Here, we detail the use of the IDP-LZerD package and protocol.

Keywords

protein docking; protein-protein docking; intrinsic disorder; intrinsically disordered protein; protein flexibility; LZerD; IDP-LZerD

1. Introduction

Protein-protein docking is a common task in computational structural biology. The problem of rigid-body docking, where subunit conformations are assumed to not change on binding, is well-studied, and many methods with available software exist[1–3]. However, this assumption does not hold for intrinsically disordered proteins, which do not take on a stable conformation until they bind to a receptor. Unlike the rigid-body case, where the number of degrees of freedom remains fixed regardless of the lengths of the proteins involved, the number of degrees of freedom in the disordered case increases with the number of residues. One flexible docking approach has been to use normal mode analysis [4] or other coarse sampling of the conformational space[5]. However, these approximations are severely limited in the types and extent of conformational change they can model. Molecular dynamics force fields have been designed specifically for modeling IDPs [6], but performing an ab initio simulation to assemble a disordered protein-protein interaction from scratch is not computationally feasible. Many docking methods specific to protein-peptide interactions exist [7,8]. However, these methods are generally limited to short peptides, and cannot model longer amino acid chains.

Unlike rigid-body and peptide docking methods, IDP-LZerD [9,10] has been shown to reliably dock disordered proteins up to 69 residues long. IDP-LZerD follows the idea of

the dock-and-coalesce model of IDP binding [11]. In dock-and-coalesce, some segment of the IDP initially binds to the ordered receptor and acts as an anchor. Then, the rest of the IDP coalesces into the bound conformation. While this mechanism can be seen from a biophysical standpoint having thermodynamic and kinetic advantages, it is also computationally advantageous. Initially and in parallel, small segments of the IDP can be docked to the receptor, and then the full IDP chain can be modeled around these. Indeed, this is the approach IDP-LZerD takes. IDP-LZerD models segments of the IDP as rigid fragments, docks them to a receptor structure, selects combinations of consecutive fragments according to geometric constraints and a scoring function, and uses molecular dynamics simulation to convert these into physical models.

IDPs offer many possibilities for practical applications. It has been estimated that 15–45% of all protein-protein interactions (PPIs) involve IDPs, meaning IDPs are often hubs in the PPI network [12]. Moreover, post-translational modifications are more frequent in IDPs, and IDPs are major players in allosteric regulation [13,14]. Thus, the capability of modeling disordered PPIs can inform the design of drugs, whether by facilitating the design of artificial IDPs and peptides, by informing the design of IDP-mimetic drugs, or by simply making it easier to model the effects of a drug that modulates a disordered PPI.

In this chapter, we explain how to use the IDP-LZerD software we developed with three examples.

2. Materials

To dock an IDP, IDP-LZerD takes an IDP sequence and receptor structure as input. IDP fragment structures are first generated, ideally using secondary structure information, and are then docked to the receptor. Fragments are then combined to generate full-sequence structures of the IDP bound to the receptor. These coarse structures are then relaxed into something more physical with a short molecular dynamics simulation. In this section, we detail the materials required to carry out these procedures.

2.1. Input files

To run IDP-LZerD, only the sequence of the IDP and the structure of its receptor are needed as input. The IDP sequence should be formatted in FASTA format. The receptor structure should be formatted in PDB format.

2.2. IDP-LZerD software

Of course, IDP-LZerD itself is required. IDP-LZerD can be downloaded from <http://kiharalab.org/proteindocking/idplzerd.php> or from https://github.com/kiharalab/idp_lzerd.

2.3. Python and Python dependencies

IDP-LZerD is implemented in Python, and requires a Python install with NumPy, SciPy, pandas, apws, and BioPython to run. To install these libraries in an Anaconda environment, simply run:

```
conda install numpy scipy pandas biopython
```

```
conda install -c conda-forge apsw
```

2.4. BLAST software and database

IDP-LZerD uses NCBI BLAST (available from <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [15] as part of the Rosetta fragment picker input in the fragment modeling stage. An installation with both the software and a recent copy of the NR database is necessary.

