Methods for molecular modelling of protein complexes

Tejashree Rajaram Kanitkar, Neeladri Sen, Sanjana Nair, Neelesh Soni, Kaustubh Amritkar, Yogendra Ramtirtha, and M.S. Madhusudhan*

Indian Institute of Science Education and Research Pune,
Dr Homi Bhabha Road, Pashan, Pune 411008, India

* To whom the correspondence should be made

madhusudhan@iiserpune.ac.in

Abstract

Biological processes are often mediated by complexes formed between proteins and various biomolecules. The 3D structures of such protein-biomolecule complexes provide insights into the molecular mechanism of their action. The structure of these complexes can be predicted by various computational methods. Choosing an appropriate method for modelling depends on the category of biomolecule that a protein interacts with and the availability of structural information about the protein and its interacting partner. We intend for the contents of this chapter to serve as a guide as to what software would be the most appropriate for the type of data at hand and the kind of 3D complex structure required. Particularly, we have dealt with protein-small molecule ligand, protein-peptide, protein-protein, and protein-nucleic acid interactions.

Most, if not all, model building protocols perform some sampling and scoring. Typically, several alternate conformations and configurations of the interactors are sampled. Each such sample is then scored for optimization. To boost the confidence in these predicted models, their assessment using other independent scoring schemes besides the inbuilt/default ones would prove to be helpful. This chapter also lists such software and serves as a guide to gauge the fidelity of modelled structures of biomolecular complexes.

Introduction

All biological processes are mediated by various molecular interactions. These include interactions between protein and protein, protein and small molecule ligands, protein and DNA, etc. Characterizing these interactions is essential for gaining biological insights. Experimental characterization is often cumbersome, expensive, and/or difficult to perform. Computational methods [1–4] are hence routinely used to model the 3D structures of the complexes resulting from such interactions.

The computational methods fall into two broad categories – a) Those that exploit information from a related or homologous template structure (henceforth referred to as template based methods) or b) methods that attempt to model the 3D structures of complexes without any pre-determined structural bias. Such methods are often referred to as *ab initio* or as template-free algorithms, include the various docking programs. Many contemporary algorithms make use of a hybrid of methods a) and b) to predict the structure of the interacting 3D complex. Figure 1 shows a spreadsheet of many such methods.

Most computational methods employ similar protocols for predicting the structures of the complexes - viz, sampling different conformations and then evaluating/scoring them to find the most optimal mode of association. Each of the algorithms differs in strategies they use for these sampling and scoring steps [5]

This chapter is written to serve as a practical guide to model complexes of 1) protein-small molecule ligands, 2) protein-peptide, 3) protein-protein, 4) protein-nucleic acid (DNA/RNA), and 5) macromolecular assemblies. In each subsection, one or a few representative methods are highlighted while some information is provided about alternate techniques. The choice of representative method has been based on our familiarity, the ease of access (with a preference for freeware) and overall popularity. We believe that once predictions have

been made their assessment is crucial in deciding their benefit or applicability and we list a few such software that can be used for assessment.

We believe that the relevance of this chapter is enhanced given the current circumstances, when there is an all-out effort to discover or design therapeutic agents and vaccines against SARS CoV2.

1. Modelling protein-small molecule complexes

Modelling protein-small molecule complexes is important for a wide range of applications from gaining insights into processes such as metabolism to designing therapeutics. While naturally occurring small molecules (~50-1500 daltons [6]) are integral components of metabolic and sensory pathways[7], synthetic small molecules (>500 daltons) find applications in designing therapeutic agents.

We envisage two different situations that would warrant the need for modelling protein-small molecule ligand complexes – a) to find a suitable small molecule ligand for a given target protein, and b) to find protein targets of a given small molecule. In both cases, we would also want to find the exact binding pose of a small molecule onto a particular target protein. The sections below cover the situation (a) in some length along with an illustrative example of finding suitable small molecule inhibitors to the Nipah virus glycoprotein[8]. The issues discussed in sections 1.2 -1.4 below are also applicable to the situation (b).

1.1 Selecting the small molecule library

When searching for putative binding small molecule ligands of given target proteins, it is essential to utilize a screening library. Two such popular libraries are PubChem[9] and ZINC [10]. PubChem hosts ~103 million chemical compounds annotated by physical and chemical properties, biological activities, toxicity, etc. One can create appropriate subsets based on the desired properties of the small molecules. The ZINC database hosts ~230 million commercially available compounds categorized into pre-created subsets such as FDA approved drugs, derivatives of natural products and so on. The compounds in the ZINC database are also available in docking friendly file formats. User defined subsets based on physical and/or biological properties can also be easily created.

For our example of finding an appropriate inhibitor to the Nipah glycoprotein, we selected the ZINC12 clean drug-like subset. More on this in section 1.2 to 1.4

Small molecules can also be selected from various other online libraries such as DrugBank, ChEMBL, ChemSpider, KEGG, ChEBI and Ligand Depot [11–16].

1.2 Predicting small molecule binding pockets on the target protein.

Many docking software that attempt to predict/build the complexes of proteins with their small molecule ligands often scan the entire protein surface for suitable binding pockets for the ligands. This exercise makes screening a large number of compounds computationally expensive and time consuming. This problem can be circumvented by localizing potential small molecule binding sites and then having the software scan these sites to conformationally optimize the protein-ligand complex.

A small molecule binding pocket is a cavity on or inside the protein that can potentially harbour a ligand [17]. Several methods such as ProBiS-CHARMMing, 3DLigandSite, PrankWeb, PockDrug-Server are among others that predict the binding pockets given a 3D structure of a protein predict the binding pockets given a 3D structure of a protein [17–27]. DEPTH (http://cospi.iiserpune.ac.in/depth/htdocs/index.html) is one such method that uses the depth of amino acid residues along with the evolutionary information to predict putative binding pockets. The DEPTH server takes 3D structure of a protein as input (See Note 1 and 2) and assigns probability scores to each of the amino acids to be a part of a binding pocket. A user tuneable cut-off score can be used to select binding pockets. These predicted binding pockets can then be used as an input to docking programs.

For instance, DEPTH predicts two binding pockets on the surface of Nipah glycoprotein (PDB ID: 3D11). Interestingly, one of the predicted pockets overlaps with the region where

the glycoprotein interacts with host cells proteins. Each of these pockets can be used for docking.

1.3 Docking small molecules on a target protein: Sampling the ligand conformation and scoring

Molecular docking, similar to other computational procedures, involves a sampling and scoring protocol. There are various sampling schemes such as the systematic incremental approach[29], shape based sampling[30], genetic algorithms[31], fragment based approaches[32] and Monte Carlo simulations[33]. The sampling generates various conformations of the small molecules called poses that are evaluated by a scoring scheme. The scoring includes physics based scoring schemes, empirical scoring functions or knowledge based potentials[34].

Autodock[35] is one of several popular docking programs (Refer Table1 for other docking methods) that uses a Lamarckian genetic algorithm for sampling conformations. A semi-empirical free energy force field is used to predict the binding free energy. Binding poses of a small molecule can be sampled on the entire protein surface or it can be restricted to binding pockets (such as the 2 pockets predicted by DEPTH for Nipah glycoprotein). Along with exploring the poses of the small molecule, protein side chain conformations can also be sampled to account for their flexibility (flexible docking). The tutorial http://autodock.scripps.edu/faqs-help/tutorial/using-autodock-4-with-autodocktools/2012_ADTtut.pdf describes the docking procedure elaborately.

1.4 Shortlisting the compounds

One method of selecting potential small molecule ligands is based on the energy values of their docked poses. If a small molecule is already known to bind a given pocket (control molecule), the score of its complex with the protein can be used as a cut-off to shortlist other ligands. All the complexes where small molecules are docked at this pocket that have energy

better than the cut-off can be considered as potential binding molecules. In cases where control molecules are unknown, shortlisting the ligands is challenging and a consensus of more than one docking method can be employed to shortlist the ligands. The intersection of the top ‘N’ best scoring ligands from various docking software can be further subjected to the structural superimposition of the protein to calculate the ligand RMSD between poses predicted by different docking tools. All compounds that have ligand RMSD better than a preset threshold can be shortlisted for further validation. Such a jury approach ensures predictions with increased confidence[8, 36]

For the Nipah glycoprotein, a subset of small molecules from the ZINC12 database was scanned on the DEPTH predicted binding sites using two docking software, Dock and AutoDock. 9 putative ligands were identified from the top scoring 150 molecules that overlap between Dock and Autodock runs. Such small molecules can then be experimentally tested to confirm their inhibitory activity.

An alternative to docking for finding the exact binding pose of a particular small molecule onto a given target protein, is searching a structural database for regions of geometric and physico-chemical similarity of the binding pocket [37].

2. Modelling protein-peptide complexes

Several proteins such as MHCs, membrane proteins, etc, interact with peptides[38, 39]. Such interactions are estimated to account for 15% to 40% of known protein-protein interactions[40]. Because peptides are usually associated with low levels of toxicity and are easy to synthesize [41], they make for attractive therapeutic agents[42]. In this section, we explore the different approaches for modelling protein-peptide complexes.

