



*Review*

# Bioinformatics Tools and Benchmarks for Computational Docking and 3D Structure Prediction of RNA-Protein Complexes

Chandran Nithin <sup>1,†</sup> , Pritha Ghosh <sup>1,†</sup>  and Janusz M. Bujnicki <sup>1,2,\*</sup>

<sup>1</sup> Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland; nchandran@genesilico.pl (C.N.); pghosh@genesilico.pl (P.G.)

<sup>2</sup> Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, PL-61-614 Poznan, Poland

\* Correspondence: iamb@genesilico.pl

† These authors have contributed equally and should be considered joint first authors.

Received: 14 May 2018; Accepted: 21 August 2018; Published: 25 August 2018



**Abstract:** RNA-protein (RNP) interactions play essential roles in many biological processes, such as regulation of co-transcriptional and post-transcriptional gene expression, RNA splicing, transport, storage and stabilization, as well as protein synthesis. An increasing number of RNP structures would aid in a better understanding of these processes. However, due to the technical difficulties associated with experimental determination of macromolecular structures by high-resolution methods, studies on RNP recognition and complex formation present significant challenges. As an alternative, computational prediction of RNP interactions can be carried out. Structural models obtained by theoretical predictive methods are, in general, less reliable compared to models based on experimental measurements but they can be sufficiently accurate to be used as a basis for formulating functional hypotheses. In this article, we present an overview of computational methods for 3D structure prediction of RNP complexes. We discuss currently available methods for macromolecular docking and for scoring 3D structural models of RNP complexes in particular. Additionally, we also review benchmarks that have been developed to assess the accuracy of these methods.

**Keywords:** ribonucleoprotein; RNP; macromolecular complex; computational modelling; structural bioinformatics

## 1. Introduction

Ribonucleic acid (RNA) plays major roles in various biological processes including protein synthesis and gene regulation at the transcriptional and post-transcriptional level. RNA molecules are involved in catalysing biological reactions, controlling gene expression, or sensing and communicating responses to cellular signals [1–3]. The majority of the known RNAs exert their function in conjunction with proteins to form RNA-protein (RNP) complexes, at one or more stages of their life cycle. The strength of these complexes can vary from being stable, like in the case of the individual subunits of the ribosome [4], or being able to undergo extensive rearrangements like the spliceosome [5] to being transient, enabling their assembly and disassembly as is observed in the exon junction complex [6]. RNA-protein complexes are involved in many cellular processes, including the maintenance of chromosome ends, transcription, RNA transport and processing, regulation of gene expression, protein synthesis [1–3,7], alternative splicing [8], RNA modification and polyadenylation [9,10]. Moreover, the protein non-coding RNAs (ncRNAs) act as scaffolds during macromolecular assembly [11]. For instance, 7SK ncRNA acts as a scaffold for the formation of multiple RNPs and is a major player

in the regulation of eukaryotic transcription [12]. Furthermore, RNPs help govern the association of sister chromatid cohesion proteins with genes and enhancers [13]. Defects in RNP interactions are implicated in many diseases ranging from neurological disorders to cancer [14,15]. RNP interactions are thus essential for the critical aspects of cellular metabolism.

RNA-binding proteins (RBPs) often contain structurally and functionally distinct modules. For instance, in all enzymes that act on RNA, RNA-binding is a common feature of catalytic domains that assume various well-defined three-dimensional (3D) structural folds. Some of these domains can bind RNA on their own, while others require dedicated RNA-binding domains (RBDs), which enable the recognition of substrate RNAs [16,17]. Examples of well-studied RBDs include the RNA recognition motif (RRM) [18], the heterogeneous nuclear ribonucleoprotein K homology domain (KH) [19], the double-stranded RNA-binding motif (dsRBM) [20] and the zinc finger domain [21], to name a few. RNA-binding domains are also typical components of proteins involved in the formation of large RNP complexes such as the ribosome or the spliceosome [22,23] and they also occur in proteins that regulate the function of RNAs [24]. Proteins that simultaneously bind multiple sites in RNA often include multiple RBDs [24,25]. Here, we point our readers to a review from the Varani group, on the various RNA-binding strategies of RBPs that exploit the modular nature of RBDs [16]. Apart from RBDs with well-defined 3D structures, RNAs can also be recognized and bound by structurally disordered regions, as in ribosomal proteins, which assume folded conformation only upon binding RNA [26].

Three dimensional structures of RNP interactions have provided important insights into the molecular intricacies governing these interactions, including the specificity of mutual recognition by protein and RNA components and assist in studying the physicochemical principles of RNP interactions. Although 3D structure-derived information is important for understanding biological roles of RNP interactions, experimental determination of RNP complex structures is a slow and laborious process [27,28]. Firstly, many RNA-protein interactions are transient resulting in formation of short-lived complexes. Secondly, there are difficulties associated with the chemical character of the RNA component(s) of the complex. RNA is conformationally more flexible than proteins, and RNA molecules are often structurally heterogeneous. In addition, RNAs are often elongated in shape and in contrast to proteins, exhibit few elements that can stabilize crystal contacts hindering crystal packing. Furthermore, the negatively-charged sugar-phosphate backbone contributes to the repulsion between the molecules. These factors collectively make RNA and RNP structure determination more challenging than protein structure determination.