2.5. Secondary structure prediction software

It is recommended to use secondary structure information in the fragment modeling step of IDP-LZerD to bias the fragment structures. IDP-LZerD can use any secondary structure information given in PSIPRED format, including from secondary structure prediction methods. Some available secondary structure prediction methods include:

- PSIPRED (available at <http://bioinf.cs.ucl.ac.uk/psipred/>) [16]
- JPred (available at <http://www.compbio.dundee.ac.uk/jpred/>) [17]
- Porter (available at <http://distillf.ucd.ie/porterpaleale/>) [18]
- SSPro (available at <http://scratch.proteomics.ics.uci.edu/>) [19]
- DeepCNF/RaptorX Property (available at <http://raptorx.uchicago.edu/StructurePropertyPred/predict/>) [20]
- SPIDER3 (available at <http://sparks-lab.org/server/SPIDER3/>) [21]
- DNSS (available at http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html) [22]

2.6. Fragment modeling software

IDP-LZerD uses the Rosetta fragment picker [23] to model IDP fragments, although any fragment modeling software can in principle be used. The Rosetta fragment picker is available as part of the Rosetta package, available from <https://www.rosettacommons.org>.

2.7. Full-atom reconstruction software

The Rosetta fragment picker does not generate full-atom models, which are necessary for the fragment docking step. IDP-LZerD will generate C-alpha coordinates from the Rosetta fragment picker output, and requires backbone and sidechain modeling software to generate complete models. Pulchra [24] (available from <http://cssb.biology.gatech.edu/PULCHRA>) is recommended for modeling the rest of the backbone. These models should then be input to side-chain modeling software. Side-chain modeling methods are more diverse, but currently OSCAR-star[25] (available from <http://sysimm.ifrec.osaka-u.ac.jp/OSCAR>) and SCWRL4 [26] (available from <http://dunbrack.fccc.edu/scwrl4>) are recommended.

2.8. Protein-protein docking software

A crucial component of IDP-LZerD is docking IDP fragments to the receptor. IDP-LZerD uses our lab's rigid-body docking method, LZerD [3] (available from <http://kiharalab.org/proteindocking/lzerd.php>) to dock each fragment. In principle, however, any protein docking

method, such as ZDOCK [2] (available from <http://zdock.umassmed.edu/software/>) could be used.

2.9. Model scoring software

IDP-LZerD depends on various scoring functions at all stages, including for fragment scoring, path scoring, and final model scoring. From outside the IDP-LZerD package, the statistical scoring functions DFIRE [27], GOAP [28], and ITScorePro [29] are required. DFIRE and GOAP are both calculated by the GOAP package (available from <http://cssb.biology.gatech.edu/GOAP/index.html>). ITScorePro is available from <http://zoulab.dalton.missouri.edu>.

2.10. Molecular dynamics software

The path selection stage of IDP-LZerD only selects which IDP fragments should be included in a path. The atom coordinates, bond lengths, and bond angles at fragment boundaries may not necessarily agree with each other. IDP-LZerD uses CHARMM [30] to knit the fragments together in a physical way. CHARMM is available from <https://www.charmm.org>.

3. Methods

In the general workflow of IDP-LZerD, illustrated in Figure 1, users generate IDP structure fragments, dock the fragments to the receptor, assemble the docked fragments into complete chains, and finally refine the structure. Here, we detail the process of running IDP-LZerD. The programs used and the general flow of data are illustrated in Figure 2.

3.1. Fragment model generation

IDP-LZerD uses the Rosetta fragment picker to generate fragment models for each sequence window. The Rosetta fragment picker can generate models without secondary structure information. However, it is recommended to pass secondary structure information, for example from secondary structure prediction methods, to the fragment picker. To generate the fragments, simply run, for example:

```
run_rosetta.py $PDBID -l $L_CH --psipred $PSIPRED_FILE --porter  
$PORTER_FILE --jpred $JPRED_FILE --sspro $SSPRO_FILE -d.
```

3.2. Fragment-receptor docking

Once fragment models have been generated for each sequence window, they must be docked to the receptor structure. IDP-LZerD uses LZerD for docking, but in principle, any docking program can be used. For each fragment, simply run LZerD as in the LZerD distribution.