2.1 Predicting binding sites for peptide ligands

Similar to the modelling approaches described in protein-small molecule ligand modelling (Refer section 1), some protein-peptide complex modelling methods require the binding site information. ACCLUSTER (http://zougrouptoolkit.missouri.edu/accluster) [43] is one of the several software[44–46] that can be used to predict peptide binding sites on the surface of a given protein(See Note 3). ACCLUSTER uses the standard 20 amino acids as probes to detect the poses that form stable chemical interactions with the protein surface. These poses are spatially clustered, and the largest clusters are predicted as potential binding sites. We tested the ability of ACCLUSTER to predict the peptide binding sites on HLA-B27 major histocompatibility complex that is known to bind to antigenic peptides. Starting with a crystal structure of HLA-B27 (PDB id 6PYL), without its peptide ligand, the true antigenic peptide binding site was one of the predictions.

2.2 Modelling protein-peptide complexes

As with most of the methods that deal with modelling complexes, the input here is the known 3D structure of the target protein. The method of choice would depend on the information available about the peptide. If the structure and sequence of the peptide is not known, the structure of protein-peptide complex can be predicted using tools such as SPOT peptide(Section 2.2.1) [47]. If the sequence of the binding peptide is known, the 3D structure of the complex can be modelled using tools such as GalaxyPepDock(Section 2.2.2.1) [48], Rosetta FlexPepDock (Section 2.2.2.2)[49], and HADDOCK [50] (Refer table 2 for various methods of protein-peptide complex modelling).

2.2.1 Predicting the sequence of the peptides and the structure of the protein-peptide complex

The methods in this category fall into two classes, a) knowledge-based [51, 52] and b) *de novo*[53–55]. Knowledge based methods make use of known structural information to predict the structure. The *de novo* methods, however, are independent of the known structural information and generally make use of physics based principles to predict the structure of the complex. In this section, we describe a prediction of the peptide that is most likely to bind histone transferase (Histone-lysineN-methyl transferase 2A) using the knowledge based method, SPOT-Peptide (http://sparks-lab.org/tom/SPOT-peptide)[47]. The 3D structure of the histone transferase is used as the target protein. SPOT-Peptide superimposes this target protein on a library of known peptide binding proteins to identify suitable templates. Models are built using these templates and are assessed using DFIRE[56] and evolutionary alignment score. The models that are favourably scored from either of the scoring schemes are then filtered by a score based on template similarity, SP-score to get final predicted models.

All the predictions are associated with the three assessment scores and a list of residues of the protein that interact with the peptide. SPOT-peptide was able to successfully reproduce the transferase and peptide complex as one of the top predictions. The predicted complex is comparable to the crystal structure, with a peptide backbone RMSD of ~2.5Å.

2.2.2 Docking peptides onto target proteins

Docking a given peptide onto a protein can be guided by a template. Template based methods rely on structures of homologous complexes to model the 3D structure. If homologous templates are unavailable, template independent docking algorithms are employed [56].

2.2.2.1 Template based docking of protein-peptide complexes

The GalaxyPepDock server (http://galaxy.seoklab.org/pepdock) [57] is a template based docking program that uses structural similarity of the protein and sequence similarity of the peptide to identify the templates. To predict the complex of ubiquitin Nedd4 with the peptide PPXY (a motif of arrestin-related domain-containing protein-3), GalaxyPepDock takes a 3D structure of the ubiquitin and the sequence PPXY as inputs. Multiple models are generated by GalaxyTBM[58, 59] for each homologous template identified by structural and interaction similarity. Top 10 best energy models for each template are refined by energy based optimization and are presented as final predicted models. The predicted complexes are associated with details such as templates used for protein and peptide, sequence alignments, structure similarity score, interaction similarity score, accuracy and the residues on the protein predicted to interact with the peptide. The predicted model for ubiquitin and motif peptide complex (excluding the known crystal structure template), was built using a template with high structural similarity assessed by a metric called TM-score [60].

2.2.2.2 Local docking of protein-peptide complexes

Given a peptide sequence and a protein structure on which a binding pocket has been identified (Refer to section 2.1), local docking can be used to predict the 3D structure of the complex. One such method is Rosetta FlexPepDock (http://flexpepdock.furmanlab.cs.huji.ac.il/) [49, 61]. The input is an approximate protein-peptide complex (See Note4) where the peptide is placed near the binding pocket. The initial complex can be built using standard homology modelling tools. If the structure of homologs is not available, an initial peptide conformation can be manually constructed and placed in the vicinity of the binding site using tools such as Chimera[62]. Rosetta FlexPepDock refines the initial complex structure in 200 independent FlexPepDock simulations. 100 of these are

performed in a high-resolution mode, whereas, the other 100 are performed with a low-resolution pre-optimization followed by a high-resolution refinement step. These are then ranked according to the rosetta full-atom energy score. Ten best scoring complexes are presented as final predictions.

Along with the initial approximate model, atomic constraints, if known, can also be provided. To better assess the predicted structure, a reference structure can be used as a comparison standard. The reference structure is often a structure of a similar interaction and is used to calculate RMSDs of the predicted complex. If the reference structure is not given as an input, RMSDs are calculated with respect to the starting conformation(input protein-peptide complex). Users can select the representative atoms for RMSD calculation, the default selection is peptide backbone heavy atoms.

2.2.2.3 Blind docking of protein-peptide complexes

When little or nothing is known about the peptide binding site or the peptide conformation, we can take recourse to blind docking. The software AnchorDock performs blind docking by employing a variation of molecular dynamics (MD) simulations[62, 63]. The inputs are the structures of the target protein and a peptide with an extended initial conformation. A free peptide folding simulation is performed with explicit solvent to get a peptide conformation for docking. It localizes the conformational space by identifying the most probable peptide binding regions on the surface of the protein called anchoring spots using ANCHORSMAP[64]. Once the anchoring spots are identified, an anchor-driven simulated annealing simulation is applied to the free peptide conformation around the anchoring spots. The simulation trajectories are clustered based on backbone RMSD and ranked based on the average potential energy of the system to get the final protein-peptide complexes (the one with the least energy). Refer to https://link.springer.com/protocol/10.1007/978-1-4939-6798-8_7 for a detailed protocol [65].

2.3 Assessing predicted models with various scoring schemes

The modelled complexes can then be assessed by various protein-peptide complex scoring schemes such as the FoldX suite, that computes the interaction energy. In principle all the protein-protein assessment scores can also be used here. For more details on this, refer to section 3.3)

3. Modelling protein-protein complexes

Only ~6% of all estimated protein interactions have experimentally solved structures in the PDB leaving a substantial number of them structurally uncharacterized[65]. Computational methods can aid in modelling these uncharacterized structures. Similar to protein-peptide complex modelling methods, the protein-protein complex modelling methods have two broad categories i.e., template based prediction and docking. The section below describes some of these methods for dimeric complexes. Modelling of multimeric interactions is covered in section 5.

3.1.1 Template based prediction of structure of a protein – protein complex given structures of the target proteins

PRISM [66](http://cosbi.ku.edu.tr/prism/) is one of the several (Refer table 3) template based docking programs that predict the structure of the complex when the structures of both the target proteins are known. PRISM was used to model the falcipain-cystatin complex [66]. Falcipain is a cysteine protease that is inhibited by cystatin. PRISM takes the structures of the targets falcipain and cystatin as inputs. The surface of the targets is then scanned through a library of known protein-protein interfaces to identify a template interface (based on the structural match). Models are built using the identified template interface and are

assessed using an energy function, FiberDock[67]. The lower the energy, the better is the model. PRISM built the best scoring model for the falcipain-cystatin complex using a template of Cathepsin B (a cysteine protease) and stefin A (inhibitor of cysteine protease) complex. The predicted model had the binding regions and relative orientation of the two proteins similar to that of the native falcipain-cystatin complex (PDB ID:1YVB) [68].

In addition to the structure of targets, the template interface, if known, can also be provided as an input. PRISM will then only sample over the specified interface instead of the entire library of protein-protein interfaces.

3.1.2 Template based prediction of structure of a protein complex when the structures of the constituent target proteins are not known

To predict the structure of the protein-protein complex, where the structure of the two (or more) target proteins and that of their complex is unknown, template based prediction methods (Refer table 3) can be used. HOMCOS (http://homcos.pdbj.org/) [69, 70] is one such method that is based on dimeric threading. It uses homologous dimeric templates to predict structures of complexes. Here, we use an example of constructing a complex of kinase CDK5 and Cyclin B1, with which it is known to interact specifically, as described comprehensively elsewhere [70]. To predict the structure of their complex, their structures (if structures are known) or sequences are inputs to the HOMCOS server. The HOMCOS server identifies the homologous dimeric templates for the target proteins CDK5 and Cyclin B1, by performing 2 rounds of BLAST over the PDB database, one for each of the given target. Of the detected dimeric templates, only those that involve homologs of both, CDK5 and Cyclin B1 are used to build models. The models are associated with statistics such as the percentage identity of aligned residues and contact residues, the number of

aligned contact residues and number of contact residues in the template homolog. Model selection can be assisted by these statistics and is at the discretion of the user.