Until the end of the 20th century there were only a handful of high-resolution structures of RNP complexes available but the number of such structures has exponentially increased over the past decade due to significant improvements in established techniques such as X-ray crystallography [29], as well as the advent of newer technologies like electron microscopy (EM). Several groups developed hybrid techniques and for instance combined the nuclear magnetic resonance (NMR), small angle scattering (SAS), analytical ultracentrifugation (AUC), and/or electron paramagnetic resonance (EPR) experiments [30–32]. Despite all these advances, the number of structures for RNP complexes is much lower compared to that of DNA-protein complexes. As of April 2018, 4265 DNA-protein complex structures were available in the Protein Data Bank (PDB). On the other hand, only 2194 macromolecular complexes involving both protein and RNA components (but excluding RNA/DNA hybrids) were available in the PDB. Of these structures, 1642 were solved by X-ray crystallography, 426 by EM, 120 by solution NMR spectroscopy and 6 by other methods such as fibre diffraction. These structures contained 51,101 protein chains interacting with RNAs but many proteins were highly similar to each other. After removing redundant protein chains with sequence identity >90% or >40%, only 2862 or 1302 experimentally-determined RNA-bound proteins respectively, remained.

Over the past decade, various research groups have utilized different tools and techniques to estimate the total number of RBPs in the human proteome. The database of RNA-binding protein specificities (RBPDB) database documents 397 manually curated examples of RBPs from the literature [33]. In 2012, independent reports from the Landthaler and Hentze groups identified

797 and 860 mRNA-binding proteins (mRBPs) in the human proteome, from the HEK293 and the HeLa cells, respectively [34,35]. Many of these proteins lack typical RBDs or motifs, or are known to exhibit other functions not related to RNA metabolism [36]. On the one hand, this could indicate that many proteins take part in RNP formation as a part of their life cycle; while on the other hand, such large-scale experimental screens may identify proteins that interact with RNA indirectly, for example, as components of larger protein-protein complexes. The Zhou group used the approach of fold recognition for protein structure prediction and the SPOT-seq technique for binding affinity-based RNA-binding prediction. Using these techniques, they identified 2937 RBPs in humans [37], with a 43.6% coverage of the experimentally reported mRBPs by the Hentze group [35]. More recently, the Tuschl group has consolidated a list of 1542 RBPs in the human genome that have been identified by a combination of bioinformatics approaches and curation of experimental data from literature [38]. Work done by the Sowdhamini group has resulted in the computational prediction of 2625 RBPs in the human proteome [39]. The hRBPome database compares and contrasts the RBPs as reported by the above studies [40]. In 2016, work done by the Preiss, Ørom and Ostareck-Lederer groups have reported 1148 RBPs in cardiomyocytes, 382 RBPs in the nucleus and 402 RBPs in macrophages, respectively [41–43]. Obviously, for the majority of these complexes, no structural data is available.

Given the scarcity of experimentally-determined structures of RNP complexes, computational techniques can complement existing data to help elucidation of RNP complex 3D structures. However, while the methodology for prediction and modelling of 3D structures for individual proteins and protein-protein complexes is very well established [44–49], there are much fewer methods for predicting and modelling 3D structures of RNA molecules and RNP complexes [50–53]. In this article, we review computational approaches for modelling of RNP complex structures. We focus on RNP docking, scoring functions and on methods for evaluating the accuracy of predictions, in particular, docking benchmarks and affinity datasets.