To use another docking program, simply pass the fragment PDB file and the receptor PDB file to the program like any other structures. Then, generate or rename output PDB files to have names like model1.pdb, model2.pdb, etc. For example, to rename ZDOCK models, which have file names like complex.1, complex.2, etc., you can use the Perl rename command like:

```
rename "s/complex\./model/g; s/$/.pdb/g"
```

When using LZerD models, the top 4500 by DI score (see Note 1) should be retained.

3.3. Fragment assembly

To generate full-chain IDP structures, IDP-LZerD must select which docked fragment models to combine. IDP-LZerD greedily assembles fragments into paths. The set of paths is initialized with the fragments of the first sequence window. Then, each fragment of the next window is added to each path in the set, except those that would violate the heuristic geometric constraints (see Note 2). The path set is then clustered to remove redundant paths. To perform fragment assembly, the following commands should be run in the top-level fragment directory:

```
python $PACKAGE_DIR/scripts/create_database.py -p $PDBID -i "$  
{PDBID}_data.csv" -d.  
  
mv "scores_${PDBID}.db" "scores_${complexid}.db"  
  
python $PACKAGE_DIR/scripts/load_model_scores.py -b "scores_${complexid}.db"  
-p $PDBID -r $R_CH -l $L_CH -d.  
  
python $PACKAGE_DIR/scripts/find_paths_stepwise.py $complexid -n  
$N_WINDOWS -d.
```

where the environment variables have been set as: `PACKAGE_DIR` to the location of the IDP-LZerD distribution, `PDBID` to the ID of the structure being modeled, `R_CH` to the receptor chain IDs, `L_CH` to the IDP chain ID, `complexid` to `{PDBID}${R_CH}${L_CH}`, and `N_WINDOWS` to the number of sequence windows. The output structures will be in the subdirectories of `{complexid}${N_WINDOWS}`.

3.4. Model refinement

IDP-LZerD output structures are complete but typically have severe bond length and angle issues. To remedy this, the structures are refined using CHARMM, or any other molecular dynamics software that the protocol can be translated to. The IDP-LZerD distribution comes with a CHARMM input file, `idp_relax.inp`, which contains the entire relaxation protocol. To use the protocol, simply replace the fields in braces in the lines

```
set charmmsredir {charmm_dir}  
  
set dir {working_dir}  
  
set nam {filename}  
  
set rch {receptor_chain}  
  
set lch {ligand_chain}
```

with the relevant information. For example:

```
set charmmsredir /path/to/charmm/data  
  
set dir /path/to/working/dir  
  
set nam /filename/to/refine
```

set rch X

set lch Y

After the models have been refined, the model score can be calculated and the top models selected.

3.5. Model scoring

IDP-LZerD uses a combined scoring function optimized for disordered PPIs to score the refined models. This model score is defined as the weighted sum of Z-scores from ITScorePro, DFIRE, GOAP, and our lab's molecular mechanics score:

$$S_{model} = w_5(w_1 Z(ITScorePro) + w_2 Z(DFIRE) + w_3 Z(MolMech) + w_4 Z(GOAP)) + (1 - w_5) \min\{Z(ITScorePro), Z(DFIRE), Z(MolMech), Z(GOAP)\}$$

The ranking of the refined models by this score is the final output of IDP-LZerD.

4. Case Studies

Here we provide three examples where IDP-LZerD successfully models the structure of disordered PPIs.