The HOMCOS server depends on dimeric homologous templates to predict structures. In the absence of such templates, monomer threading followed by oligomer mapping approaches can be used. SPRING(https://zhanglab.ccmb.med.umich.edu/spring/) [71] is one such method (Refer table 3) that models dimeric complexes. SPRING was used to model a homodimeric complex of a peroxidase, 1-Cys peroxiredoxin [71]. To model the homodimeric complex, SPRING takes sequences of the two targets as inputs. In this case, the sequence of 1-Cys peroxiredoxin is used as both the targets. For each of the query proteins, the SPRING algorithm searches templates for threading. The target sequences are threaded onto each of the interacting monomers of the template. The models are then evaluated based on the SPRING score that is a composite of a threading Z-score, a structural alignment score (TM-align score) and a contact based potential. Models are ranked based on the SPRING score. The best scoring model of the dimeric complex of 1-Cys peroxiredoxin had a TM-score of 0.75 and interface RMSD of 3 Å (See Note 5)

3.2 Protein-protein Docking

Protein protein docking can be used when no suitable templates for modelling the protein-protein complexes are available. Docking samples various conformations/configurations in which the two proteins can associate with each other and scores them to identify the most probable mode/pose of association.

Similar to protein-small molecule docking, protein-protein docking can be either local, here the search is localized with the help of user provided restraints, or blind, where the entire surface of the protein is sampled. The following section deals with local and blind docking of protein-protein complexes

3.2.1 Restraint based local docking of protein-protein complexes

As mentioned earlier (section 1.2), local docking methods try to localize (restrict) the conformational sampling space. In protein-protein local docking, the search space can be restricted using user provided restraints. The restraints can be a list of interacting residues of the two proteins, or more specifically be distances between specific amino-acids. Such restraints are often extracted from experimental data. Computational methods such as CPORT[72] (See Note 6), BIPSPI[73], EVcoupling complex[74] (See Note 7) etc. can also be employed to predict the restraints.

Local docking can now be performed using the identified restraints. HADDOCK [75](http://milou.science.uu.nl/services/HADDOCK2.2/) is one of the several software/webservers[refs] that perform local docking. The structure of the two target proteins and the restraints are inputs to HADDOCK. The residues that are known to contribute to the interaction but are of limited importance, called passive residues, can also be specified or HADDOCK can automatically select them. HADDOCK samples docking poses and performs clustering based on the pose similarity. All the clusters are provided as output. The best cluster is the one with the lowest HADDOCK and Z-score. The server also provides values for electrostatic, desolvation, van Der Waals and restraint violation energies (See Note 8).

3.2.2 Blind docking of protein-protein complexes

In the absence of reliable restraints, blind docking can be performed. Blind docking involves prediction of the structure of the protein-protein complex without any prior knowledge of interacting residues or restraints. The Z-dock webserver [76] (http://zdock.umassmed.edu/) performs such blind docking (Refer table 4 for other methods). It takes the structure of the two target proteins as input. If the structure is provided in the form of PDB IDs, the entire biological assembly or specific chains can be used for docking. It uses rigid body docking to

sample conformations of the two targets onto each other. The docking poses are evaluated based on a score that involves shape complementarity, electrostatics and statistical potential terms. The top 'N' docking poses can be further evaluated based on the user's choice. Z-DOCK also provides the facility to select residues that can be part of the binding site or can be excluded from the binding site.

3.3 Evaluating protein-protein complexes

Most, if not all models that are built or predicted are scored based on their in house/known scoring schemes. The complexes can be evaluated by various independent scoring schemes to gain higher confidence in the prediction. The PIZSA [77, 78] webserver (http://cospi.iiserpune.ac.in/pizsa/) predicts if the complex is a binder/non-binder using a knowledge based statistical potential. The predicted complexes can be uploaded on the webserver(See Note 1). A distance cut-off threshold for interface residue definition can be chosen between 4 Å, 6 Å and 8 Å. The best results are obtained at 4 Å. A Z-score value of greater than 1.2 indicates a stable association.

Another scoring scheme from the FoldX suite [79](http://foldxsuite.crg.eu/) can be used to assess the interaction by calculating the binding free energy. FoldX is an empirical force field developed for the fast evaluation of protein complexes. The standalone version can be installed and the protein complex can be evaluated using it. A negative value indicates a feasible interaction.

4. Modelling protein-nucleic acid complexes

Protein-nucleic acid interactions regulate various processes such as gene expression, DNA repair, replication of the DNA/RNA and several others [80]. Structures of protein-nucleic acid complexes are hence vital to get insights into the molecular mechanism of these processes. This section describes computational methods for predicting the structures of complexes of proteins with DNA/RNA. As with other sections, protein-Nucleic acid modelling also has two broad categories, template based modelling and docking.

4.1 Template based modelling of protein-nucleic acid complexes

Template based modelling is preferred over docking in the presence of a suitable template[81]. The template based methods are of two types, homology modelling[82] and fragment based assembly[80]

4.1.1 Homology modelling of protein-nucleic acid complexes

In the presence of homologous templates, methods such as TFmodeller[82] (Refer table 4 for other methods) can be used to model a protein-DNA complex. TFmodeller takes the FASTA sequence of a protein as input. The template to be used if known, can be provided as an input. If not, TFmodeller identifies homologous templates using PSI-BLAST. The homologs are searched in a library of protein-DNA complexes obtained from the PDB. Each of the identified templates is used to build a complex of the query protein with the template DNA. The predicted models, a matrix of homologous interface contacts, the alignment used for the creation of the complex and a list of query positions interacting with the nucleotide along with their conservation are presented as output.

For modelling protein-RNA complexes (Refer table 4 for various methods of protein-RNA complex modelling), MPRDock (http://huanglab.phys.hust.edu.cn/mprdock/) [83] uses a

combination of template based modelling and docking. MPRDock allows flexibility of protein side-chains by considering an ensemble of protein structures that are modelled based on homologous templates. The RNA is considered as a rigid entity and is docked on each protein from the ensemble. The docked complexes are evaluated by an inbuilt scoring function. The lower the score, the better is the model. The input for MPRDock is the structure of RNA and structure or FASTA sequence of the protein. The binding interface and distance restraints (between amino acid and nucleotide residues) if known, can also be provided. The output consists of all the modelled Protein-RNA complexes along with their energy values.

4.1.2 Fragment based modelling of protein-nucleic acid complexes

Protein Assisted DNA Assembly[80] is a fragment based method that can be used to predict the DNA-protein complex or DNA-binding site on a protein. It has a library of small fragments of proteins (length of 6 to 12 amino acids) along with their interacting dsDNA (length of 4 to 8 base pairs) obtained from the known DNA-protein complexes. An empirical interaction model generator performs docking using this library to build docking models. The models are then scored and filtered using a statistical knowledge based force-field (See Note 9)

Similar to protein assisted DNA assembly, RNAx [84] is a fragment based method for docking of RNA fragments. Refer to the tutorial http://modelx.crg.es/PADA1Tutorial for details of the commands for both RNAx and protein assisted DNA assembly.

4.2 Docking of protein-nucleic acid complexes

Protein-nucleic acid docking methods can be knowledge based or *ab initio*. Knowledge based methods can be applied if the information of the interface region is known, otherwise, *ab initio* methods are used.

4.2.1 Knowledge based docking of protein-nucleic acid complexes

Knowledge based docking uses information about the interface residues in the protein. The interface residues can be inferred from experiments or can be predicted computationally. A variety of sequence and structure based algorithms can be used to predict these interface residues on DNA/RNA and on protein (Refer table 4 for the methods). These interface residues can be specified as inputs to HADDOCK (https://milou.science.uu.nl/services/HADDOCK2.2/haddock.php) for docking DNA/RNA on protein[75, 85]. HADDOCK takes structures of both, the protein and the DNA/RNA as inputs (See Note 10). With the specified input structures and restraints, HADDOCK performs rigid docking followed by semi-flexible and solvent refinements. The docking models are clustered based on structural similarity to one another (RMSD). The final clusters (predicted models) are selected based on the HADDOCK scoring function.

4.2.2 Blind docking of protein-nucleic acid complexes

NPDock [86] is an exclusively designed nucleic acid-protein docking method that can be used when the sequence of DNA that binds the protein of interest is known. It has been employed to characterize novel transcription factors such as PvDREB1A[87]. NPDock accepts structures of DNA/RNA and proteins as inputs. The DNA/RNA-protein rigid body docking is performed using GRAMM [88] and scored using statistical potentials, DARS-RNP and QUASI-RNP [89] for RNA-protein complexes and a combination of QUASI-DNP, DFIRE[56] and Varani group potential[25] for DNA-protein complexes. The best scoring models are clustered based on structural similarity and refined using a simulated annealing protocol. The predicted models are the best scoring complexes in the three biggest clusters. The clash score of the best model of the biggest cluster is provided along with the plot for the change in score across the duration of the simulation.

To get better confidence in the models generated by various software, the models can be further assessed using the EvaluateComplex function of ModelX. The command line parameters to be used are mentioned in the ModelX tutorial (http://modelx.crg.es/PADA1Tutorial)

5. Modelling macromolecular assemblies containing various biomolecules

Macromolecular assemblies are biological structures with dimensions in the range of few nanometers to micrometers. They consist of various proteins, peptides, nucleotides, etc. that together act as a functional unit. Elucidating the 3D structures of these macromolecular assemblies is crucial to understand their mechanism. Experimental methods of determining structures of assemblies are challenging due to the complexity and heterogeneity of the assemblies. Computational methods such as integrative modelling can aid in determining the structure of these assemblies. Integrative modelling uses various inputs obtained from multiple experiments, statistical analysis etc. to model the structure of the assembly[90–92]. It follows a four stage process that involves data collection, representation and evaluation of models, sampling conformations and validation. These four stages are iterated until ensemble/s of structures that satisfy the input restraints are found. The following sections describe each of these steps.