4.1. MHC class II antigen DRA/DRB5 with myelin basic protein

In this example, we model the formation of a complex between human MHC class II antigen DRA/DRB5 and a 20-residue segment of the IDP human myelin basic protein. This protein complex is relevant to multiple sclerosis in humans [31]. PDB ID 1FV1 is a bound structure of this complex, but we use the unbound receptor structure found in PDB ID 4AH2 with engineered residues removed for modeling. Using for example JPRED, the predicted secondary structure of the bound myelin basic protein is:

PSIPRED VFORMAT (PSIPRED V3.3)

1. N C 0.963 0.012 0.018
2. P C 0.912 0.032 0.022
3. V H 0.200 0.475 0.075
4. V H 0.092 0.621 0.123
5. H H 0.073 0.707 0.093
6. F H 0.094 0.766 0.065
7. F H 0.117 0.782 0.054
8. K H 0.192 0.755 0.051
9. N H 0.319 0.624 0.065
10. I H 0.585 0.418 0.043
11. V C 0.736 0.225 0.055

12. T C 0.830 0.129 0.054
 13. P C 0.816 0.095 0.081
 14. R C 0.814 0.095 0.085
 15. T C 0.818 0.078 0.089
 16. P C 0.840 0.069 0.079
 17. P C 0.862 0.053 0.071
 18. P C 0.898 0.036 0.059
 19. S C 0.924 0.023 0.051
 20. Q C 0.936 0.017 0.044

The first two columns in this format are the sequence position and residue type, respectively. The third column is the predicted secondary structure class for that residue. IDP-LZerD uses the three-way classification of secondary structures into coil (C), helix (H), and sheet (E). The next three columns are the predicted probabilities that the residue is coil, helix, or sheet, respectively. When using a secondary structure prediction method that does not output probabilities for each class, IDP-LZerD assumes the predicted class has predicted probability 0.67, and assumes the other two classes have a predicted probability of 0.15 each.

The final result of IDP-LZerD for this target is illustrated in Figure 3. The rank 1 model shown is of CAPRI-acceptable quality [32], with an fnat (fraction of the native residue contacts at the docking interface) of 0.40, an I-RMSD (the root mean square deviation at the docking interface) of 2.5 Å, and an L-RMSD (RMSD of the ligand protein that is computed after the native and the modelled receptor proteins are superimposed) of 5.8 Å. Not shown are the other top 10 models; altogether, 7 of the top 10 output models are of acceptable quality.

4.2. Proteinase A with protease A inhibitor 3

In this example, we model the formation of a complex between *S. cerevisiae* proteinase A and a 31-residue segment of *S. cerevisiae* protease A inhibitor 3. Aspartic proteinases are relevant to a number of human pathologies, including hypertension [33]. PDB ID 1G0V is a bound structure of this complex, but we use the unbound receptor structure found in PDB ID 1FMX for modeling. Using for example Porter, the predicted secondary structure is:

PSIPRED VFORMAT (PSIPRED V3.3)

1. M C 0.670 0.150 0.150
 2. N C 0.670 0.150 0.150
 3. T C 0.670 0.150 0.150
 4. D C 0.670 0.150 0.150
 5. Q H 0.150 0.670 0.150
 6. Q H 0.150 0.670 0.150

7. K H 0.150 0.670 0.150
8. V H 0.150 0.670 0.150
9. S H 0.150 0.670 0.150
10. E H 0.150 0.670 0.150
11. I H 0.150 0.670 0.150
12. F H 0.150 0.670 0.150
13. Q H 0.150 0.670 0.150
14. S H 0.150 0.670 0.150
15. S H 0.150 0.670 0.150
16. K H 0.150 0.670 0.150
17. E H 0.150 0.670 0.150
18. K H 0.150 0.670 0.150
19. L H 0.150 0.670 0.150
20. Q C 0.670 0.150 0.150
21. G C 0.670 0.150 0.150
22. D C 0.670 0.150 0.150
23. A H 0.150 0.670 0.150
24. M H 0.150 0.670 0.150
25. V H 0.150 0.670 0.150
26. V H 0.150 0.670 0.150
27. S H 0.150 0.670 0.150
28. D H 0.150 0.670 0.150
29. A C 0.670 0.150 0.150
30. F C 0.670 0.150 0.150
31. K C 0.670 0.150 0.150

As shown in Figure 4, the native conformation of the IDP in this complex has entirely helical secondary structure. The rank 1 model from the final output of IDP-LZerD follows the mostly-helical secondary structure prediction and agrees well with the native structure. In fact, this model is CAPRI-acceptable, with an fnat of 0.29, an I-RMSD of 4.8 Å, and an L-RMSD of 5.4 Å. Not shown in the figure are the other top 10 models; 6 of the top 10 models are of at least acceptable quality, and 4 of these are of medium quality. If confident secondary structure information for your target disordered complex is available, it is strongly recommended that it be included in your IDP-LZerD input.

4.3. Outer membrane virulence protein YopE with YopE chaperone SycE

In this last example, we model the formation of a complex between *Y. Pseudotuberculosis* outer membrane virulence protein YopE and a 69-residue segment of *Y. Pseudotuberculosis* YopE chaperone SycE, relevant to a large number of bacterial pathogens [34]. This example represents the extreme end of what lengths of IDP can currently be reliably modeled, and illustrates why IDP-LZerD currently represents the state of the art. PDB ID 1L2W is a bound structure of this complex, but we use the unbound receptor structure from PDB ID 1JYA for modeling. As shown in Figure 5, when SycE binds to YopE, it wraps most of the way around YopE, and its conformation contains many nonlinearities. The rank 2 model from IDP-LZerD shown in Figure 5 is CAPRI-acceptable with an fnat of 0.21, an I-RMSD of 5.4 Å, and an L-RMSD of 10.0 Å.

5. Notes

5.1. DI score

The DI score is defined as the sum of the Z-scores of the DFIRE and ITScorePro scores. See Supplemental Table S3 in the original paper[9].

5.2. Fragment pair geometric constraints

Any fragment pair A, B is removed from consideration in the path search if any of the following are true:

- Any atom of A is closer than 3 Å to any atom of B
- A and B are adjacent, and their midpoints are closer than 6.5 Å
- The midpoints of A and B are closer than 3.8 Å
- The midpoints of A and B are further apart than 18.5 Å times the distance between their sequence windows
- A and B are adjacent, and their overlap distance is not between 5.2 Å and 13.6 Å
- A and B are adjacent, and all their overlap pair distances are greater than 6 Å
- A and B are adjacent, and any of their overlap pair distances is greater than 10 Å
- A and B are adjacent, and their overlap angle does not satisfy $\cos(\theta) \geq 0.1$

Figure 6 illustrates some of these constraints.

Acknowledgments

This work was supported by the National Institutes of Health (R01GM123055). DK also acknowledges support from the National Science Foundation (DMS1614777, CMMI1825941). CC is supported by NIGMS-funded predoctoral fellowship (T32 GM132024).

References

1. Ritchie DW, Kemp GJ (2000) Protein docking using spherical polar Fourier correlations. *Proteins* 39 (2):178–194 [PubMed: 10737939]