5.1 Data collection

This stage involves finding data that describes the assembly. The description involves identifying copy number, shape and localization of each unique component, shape and symmetry of the overall assembly, relative orientations, envelope surface and contacts between the components. These data can be obtained from different independent experiments. For instance, the overall shape and symmetry of the macromolecular assembly

can be obtained by Electron Microscopy (EM) or Cryo-Electron Microscopy (Cryo-EM) (Refer Table 5 for data that can be extracted from various experiments). Along with experimental data, computational data such as homology models of individual components, statistical inferences from bioinformatics data, etc. can also be used to model the 3D structure of the assembly. The quality and quantity of the collected data affect the accuracy of the generated models (See Note11 and Note13).

5.2 Data representation and model evaluation

The data collected in the previous stage is represented as spatial restraints for modelling (Refer table 5). In cases where experimental data are not available, computational techniques play a dominant role in determining the inter-component structural data. Several methods that model Protein-Protein, Protein-DNA/RNA complexes [Refer Section 3.2.1 and 4.2.1] can provide spatial restraints between the interacting components that can be used for macromolecular complex modelling.

The features being restrained include angles, distances and relative orientations. These restraints are in the form of probability density functions that describe the assembly. All the specified restraints are combined into a scoring scheme and used to evaluate the generated conformations.

Integrative modelling Platform (IMP) is one of the earliest software to perform integrative modelling. We use IMP to illustrate the workflow of the integrative modelling method. IMP provides IMP:Model and IMP:Restraint modules [93, 94] to facilitate the representation of experimental data into spatial restraints. These modules can represent different experimental data to a single and compatible platform for representation and scoring.

5.3 Sampling and optimization

Two different protocols can be followed depending on the symmetric or non symmetric nature of the macromolecular assemblies. In a symmetric complex, individual components follow a symmetrical pattern such as linear, spiral, circular (example: Rad51, Microtubules, Actin filaments). In a non symmetric complex, the different components do not follow a regular pattern (example: Ribosome, Proteasome, Chromatin, Intermediate Filaments).

If the macromolecular complex is symmetric, then the symmetry restraint between the repeating units provide a symmetrical axis. Rigid body transformation of the repeating units around the symmetrical axis can be done using CLICK (a topology independent structural superimposition program) [95, 96]to create a complete model of the multi-component macromolecular structure.

If the macromolecular complex is not symmetric, then various conformations are sampled followed by optimization (See Note12). The computational assembly starts with sampling a random configuration. The scoring scheme constructed in section 5.2 evaluates the 3D structure/model. An optimizer minimizes the violated restraints, and the final score defines the quality of the optimized models.

Depending on the type of experimental data, several methods exist in the IMP package and other softwares for computational optimization[97]. For instance, the IMP:MultiFit module for multi-component molecular docking and fitting on EM maps[1], IMP:EmageFit module uses available subunit structures and EM class averages[98], IMP:MultiFoXS for multi-state models using SAXS data [49]

5.4 Ensemble analysis

Models from section 5.3 are clustered based on structural similarity to get ensembles. Analyzing these ensembles allows us to evaluate the quality of the models. The analysis involves the assessment of probability distributions of component properties such as positions, contacts, and localization. Single peak distributions with a small standard deviation indicate precise input information. Lack of such single peak distributions indicates the possibility of alternate configurations/conformations or inconsistent input data. In such cases, the entire exercise can be repeated leaving out fewer or different set of restraints for validation or alternatively by getting more information about the assembly to get more restraints. If the ensemble analysis shows satisfactory results (See Note13 and 14), then the model can be further validated by experimental testing.

Notes

1. The computational techniques are sensitive to clashes and orientations of side chains in the initial input models. The input models should be free of clashes. To remove the clashes chimera can be used. Open the structure in Chimera and go to Tools -> Structure editing -> Energy minimize. A more elaborate energy minimization can also be done to remove clashes and improve interactions using GROMACS [99, 100] (Please follow http://www.mdtutorials.com/ till energy minimization).

2. Model the protein structure to fill in missing atoms/residues (complete PDB) before predicting the binding pockets. Do not add hydrogens to the structure while predicting the binding pockets using DEPTH server. Use the complete PDB with no missing atom for docking.

3. It is recommended to pre-check the PDB file for the presence of mutated non-standard residues. The PDB file should have at least 31 and maximum 1000 amino acids. Additional inputs such as the peptide sequence and the residues that are away from binding site if known, can be provided. These additional inputs improve the computational efficiency of the method.

4. The protein-peptide complex for Rosetta FlexPepDock should not contain any heteroatoms.

5. The models built using SPRING only contains C-alpha atoms. The complete models can be further built using the predicted model as a template using MODELLER (Please follow MODELLER tutorial on Basic modelling at https://salilab.org/modeller/tutorial/basic.html).

6. CPORT over predicts the interface residues. In cases if a large number of interface residues are predicted, one can just take the interface residue prediction from any of the servers that CPORT uses to make the prediction.

7. The coupling file provided by evcouplings contains information of both inter and intra target protein couplings. It is important to filter the table to only extract information about inter protein coupling.

8. HADDOCK server has multiple services based on the type of restraint data. The easy interface is used when the number of interface residues are less and we are confident about them. The prediction interface should be used with tools that over predict the interface such as CPROT. Several restraints such as ambiguous interaction restraints, dipolar coupling restraints, pseudo contact restraints etc. can also be utilized in the Expert and Guru interface. Details about restraints that can be set up using the Expert and Guru Interface can be found in HADDOCK manual (https://www.bonvinlab.org/software/haddock2.2/manual/). Increasing the number of restraints can help reduce docking sampling space and improve prediction accuracy.

9. The MYSQL dumps are extremely large and you may need to install MYSQL in an external device with at least 50GB space available.

10. Important to note that the web server asks for residue numbers. So if the protein has multiple chains and the residue numbers overlap, it can create a problem. So the residues must be renumbered so that they are unique.

11. The amount and quality of the data collected can significantly increase or decrease the accuracy of the models. Thus, the data for building and validating the models should be balanced in terms of quality and quantity.

12. Non-symmetric macromolecular complexes need to be sampled extensively during optimization compared to symmetric complexes. Inappropriate sampling and scoring strategy may present convergence issues to optimizing algorithms and can lead to incorrect models. Thus, obtaining symmetry restraints (if present) can significantly improve the model quality.

13. The clustering of the ensemble can lead to three possible outcomes. (1) A single cluster satisfies all restraints; This implies that the data is sufficient for determining the unique native structure. (2) Two or more clusters satisfy the restraints; Implies that data is insufficient to resolve a unique native structure or there are multiple conformations of the system. 3) No cluster satisfies the restraints; Implies that either the data is wrong or there has been an error in data interpretation.

14. Integrative modelling uses experimental data having different resolutions to construct a 3D model. Thus, different parts of the macromolecular complex have different resolution and accuracy.

References

1. Lasker K, Topf M, Sali A, Wolfson HJ (2009) Inferential Optimization for Simultaneous Fitting of Multiple Components into a CryoEM Map of Their Assembly. J Mol Biol. https://doi.org/10.1016/j.jmb.2009.02.031

2. Greenberg CH, Kollman J, Zelter A, et al (2016) Structure of γ-tubulin small complex based on a cryo-EM map, chemical cross-links, and a remotely related structure. J Struct Biol 194:303–310. https://doi.org/10.1016/j.jsb.2016.03.006

3. Carlsson J, Coleman RG, Setola V, et al (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. Nat Chem Biol 7:769–78. https://doi.org/10.1038/nchembio.662

4. Kuroda D, Shirai H, Jacobson MP, Nakamura H (2012) Computer-aided antibody design. Protein Eng Des Sel 25:507–522. https://doi.org/10.1093/protein/gzs024

5. Soni N, Madhusudhan MS (2017) Computational modeling of protein assemblies. Curr. Opin. Struct. Biol. 44:179–189

6. https://www.ebi.ac.uk/training/online/course/introduction-metabolomics/what-metabolomics/no-glossary-small-molecules-no-glossary. Accessed 30 Jun 2020

7. McFedries A, Schwaid A, Saghatelian A (2013) Methods for the Elucidation of Protein-Small Molecule Interactions. Chem Biol 20:667–673. https://doi.org/10.1016/j.chembiol.2013.04.008

8. Sen N, Kanitkar TR, Roy AA, et al (2019) Predicting and designing therapeutics against the Nipah virus. PLoS Negl Trop Dis 13:e0007419. https://doi.org/10.1371/journal.pntd.0007419

9. Kim S, Thiessen PA, Bolton EE, et al (2016) PubChem Substance and Compound databases. Nucleic Acids Res 44:D1202. https://doi.org/10.1093/NAR/GKV951

10. Irwin JJ, Shoichet BK ZINC-A Free Database of Commercially Available Compounds for Virtual Screening

11. Wishart DS, Feunang YD, Guo AC, et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074–D1082. https://doi.org/10.1093/nar/gkx1037

12. Gaulton A, Hersey A, Nowotka M, et al (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945. https://doi.org/10.1093/NAR/GKW1074

13. Pence HE, Williams A (2010) ChemSpider: An Online Chemical Information Resource. J Chem Educ 87:1123–1124. https://doi.org/10.1021/ed100697w

14. Kanehisa M, Furumichi M, Tanabe M, et al (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353. https://doi.org/10.1093/NAR/GKW1092

15. Hastings J, de Matos P, Dekker A, et al (2012) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res 41:D456–D463. https://doi.org/10.1093/nar/gks1146

16. Feng Z, Chen L, Maddula H, et al (2004) Ligand Depot: a data warehouse for ligands bound to macromolecules. Bioinformatics 20:2153–2155. https://doi.org/10.1093/bioinformatics/bth214

17. Stank A, Kokh DB, Fuller JC, Wade RC (2016) Protein Binding Pocket Dynamics. Acc Chem Res 49:809–815. https://doi.org/10.1021/acs.accounts.5b00516

18. Tan KP, Nguyen TB, Patel S, et al (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res 41:. https://doi.org/10.1093/nar/gkt503

19. Konc J, Miller BT, Štular T, et al (2015) ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. J Chem Inf Model 55:2308–2314. https://doi.org/10.1021/acs.jcim.5b00534

20. Jendele L, Krivak R, Skoda P, et al (2019) PrankWeb: a web server for ligand binding site prediction and visualization. Nucleic Acids Res 47:W345–W349. https://doi.org/10.1093/nar/gkz424

21. Hussein HA, Borrel A, Geneix C, et al (2015) PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 43:W436–W442. https://doi.org/10.1093/nar/gkv462

22. Xu Y, Wang S, Hu Q, et al (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46:W374–W379. https://doi.org/10.1093/nar/gky380

23. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: An open source platform for ligand pocket detection. BMC Bioinformatics 10:168. https://doi.org/10.1186/1471-2105-10-168

24. Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model 15:359–363. https://doi.org/10.1016/S1093-3263(98)00002-3

25. Xu B, Yang Y, Liang H, Zhou Y (2009) An all-atom knowledge-based energy function for protein-DNA threading, docking decoy discrimination, and prediction of

transcription-factor binding profiles. Proteins 76:718–30. https://doi.org/10.1002/prot.22384

26. Brady Jr. GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Des 14:383–401. https://doi.org/10.1023/A:1008124202956

27. Yang J, Roy A, Zhang Y (2013) Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29:2588–2595. https://doi.org/10.1093/bioinformatics/btt447

28. Wass MN, Kelley LA, Sternberg MJE (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res 38:W469–W473. https://doi.org/10.1093/nar/gkq406

29. Repasky MP, Shelley M, Friesner RA (2007) Flexible Ligand Docking with Glide. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA

30. Lang PT, Brozell SR, Mukherjee S, et al (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. RNA 15:1219–30. https://doi.org/10.1261/rna.1563609

31. Verdonk ML, Cole JC, Hartshorn MJ, et al (2003) Improved protein-ligand docking using GOLD. Proteins Struct Funct Bioinforma 52:609–623. https://doi.org/10.1002/prot.10465

32. Cross* SSJ (2005) Improved FlexX Docking Using FlexS-Determined Base Fragment Placement. https://doi.org/10.1021/CI050026F

33. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 21:289–307. https://doi.org/10.1016/S1093-3263(02)00164-X

34. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. J Comput Aided Mol Des 16:151–166. https://doi.org/10.1023/A:1020155510718

35. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785–91. https://doi.org/10.1002/jcc.21256

36. Zhou Y, Elmes MW, Sweeney JM, et al (2019) Identification of Fatty Acid Binding Protein 5 Inhibitors Through Similarity-Based Screening. Biochemistry 58:4304–4316. https://doi.org/10.1021/acs.biochem.9b00625

37. Nguyen MN, Sen N, Lin M, et al (2019) Discovering Putative Protein Targets of Small Molecules: A Study of the p53 Activator Nutlin. J Chem Inf Model 59:1529–1546. https://doi.org/10.1021/acs.jcim.8b00762

38. Krumm BE, Grisshammer R (2015) Peptide ligand recognition by G protein-coupled receptors. Front Pharmacol 6:48. https://doi.org/10.3389/fphar.2015.00048

39. Antunes DA, Devaurs D, Moll M, et al (2018) General Prediction of Peptide-MHC Binding Modes Using Incremental Docking: A Proof of Concept. Sci Rep 8:1–13. https://doi.org/10.1038/s41598-018-22173-4

40. Cunningham AD, Qvit N, Mochly-Rosen D, Opin Struct Biol Author manuscript C (2017) Peptides and peptidomimetics as regulators of protein-protein interactions HHS Public Access Author manuscript. Curr Opin Struct Biol 44:59–66. https://doi.org/10.1016/j.sbi.2016.12.009

41. Du Q-S, Xie N-Z, Huang R-B (2015) Recent Development of Peptide Drugs and Advance on Theory and Methodology of Peptide Inhibitor Design. Med Chem (Los Angeles) 11:235–247. https://doi.org/10.2174/1573406411666141229163355

42. Lau JL, Dunn MK (2018) Therapeutic peptides: Historical perspectives, current development trends, and future directions. Bioorganic Med Chem 26:2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052

43. Yan C, Zou X (2015) Predicting peptide binding sites on protein surfaces by clustering chemical interactions. J Comput Chem 36:49–61. https://doi.org/10.1002/jcc.23771

44. Taherzadeh G, Zhou Y, Liew AW-C, Yang Y (2018) Structure-based prediction of protein– peptide binding regions using Random Forest. Bioinformatics 34:477–484. https://doi.org/10.1093/bioinformatics/btx614

45. Trabuco LG, Lise S, Petsalaki E, Russell RB (2012) PepSite: prediction of peptide-binding sites from protein surfaces. Nucleic Acids Res 40:W423–W427. https://doi.org/10.1093/nar/gks398

46. Lavi A, Ngan CH, Movshovitz-Attias D, et al (2013) Detection of peptide-binding sites on protein surfaces: The first step toward the modeling and targeting of peptide-mediated interactions. Proteins Struct Funct Bioinforma 81:2096–2105. https://doi.org/10.1002/prot.24422

47. Litfin T, Yang Y, Zhou Y (2019) SPOT-Peptide: Template-Based Prediction of Peptide-Binding Proteins and Peptide-Binding Sites. J Chem Inf Model 59:924–930. https://doi.org/10.1021/acs.jcim.8b00777

48. 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:W431–W435. https://doi.org/10.1093/nar/gkv495

49. London N, Raveh B, Cohen E, et al (2011) Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. Nucleic Acids Res 39:W249–W253. https://doi.org/10.1093/nar/gkr431

50. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/J.JMB.2015.09.014

51. Verschueren E, Vanhee P, Rousseau F, et al (2013) Protein-peptide complex prediction through fragment interaction patterns. Structure 21:789–797. https://doi.org/10.1016/j.str.2013.02.023

52. Vanhee P, Stricher F, Baeten L, et al (2009) Protein-Peptide Interactions Adopt the Same Structural Motifs as Monomeric Protein Folds. Structure 17:1128–1136. https://doi.org/10.1016/j.str.2009.06.013

53. Unal EB, Gursoy A, Erman B (2010) VitAL: Viterbi Algorithm for de novo Peptide Design. PLoS One 5:e10926. https://doi.org/10.1371/journal.pone.0010926

54. Petsalaki E, Stark A, García-Urdiales E, Russell RB (2009) Accurate Prediction of Peptide Binding Sites on Protein Surfaces. PLoS Comput Biol 5:e1000335. https://doi.org/10.1371/journal.pcbi.1000335

55. Frenkel D, Clark DE, Li J, et al (1995) PRO_LIGAND: An approach to de novo molecular design. 4. Application to the design of peptides. J Comput Aided Mol Des 9:213–225. https://doi.org/10.1007/BF00124453

56. Zhou H, Zhou Y (2009) 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. https://doi.org/10.1110/ps.0217002

57. 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:W431–W435. https://doi.org/10.1093/nar/gkv495

58. Ko J, Park H, Heo L, Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 40:W294–W297. https://doi.org/10.1093/nar/gks493

59. Ko J, Park H, Seok C (2012) GalaxyTBM: Template-based modeling by building a reliable core and refining unreliable local regions. BMC Bioinformatics 13:198. https://doi.org/10.1186/1471-2105-13-198

60. Lee H, Seok C (2017) Template-Based Prediction of Protein-Peptide Interactions by Using GalaxyPepDock. Humana Press, New York, NY, pp 37–47

61. Raveh B, London N, Schueler-Furman O (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins Struct Funct Bioinforma 78:NA-NA. https://doi.org/10.1002/prot.22716

62. Slutzki M, Ben-Shimon A, Niv MY (2017) AnchorDock for Blind Flexible Docking of Peptides to Proteins. Humana Press, New York, NY, pp 95–108

63. Ben-Shimon A, Niv MY (2015) AnchorDock: Blind and Flexible Anchor-Driven Peptide Docking. Structure 23:929–940. https://doi.org/10.1016/J.STR.2015.03.010

64. Ben-Shimon A, Eisenstein M (2010) Computational Mapping of Anchoring Spots on Protein Surfaces. J Mol Biol 402:259–277. https://doi.org/10.1016/j.jmb.2010.07.021

65. Stein A, Mosca R, Aloy P (2011) Three-dimensional modeling of protein interactions and complexes is going 'omics. Curr. Opin. Struct. Biol. 21:200–208