2. Pierce BG, Hourai Y, Weng Z (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* 6 (9):e24657. doi:10.1371/journal.pone.0024657 [PubMed: 21949741]
3. Venkatraman V, Yang YD, Sael L, Kihara D (2009) Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics* 10:407. doi:10.1186/1471-2105-10-407 [PubMed: 20003235]
4. Moal IH, Bates PA (2010) SwarmDock and the use of normal modes in protein-protein docking. *Int J Mol Sci* 11 (10):3623–3648. doi:10.3390/ijms11103623 [PubMed: 21152290]
5. Marze NA, Roy Burman SS, Sheffler W, Gray JJ (2018) Efficient flexible backbone protein-protein docking for challenging targets. *Bioinformatics* 34 (20):3461–3469. doi:10.1093/bioinformatics/bty355 [PubMed: 29718115]
6. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14 (1):71–73. doi:10.1038/nmeth.4067 [PubMed: 27819658]
7. Lee H, Heo L, Lee MS, Seok C (2015) GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res* 43 (W1):W431–435. doi:10.1093/nar/gkv495 [PubMed: 25969449]
8. Xu X, Yan C, Zou X (2018) MDockPeP: An ab-initio protein-peptide docking server. *J Comput Chem* 39 (28):2409–2413. doi:10.1002/jcc.25555 [PubMed: 30368849]
9. Peterson LX, Roy A, Christoffer C, Terashi G, Kihara D (2017) Modeling disordered protein interactions from biophysical principles. *PLoS Comput Biol* 13 (4):e1005485. doi:10.1371/journal.pcbi.1005485 [PubMed: 28394890]
10. Christoffer C, Kihara D (2019, in press) Modeling protein-protein interactions with intrinsically disordered proteins. In: Salvi N (ed) *Intrinsically Disordered Proteins*. Elsevier,
11. Zhou HX (2012) Intrinsic disorder: signaling via highly specific but short-lived association. *Trends Biochem Sci* 37 (2):43–48. doi:10.1016/j.tibs.2011.11.002 [PubMed: 22154231]
12. Petsalaki E, Russell RB (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr Opin Biotechnol* 19 (4):344–350. doi:10.1016/j.copbio.2008.06.004 [PubMed: 18602004]
13. Zhou J, Zhao S, Dunker AK (2018) Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J Mol Biol* 430 (16):2342–2359. doi:10.1016/j.jmb.2018.03.028 [PubMed: 29626537]
14. Darling AL, Uversky VN (2018) Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* 9:158. doi:10.3389/fgene.2018.00158 [PubMed: 29780404]
15. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25 (17):3389–3402. doi:10.1093/nar/25.17.3389 [PubMed: 9254694]
16. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292 (2):195–202. doi:10.1006/jmbi.1999.3091 [PubMed: 10493868]
17. Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43 (W1):W389–394. doi:10.1093/nar/gkv332 [PubMed: 25883141]
18. Mirabello C, Pollastri G (2013) Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* 29 (16):2056–2058. doi:10.1093/bioinformatics/btt344 [PubMed: 23772049]
19. Magnan CN, Baldi P (2014) SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* 30 (18):2592–2597. doi:10.1093/bioinformatics/btu352 [PubMed: 24860169]
20. Wang S, Peng J, Ma J, Xu J (2016) Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Sci Rep* 6:18962. doi:10.1038/srep18962 [PubMed: 26752681]
21. Heffernan R, Dehzangi A, Lyons J, Paliwal K, Sharma A, Wang J, Sattar A, Zhou Y, Yang Y (2016) Highly accurate sequence-based prediction of half-sphere exposures of amino acid