66. Baspinar A, Cukuroglu E, Nussinov R, et al (2014) PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. Nucleic Acids Res 42:W285–W289. https://doi.org/10.1093/nar/gku397

67. Mashiach E, Nussinov R, Wolfson HJ (2010) FiberDock: Flexible induced-fit backbone refinement in molecular docking. Proteins Struct Funct Bioinforma. https://doi.org/10.1002/prot.22668

68. Tuncbag N, Gursoy A, Nussinov R, Keskin O (2011) Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. Nat Protoc 6:1341–1354. https://doi.org/10.1038/nprot.2011.367

69. Kawabata T (2016) HOMCOS: an updated server to search and model complex 3D structures. J Struct Funct Genomics 17:83–99. https://doi.org/10.1007/s10969-016-9208-y

70. Fukuhara N, Kawabata T (2008) HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures. Nucleic Acids Res 36:. https://doi.org/10.1093/nar/gkn218

71. Guerler A, Govindarajoo B, Zhang Y (2013) Mapping Monomeric Threading to Protein–Protein Structure Prediction. J Chem Inf Model 53:717–725. https://doi.org/10.1021/ci300579r

72. de Vries SJ, Bonvin AMJJ (2011) CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK. PLoS One 6:e17695. https://doi.org/10.1371/journal.pone.0017695

73. Sanchez-Garcia R, Sorzano COS, Carazo JM, Segura J (2019) BIPSPI: a method for the prediction of partner-specific protein–protein interfaces. Bioinformatics 35:470–477. https://doi.org/10.1093/bioinformatics/bty647

74. Hopf TA, Schärfe CPI, Rodrigues JPGLM, et al (2014) Sequence co-evolution gives 3D contacts and structures of protein complexes. Elife 3:. https://doi.org/10.7554/eLife.03430

75. Van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014

76. Pierce BG, Wiehe K, Hwang H, et al (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 30:1771–1773. https://doi.org/10.1093/bioinformatics/btu097

77. Roy AA, Dhawanjewar AS, Sharma P, et al (2019) Protein Interaction Z Score Assessment (PIZSA): an empirical scoring scheme for evaluation of protein–protein interactions. Nucleic Acids Res 47:W331–W337. https://doi.org/10.1093/nar/gkz368

78. Dhawanjewar AS, Roy AA, Madhusudhan MS (2019) A knowledge-based scoring function to assess the stability of quaternary protein assemblies. bioRxiv 562520. https://doi.org/10.1101/562520

79. Schymkowitz J, Borg J, Stricher F, et al (2005) The FoldX web server: an online force field. Nucleic Acids Res 33:W382–W388. https://doi.org/10.1093/nar/gki387

80. Blanco JD, Radusky L, Climente-González H, Serrano L (2018) FoldX accurate structural protein-DNA binding prediction using PADA1 (Protein Assisted DNA Assembly 1). Nucleic Acids Res 46:3852–3863. https://doi.org/10.1093/nar/gky228

81. Xue LC, Rodrigues JPGLM, Dobbs D, et al (2016) Template-based protein–protein docking exploiting pairwise interfacial residue restraints. Brief Bioinform bbw027. https://doi.org/10.1093/bib/bbw027

82. Contreras-Moreira B, Branger P-A, Collado-Vides J (2007) TFmodeller: comparative modelling of protein–DNA complexes. Bioinformatics 23:1694–1696. https://doi.org/10.1093/bioinformatics/btm148

83. He J, Tao H, Huang S-Y (2019) Protein-ensemble–RNA docking by efficient consideration of protein flexibility through homology models. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz388

84. Blanco JD, Radusky LG, Cianferoni D, Serrano L (2019) Protein-assisted RNA fragment docking (RnaX) for modeling RNA–protein interactions using ModelX. Proc Natl Acad Sci U S A 116:24568–24573. https://doi.org/10.1073/pnas.1910999116

85. Karaca E, Melquiond ASJ, de Vries SJ, et al (2010) Building Macromolecular Assemblies by Information-driven Docking. Mol Cell Proteomics 9:1784–1794. https://doi.org/10.1074/mcp.M000051-MCP201

86. Tuszynska I, Magnus M, Jonak K, et al (2015) NPDock: a web server for protein–nucleic acid docking. Nucleic Acids Res 43:W425–W430. https://doi.org/10.1093/nar/gkv493

87. Vatansever R, Uras ME, Sen U, et al (2017) Isolation of a transcription factor DREB1A gene from Phaseolus vulgaris and computational insights into its characterization: protein modeling, docking and mutagenesis. J Biomol Struct Dyn 35:3107–3118. https://doi.org/10.1080/07391102.2016.1243487

88. Katchalski-Katzir E, Shariv I, Eisenstein M, et al (1992) Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci U S A 89:2195–2199. https://doi.org/10.1073/pnas.89.6.2195

89. Tuszynska I, Bujnicki JM (2011) DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. BMC Bioinformatics 12:348. https://doi.org/10.1186/1471-2105-12-348

90. Alber F, Dokudovskaya S, Veenhoff LM, et al (2007) Determining the architectures of macromolecular assemblies. Nature 450:683–694. https://doi.org/10.1038/nature06404

91. Rout MP, Sali A (2019) Principles for Integrative Structural Biology Studies. Cell

92. Braitbard M, Schneidman-Duhovny D, Kalisman N (2019) Integrative Structure Modeling: Overview and Assessment. Annu Rev Biochem. https://doi.org/10.1146/annurev-biochem-013118-111429

93. Russel D, Lasker K, Webb B, et al (2012) Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. https://doi.org/10.1371/journal.pbio.1001244

94. Alber F, Förster F, Korkin D, et al (2008) Integrating Diverse Data for Structure Determination of Macromolecular Assemblies. Annu Rev Biochem. https://doi.org/10.1146/annurev.biochem.77.060407.135530

95. Nguyen MN, Madhusudhan MS (2011) Biological insights from topology independent comparison of protein 3D structures. Nucleic Acids Res 39:e94–e94. https://doi.org/10.1093/nar/gkr348

96. Nguyen MN, Tan KP, Madhusudhan MS (2011) CLICK - Topology-independent comparison of biomolecular 3D structures. Nucleic Acids Res 39:W24–W28. https://doi.org/10.1093/nar/gkr393

97. Soni N, Madhusudhan MS (2017) Computational modeling of protein assemblies. Curr Opin Struct Biol 44:179–189. https://doi.org/10.1016/J.SBI.2017.04.006

98. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw389

99. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. https://doi.org/10.1016/0010-4655(95)00042-E

100. Pronk S, Páll S, Schulz R, et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854. https://doi.org/10.1093/bioinformatics/btt055

101. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput Chem 30:2785. https://doi.org/10.1002/JCC.21256

102. Trott,O., Olson AJ (2019) Autodock vina: improving the speed and accuracy of docking. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc.21334.AutoDock

103. Allen WJ, Balius TE, Mukherjee S, et al (2015) DOCK 6: Impact of New Features and Current Docking Performance. J Comput Chem 36:1132. https://doi.org/10.1002/JCC.23905

104. Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res 39:W270. https://doi.org/10.1093/NAR/GKR366

105. Antunes DA, Moll M, Devaurs D, et al (2017) DINC 2.0: A New Protein–Peptide Docking Webserver Using an Incremental Approach. Cancer Res 77:e55–e57. https://doi.org/10.1158/0008-5472.CAN-17-0511

106. Jain† AN (2003) Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. https://doi.org/10.1021/JM020406H

107. Schindler CEM, de Vries SJ, Zacharias M (2015) Fully Blind Peptide-Protein Docking with pepATTRACT. Structure 23:1507–1515. https://doi.org/10.1016/J.STR.2015.05.021

108. Yan C, Xu X, Zou X (2016) Fully Blind Docking at the Atomic Level for Protein-Peptide Complex Structure Prediction. Structure 24:1842–1853. https://doi.org/10.1016/J.STR.2016.07.021

109. Kurcinski M, Jamroz M, Blaszczyk M, et al (2015) CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. Nucleic Acids Res 43:W419–W424. https://doi.org/10.1093/nar/gkv456

110. Porter KA, Xia B, Beglov D, et al (2017) ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. Bioinformatics 33:3299–3301. https://doi.org/10.1093/bioinformatics/btx216

111. Alam N, Goldstein O, Xia B, et al (2017) High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. PLOS Comput Biol 13:e1005905. https://doi.org/10.1371/journal.pcbi.1005905

112. Frenkel D, Clark DE, Li J, et al (1995) PRO_LIGAND: an approach to de novo molecular design. 4. Application to the design of peptides. J Comput Aided Mol Des 9:213–25. https://doi.org/10.1007/bf00124453

113. Antes I (2010) DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. Proteins Struct Funct Bioinforma 78:1084–1104. https://doi.org/10.1002/prot.22629

114. Donsky E, Wolfson HJ (2011) PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. Bioinformatics 27:2836–2842. https://doi.org/10.1093/bioinformatics/btr498

115. Trellet M, Melquiond ASJ, Bonvin AMJJ (2013) A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking. PLoS One 8:e58769. https://doi.org/10.1371/journal.pone.0058769

116. Lamiable A, Thévenet P, Rey J, et al (2016) PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44:W449–W454. https://doi.org/10.1093/nar/gkw329

117. Mukherjee S, Zhang Y (2011) Protein-protein complex structure predictions by multimeric threading and template recombination. Structure 19:955–966. https://doi.org/10.1016/j.str.2011.04.006

118. Macindoe G, Mavridis L, Venkatraman V, et al (2010) HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res 38:W445–W449. https://doi.org/10.1093/nar/gkq311

119. Kozakov D, Hall DR, Xia B, et al (2017) The ClusPro web server for protein-protein docking. Nat Protoc 12:255–278. https://doi.org/10.1038/nprot.2016.169

120. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc 125:1731–1737. https://doi.org/10.1021/ja026939x

121. Quignot C, Rey J, Yu J, et al (2018) InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. Nucleic Acids Res 46:W408–W416. https://doi.org/10.1093/nar/gky377

122. Hosur R, Xu J, Bienkowska J, Berger B (2011) IWRAP: An interface threading approach with application to prediction of cancer-related protein-protein interactions. J Mol Biol 405:1295–1310. https://doi.org/10.1016/j.jmb.2010.11.025

123. Aloy P, Russell RB (2003) InterPreTS: protein Interaction Prediction through Tertiary Structure. Bioinformatics 19:161–162. https://doi.org/10.1093/bioinformatics/19.1.161

124. Singh R, Park D, Xu J, et al (2010) Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. Nucleic Acids Res 38:W508–W515. https://doi.org/10.1093/nar/gkq481

125. Hosur R, Peng J, Vinayagam A, et al (2012) A computational framework for boosting confidence in high-throughput protein-protein interaction datasets. Genome Biol. https://doi.org/10.1186/gb-2012-13-8-r76

126. Patro LPP, Kumar A, Kolimi N, Rathinavelan T (2017) 3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-D imensional Nu cleic Acid S tructures. J Mol Biol 429:2438–2448. https://doi.org/10.1016/j.jmb.2017.06.013

127. Magnus M, Boniecki MJ, Dawson W, Bujnicki JM (2016) SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. Nucleic Acids Res 44:W315–W319. https://doi.org/10.1093/nar/gkw279

128. Rother M, Milanowska K, Puton T, et al (2011) ModeRNA server: an online tool for modeling RNA 3D structures. Bioinformatics 27:2441–2442. https://doi.org/10.1093/bioinformatics/btr400

129. Sukumar S, Zhu X, Ericksen SS, Mitchell JC (2016) DBSI server: DNA binding site identifier. Bioinformatics 32:2853–2855. https://doi.org/10.1093/bioinformatics/btw315

130. Tjong H, Zhou H-X (2007) DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces. Nucleic Acids Res 35:1465–1477. https://doi.org/10.1093/nar/gkm008

131. Chen YC, Wright JD, Lim C (2012) DR_bind: a web server for predicting DNA-binding residues from the protein structure based on electrostatics, evolution and geometry. Nucleic Acids Res 40:W249-56. https://doi.org/10.1093/nar/gks481

132. Ozbek P, Soner S, Erman B, Haliloglu T (2010) DNABINDPROT: fluctuation-based predictor of DNA-binding residues within a network of interacting residues. Nucleic Acids Res 38:W417-23. https://doi.org/10.1093/nar/gkq396

133. Liu Z-P, Wu L-Y, Wang Y, et al (2010) Prediction of protein–RNA binding sites by a random forest method with combined features. Bioinformatics 26:1616–1622. https://doi.org/10.1093/bioinformatics/btq253

134. Li S, Yamashita K, Amada KM, Standley DM (2014) Quantifying sequence and structural features of protein-RNA interactions. Nucleic Acids Res 42:10086–98. https://doi.org/10.1093/nar/gku681

135. Eswar N, Webb B, Marti-Renom MA, et al (2006) Comparative Protein Structure Modeling Using Modeller. Curr Protoc Bioinforma 15:5.6.1-5.6.30. https://doi.org/10.1002/0471250953.bi0506s15

136. Yan J, Kurgan L (2017) DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues. Nucleic Acids Res 45:e84. https://doi.org/10.1093/nar/gkx059

137. Hwang S, Gou Z, Kuznetsov IB (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. Bioinformatics 23:634–636. https://doi.org/10.1093/bioinformatics/btl672

138. Kumar M, Gromiha MM, Raghava GPS (2008) Prediction of RNA binding sites in a protein using SVM and PSSM profile. Proteins Struct Funct Bioinforma 71:189–194. https://doi.org/10.1002/prot.21677

139. Tuvshinjargal N, Lee W, Park B, Han K (2016) PRIdictor: Protein–RNA Interaction predictor. Biosystems 139:17–22. https://doi.org/10.1016/J.BIOSYSTEMS.2015.10.004

140. Panwar B, Raghava GPS (2015) Identification of protein-interacting nucleotides in a RNA sequence using composition profile of tri-nucleotides. Genomics 105:197–203. https://doi.org/10.1016/J.YGENO.2015.01.005

141. Messeguer X, Escudero R, Farre D, et al (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18:333–334. https://doi.org/10.1093/bioinformatics/18.2.333

142. Farré D, Roset R, Huerta M, et al (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res 31:3651–3. https://doi.org/10.1093/nar/gkg605

143. Tsunoda T, Takagi T (1999) Estimating transcription factor bindability on DNA. Bioinformatics 15:622–630. https://doi.org/10.1093/bioinformatics/15.7.622

144. Kreft Ł, Soete A, Hulpiau P, et al (2017) ConTra v3: a tool to identify transcription factor binding sites across species, update 2017. Nucleic Acids Res 45:W490–W494. https://doi.org/10.1093/nar/gkx376

145. Gearing LJ, Cumming HE, Chapman R, et al (2019) CiiiDER: A tool for predicting and analysing transcription factor binding sites. PLoS One 14:e0215495. https://doi.org/10.1371/journal.pone.0215495

146. Zheng J, Xie J, Hong X, Liu S (2019) RMalign: an RNA structural alignment tool based on a novel scoring function RMscore. BMC Genomics 20:276. https://doi.org/10.1186/s12864-019-5631-3

147. Gabb HA, Jackson RM, Sternberg MJE (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information 1 1Edited by J. Thornton. J Mol Biol 272:106–120. https://doi.org/10.1006/jmbi.1997.1203

148. Banitt I, Wolfson HJ (2011) ParaDock: a flexible non-specific DNA--rigid protein docking algorithm. Nucleic Acids Res 39:e135. https://doi.org/10.1093/nar/gkr620

149. Huang Y, Li H, Xiao Y (2018) 3dRPC: a web server for 3D RNA–protein structure prediction. Bioinformatics 34:1238–1240. https://doi.org/10.1093/bioinformatics/btx742

150. Yan Y, Zhang D, Zhou P, et al (2017) HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res 45:W365–W373. https://doi.org/10.1093/nar/gkx407

151. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:W363-7. https://doi.org/10.1093/nar/gki481

152. Gruber AR, Lorenz R, Bernhart SH, et al (2008) The Vienna RNA websuite. Nucleic Acids Res 36:W70-4. https://doi.org/10.1093/nar/gkn188

153. Li S, Olson WK, Lu X-J (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. Nucleic Acids Res 47:W26–W34. https://doi.org/10.1093/nar/gkz394

154. van Dijk M, Bonvin AMJJ (2009) 3D-DART: a DNA structure modelling server. Nucleic Acids Res 37:W235-9. https://doi.org/10.1093/nar/gkp287

155. Mouradov D, Craven A, Forwood JK, et al (2006) Modelling the structure of latexin-carboxypeptidase A complex based on chemical cross-linking and molecular docking. Protein Eng Des Sel. https://doi.org/10.1093/protein/gzi070

156. Mouradov D, King G, Ross IL, et al (2008) Protein structure determination using a combination of cross-linking, mass spectrometry, and molecular modeling. Methods Mol Biol. https://doi.org/10.1007/978-1-60327-058-8_31

157. Forwood JK, Thakur AS, Guncar G, et al (2007) Structural basis for recruitment of tandem hotdog domains in acyl-CoA thioesterase 7 and its role in inflammation. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.0700974104

158. Schröder GF, Grubmüller H (2004) FRETsg: Biomolecular structure model building from multiple FRET experiments. Comput Phys Commun. https://doi.org/10.1016/j.cpc.2004.02.001

159. Alexander N, Al-Mestarihi A, Bortolus M, et al (2008) De Novo High-Resolution Protein Structure Determination from Sparse Spin-Labeling EPR Data. Structure. https://doi.org/10.1016/j.str.2007.11.015

160. Schmitz C, Vernon R, Otting G, et al (2012) Protein structure determination from pseudocontact shifts using ROSETTA. J Mol Biol. https://doi.org/10.1016/j.jmb.2011.12.056

161. Zheng W, Doniach S (2002) Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. J Mol Biol. https://doi.org/10.1006/jmbi.2001.5324

162. Zheng W, Doniach S (2005) Fold recognition aided by constraints from small angle X-ray scattering data. Protein Eng Des Sel. https://doi.org/10.1093/protein/gzi026

163. De Rosier DJ, Klug A (1968) Reconstruction of three dimensional structures from electron micrographs. Nature. https://doi.org/10.1038/217130a0

164. Nogales E, Scheres SHW (2015) Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Mol. Cell

165. Short JM, Liu Y, Chen S, et al (2016) High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. Nucleic Acids Res 44:9017–9030. https://doi.org/10.1093/nar/gkw783