- residues in proteins. *Bioinformatics* 32 (6):843–849. doi:10.1093/bioinformatics/btv665 [PubMed: 26568622]
22. Spencer M, Eickholt J, Jianlin C (2015) A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction. *IEEE/ACM Trans Comput Biol Bioinform* 12 (1):103–112. doi:10.1109/TCBB.2014.2343960 [PubMed: 25750595]
23. Gront D, Kulp DW, Vernon RM, Strauss CE, Baker D (2011) Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One* 6 (8):e23294. doi:10.1371/journal.pone.0023294 [PubMed: 21887241]
24. Rotkiewicz P, Skolnick J (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem* 29 (9):1460–1465. doi:10.1002/jcc.20906 [PubMed: 18196502]
25. Liang S, Zheng D, Zhang C, Standley DM (2011) Fast and accurate prediction of protein side-chain conformations. *Bioinformatics* 27 (20):2913–2914. doi:10.1093/bioinformatics/btr482 [PubMed: 21873640]
26. Krivov GG, Shapovalov MV, Dunbrack RL Jr. (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77 (4):778–795. doi:10.1002/prot.22488 [PubMed: 19603484]
27. Zhou H, 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 (11):2714–2726. doi:10.1110/ps.0217002 [PubMed: 12381853]
28. Zhou H, Skolnick J (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys J* 101 (8):2043–2052. doi:10.1016/j.bpj.2011.09.012 [PubMed: 22004759]
29. Huang SY, Zou X (2014) ITScoPro: 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 [PubMed: 24573475]
30. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 4 (2):187–217. doi:10.1002/jcc.540040211
31. Li Y, Li H, Martin R, Mariuzza RA (2000) Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. *J Mol Biol* 304 (2):177–188. doi:10.1006/jmbi.2000.4198 [PubMed: 11080454]
32. Mendez R, Leplae R, De Maria L, Wodak SJ (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* 52 (1):51–67. doi:10.1002/prot.10393 [PubMed: 12784368]
33. Phylip LH, Lees WE, Brownsey BG, Bur D, Dunn BM, Winther JR, Gustchina A, Li M, Copeland T, Wlodawer A, Kay J (2001) The potency and specificity of the interaction between the IA3 inhibitor and its target aspartic proteinase from *Saccharomyces cerevisiae*. *J Biol Chem* 276 (3):2023–2030. doi:10.1074/jbc.M008520200 [PubMed: 11042188]
34. Birtalan SC, Phillips RM, Ghosh P (2002) Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol Cell* 9 (5):971–980 [PubMed: 12049734]

Figure 1.

Illustration of the overall flow of IDP-LZerD.

IDP-LZerD consists of four steps. 1. fragment structure prediction, 2. fragment docking, 3. path assembly, and 4. refinement. Steps 1 and 2 correspond to “dock” and steps 3 and 4 correspond to “coalesce”. This figure is taken from Figure 1 of the original IDP-LZerD paper [9] (available under the Creative Commons CC0 public domain dedication).

Figure 2.
Programs and input data to use for each step in IDP-LZerD.

Figure 3.

Model of MHC class II antigen DRA/DRB5 and a 20-residue segment of the IDP human myelin basic protein with superimposed native structure.

Gray: unbound MHC class II antigen DRA/DRB5 from PDB 4AH2. Orange: rank 1 docked model of myelin basic protein from IDP-LZerD (CAPRI-acceptable quality with an fnat of 0.40, an I-RMSD of 2.5 Å, and an L-RMSD of 5.8 Å). Green: native structure of myelin basic protein from PDB 1FV1.

Figure 4.

Model of proteinase A and a 31-residue segment of protease A inhibitor 3 with superimposed native structure.

Gray: unbound proteinase A from PDB 1FMX. Orange: rank 1 docked model of protease A inhibitor 3 from IDP-LZerD (CAPRI-acceptable quality with an fnat of 0.29, an I-RMSD of 4.8 Å, and an L-RMSD of 5.4 Å). Green: native structure of protease A inhibitor 3 from PDB 1G0V.

Figure 5.

Model of *Y. Pseudotuberculosis* outer membrane virulence protein YopE and a 69-residue segment of *Y. Pseudotuberculosis* YopE chaperone SycE.

Gray: unbound outer membrane virulence protein YopE from PDB 1JYA. Orange: rank 2 docked model of YopE chaperone SycE from IDP-LZerD (CAPRI-acceptable with a fnat of 0.21, an I-RMSD of 5.4 Å, and an L-RMSD of 10.0 Å.). Green: native structure of YopE chaperone SycE from PDB 1L2W.

Figure 6.

Fragment geometry subject to cutoffs. Midpoint distance: between the C_α atoms of the middle residues of two fragments; overlap distance: between the C_α atoms of the residues before and after the overlapping residues; overlap pair distance: between the corresponding N , C_α , C_β and C atoms of the three overlapping residues; overlap angle: formed by the vectors from the N atom of the first overlapping residue to the C atom of the third overlapping residue. This figure is taken from Figure 8 of the original IDP-LZerD paper [9] (available under the Creative Commons CC0 public domain dedication).