166. Ho C-M, Li X, Lai M, et al (2019) Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. Nat Methods 1–7. https://doi.org/10.1038/s41592-019-0637-y

167. Hamuro Y, Burns LL, Canaves JM, et al (2002) Domain organization of D-AKAP2 revealed by enhanced deuterium exchange-mass spectrometry (DXMS). J Mol Biol. https://doi.org/10.1016/S0022-2836(02)00419-9

168. Kamal JKA, Chance MR (2007) Modeling of protein binary complexes using structural mass spectrometry data. Protein Sci. https://doi.org/10.1110/ps.073071808

169. Lees JG, Janes RW (2008) Combining sequence-based prediction methods and circular dichroism and infrared spectroscopic data to improve protein secondary structure determinations. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-9-24

Table 1: Non-exhaustive list of protein-small molecule docking methods

Tool	URL
AutoDock[101]	http://autodock.scripps.edu/downloads/autodock-registration/autodock-4-2-download-page/ (standalone)
AutoDock Vina[102]	http://vina.scripps.edu/download.html (standalone)
DOCK[103]	http://dock.compbio.ucsf.edu/Online_Licensing/index.htm (standalone)
FlexX[32]	https://www.biosolveit.de/FlexX/ (standalone)
GOLD[31]	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/ (standalone)
GLIDE[29]	https://www.schrodinger.com/glide (standalone)
LigandFit[33]	Not available
SwissDock[104]	http://www.swissdock.ch/ (standalone)

Table 2: Non-exhaustive list of protein-peptide complex modelling methods

Tool	Algorithm	URL
Pro_Ligand[112]	De novo	Not available
SPOT-Peptide[47]	Knowledge-based	http://sparks-lab.org/tom/SPOT-peptide/
GalaxyPepDock[57]	Template-based Docking	http://galaxy.seoklab.org/pepdock
Rosetta FlexPepDock[49]	Local Docking	http://flexpepdock.furmanlab.cs.huji.ac.il/
DynaDock[113]		Not available
PepCrawler[114]		http://bioinfo3d.cs.tau.ac.il/PepCrawler/
HADDOCK peptide docking[115]		http://milou.science.uu.nl/services/HADDOCK2.2/haddock.php
PEP-FOLD 3[116]		http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3
AutoDock Vina[102]		http://vina.scripps.edu/download.html (standalone)
DINC 2.0[105]		http://dinc.kavrakilab.org/
Surflex-Dock[106]		https://omictools.com/surflex-dock-tool (standalone)
pepATTRACT[107]	Blind Docking	http://bioserv.rpbs.univ-paris-diderot.fr/services/pepATTRACT/
MDockPeP[108]		http://zougrouptoolkit.missouri.edu/mdockpep/
CABS-dock[109]		http://biocomp.chem.uw.edu.pl/CABSdock

AnchorDock[62]		Not available
ClusPro PeptiDock[110]		https://peptidock.cluspro.org/
PIPER-FlexPepDock[111]		http://piperfpd.furmanlab.cs.huji.ac.il/

Table 3: Non-exhaustive list of protein-protein complex modelling methods

Tool	**Algorithm**	**URL**
Interactome 3D[65]	Template based	http://interactome3d.irbbarcelona.org/
HOMCOS[69]	Template based	http://strcomp.protein.osaka-u.ac.jp/homcos/
PRISM[66]	Template based	http://prism.ccbb.ku.edu.tr/
iWRAP [122]	Template based	http://groups.csail.mit.edu/cb/iwrap/
InterPreTS[123]	Template based	http://www.russelllab.org/cgi-bin/tools/interprets.pl
SPRING[71]	Template based	http://zhanglab.ccmb.med.umich.edu/spring/
Struct2Net[124]	Template based	http://groups.csail.mit.edu/cb/struct2net/webserver/
Coev2Net[125]	Template based	http://groups.csail.mit.edu/cb/coev2net/
COTH[117]	Template based	http://zhanglab.ccmb.med.umich.edu/COTH/
ZDOCK[76]	Docking	http://zdock.umassmed.edu/
Hex[118]	Docking	http://hexserver.loria.fr/
ClusPro[119]	Docking	https://cluspro.bu.edu/login.php
HADDOCK[120]	Docking	http://www.bonvinlab.org/software/haddock2.2/
InterEVDock2[121]	Docking	http://bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock2/

Table 4: Non-exhaustive list of some protein-nucleic acid complex modelling methods

Tool	Algorithm	Type of nucleic acid	URL
MODELLER[135]	Homology modelling	Protein/ DNA/RNA	https://salilab.org/modeller/download_installation.html (standalone)
Prime 2.0[146]		RNA	http://www.rnabinding.com/PRIME/PRIME2.0.html
MPRDock[83]		RNA	http://huanglab.phys.hust.edu.cn/mprdock/
RNA Secondary Structure Prediction			
RNAfold[152]	Dynamic Programming	RNA	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi
Modelling Nucleic Acid structure			
3DNA[153]	Geometry-Based modelling	DNA/RNA	http://web.x3dna.org/index.php/fibermodel
3D-DART[154]		DNA	http://milou.science.uu.nl/services/3DDART/
3D-Nus[126]		DNA/RNA	https://iith.ac.in/3dnus/DNA%20Mismatch.html
SimRNA[127]		RNA	https://genesilico.pl/SimRNAweb/submit
ModeRNA[128]	Template-Based modelling	RNA	http://iimcb.genesilico.pl/modernaserver/submit/model/
Binding site prediction			
DBSI[129]	Structure-Based	DNA	https://mitchell-lab.biochem.wisc.edu/DBSI_Server/upload.php
DISPLAR[130]		DNA	https://pipe.rcc.fsu.edu/displar.html
DR_Bind[131]		DNA	http://dnasite.limlab.ibms.sinica.edu.tw/
DNABINDPROT[132]		DNA	www.prc.boun.edu.tr/appserv/prc/dnabindprot/
PRNA[133]	Sequence-Based	RNA	http://doc.aporc.org/wiki/PRNA (standalone)
aaRNA[134]		RNA	https://sysimm.ifrec.osaka-u.ac.jp/aarna/
DRNAPred[136]		DNA/RNA	http://biomine.cs.vcu.edu/servers/DRNApred/
DP-Bind[137]		DNA	http://lcg.rit.albany.edu/dp-bind/
Pprint[138]		RNA	https://webs.iiitd.edu.in/raghava/pprint/submit.html
PRIdictor[139]		RNA/Protein	http://bclab.inha.ac.kr/pridictor/pridictor.html
RNApin[140]		RNA	https://webs.iiitd.edu.in/raghava/rnapin/submit.php
PROMO[141, 142]		DNA	http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3
TFBind[143]		DNA	http://tfbind.hgc.jp/
ConTra v3[144]		DNA	http://bioit2.irc.ugent.be/contra/v3/#/step/1
CiiiDER[145]		DNA	http://www.ciiider.org/ (standalone)
Modelling DNA-protein Complex (Structure of protein with DNA does not exist)			
TFModeller[82]	Homology modelling	DNA-protein	http://maya.ccg.unam.mx/~tfmodell/index.html

HADDOCK[50]	Knowledge-based Docking	Nucleic acid-protein or Protein-protein	https://milou.science.uu.nl/services/HADDOCK2.2/haddock.php
[147]	Rigid Body Docking	Nucleic acid-protein or Protein-protein	http://www.sbg.bio.ic.ac.uk/docking/download.html (standalone)
ParaDock[148]		Nucleic acid-protein	http://bioinfo3d.cs.tau.ac.il/ParaDock/php.php
NPDock[86]	Rigid Body Knowledge-based Docking	Nucleic acid-protein	http://genesilico.pl/NPDock
3dRPC[149]		RNA-protein	http://biophy.hust.edu.cn/3dRPC
HDOCK[150]		Nucleic acid-protein/Protein-protein	http://hdock.phys.hust.edu.cn/
PatchDock[151]		Nucleic acid-protein/Protein-protein	https://bioinfo3d.cs.tau.ac.il/PatchDock/

Table 5: Very few software suites do most/all steps of integrative modelling. This table is a list of methods that could be used for the conversion of experimental data into spatial restraints for macromolecular assembly modelling. References to studies where such methods were utilized is provided in the last column.

Experimental Technique	**Measured Data**	**Structural Data**	**Example Reference**
Chemical cross-linking	Mass/charge ratio of joint fragments	Upper limit on pair distance between reacted groups	[155–157]
Forster Resonance Energy Transfer (FRET)	The yield of fluorescence energy transfer	Distance between donor-acceptor pairs	[158]
Electron Paramagnetic Resonance (EPR)	Dipole-dipole coupling between electron spins	Distance between pairs of spin labels	[159, 160]
Small Angle X-Ray Scattering (SAXS)	Scattering intensity as a function of momentum transfer	Pair distribution function or shape envelope.	[161, 162]
EM and Cryo-EM	Shape envelope	Volume restraints	[163–166]
Deuterium Exchange Mass Spectroscopy (DXMS)	Rate constant of H/D exchange	Solvent exposure	[167]
Radical footprinting	Rate constant from the dose-response curve	Solvent exposure	[168]
Circular Dichroism (CD)	Mean residue ellipticity as a function of wavelength	Secondary structure content	[169]

Figure1: Spreadsheet of select protein complex structure modelling methods that can be used depending on the information available. The boxed methods that span various sections indicate applicability of the method in multiple categories.