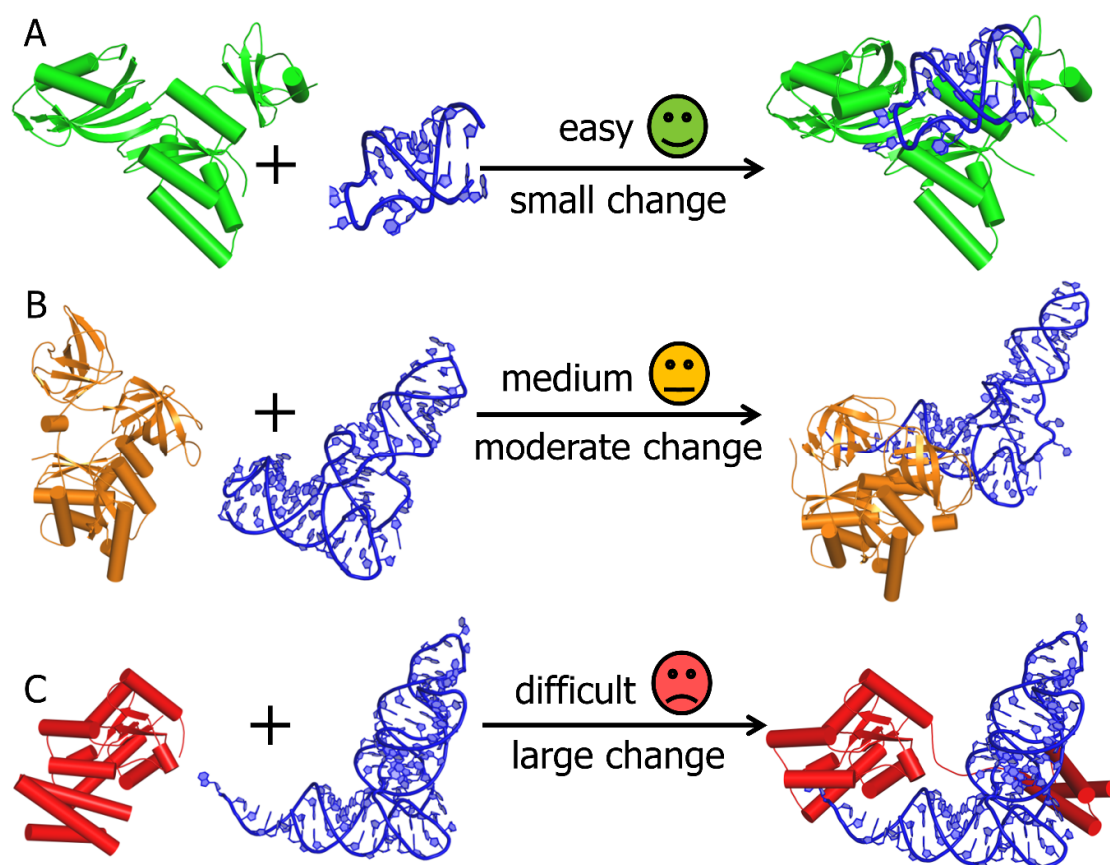
## 2. Computational Modelling of RNA-Protein Complex Structures

The prediction of 3D structures of macromolecular complexes is usually done by the computational docking approach, which requires detailed knowledge of structures for all individual components of the complex. Ideally, the docking should be based on the knowledge of high-resolution atomic structures of the components, for example, determined by X-ray crystallography or NMR. However, in many cases, experimentally determined structures of components of the complex are not available. For many applications, these can be substituted by computationally-modelled structures. To this end, a large number of computer programs have been developed, which now allow for reasonably accurate and practically useful predictions of protein 3D structures and reviewing them is beyond the scope of this article. The state-of-the-art in protein 3D structure prediction has been systematically assessed by the Critical Assessment of protein Structure Prediction (CASP) experiment [54]. More recently, the RNA Puzzles experiment has been initiated to assess the state-of-the-art in RNA 3D structure prediction [54,55]. The reader is referred to the most recent articles describing progress in these areas, for example, CASP Round XII [54–56] and RNA Puzzles Round III [57].

In computational biomolecular docking, the receptor refers to the larger molecule, while the ligand refers to the smaller one. The docking protocol comprises two steps: (i) conformational sampling, that is, searching for possible conformations and mutual orientations of the docking components that leads to the generation of docked models (called poses or decoys) and (ii) scoring of docked poses: assessing them by a mathematical function that aims to distinguish between models with different degrees of similarity to the unknown “true” structure [58]. Some of the existing docking methods combine both sampling and the scoring steps [59–61], while others only specialize in the assessment of docked poses [62–65].

A major challenge in computational docking is related to the observation that structures of binding partners often undergo conformational changes during association, in a process known as induced

fit. Despite the recent advancement of methods that take macromolecular flexibility into account [66], dealing with conformational changes involving backbone and loop rearrangements still remains the biggest challenge in the modelling of macromolecular complexes [67] (Figure 1). Docking methods differ in the details with which they model conformational changes in the receptor and the ligand. Certain methods model such changes explicitly, making such analyses computationally demanding, whereas the other set of programs focus less on molecular details and introduce a certain level of ‘fuzziness’ [68]. This can be addressed by generating ensembles of conformers. Different research groups have adopted their own strategies towards accomplishing this, such as molecular dynamics (MD) simulations, Monte Carlo (MC) simulations, Normal Mode Analysis (NMA) and use of PDB structure homologs [67].



**Figure 1.** Overview of challenges in RNA-protein (RNP) docking. (A) “easy” docking of tRNA pseudouridine synthase B (1R3F:A) and a small RNA fragment (1K8W:B); the protein undergoes a small conformational change to form the RNP complex (1K8W:A-1K8W:B). (B) “medium difficulty” docking of the Tu elongation factor (1TUI:A) and cysteinyl tRNA (1U0B:A); both components undergo a moderate conformational change to form the RNP complex (1B23:P-1B23:R). For many of the currently available docking tools, it is challenging to model this degree of conformational change. (C) “difficult” docking of L-seryl-tRNA (Sec) kinase (3A4M:A) and selenocysteine tRNA (3ADB:C); the protein undergoes a large conformational change movements to form the RNP complex (3ADB:A-3ADB:C). For most of the currently available docking tools, it is nearly impossible to model such large conformational changes.

### 3. RNP Docking Methods (Conformational Sampling with or without Scoring)

Many protein-protein docking methods have been assessed in the Critical Assessment of PRediction of Interactions (CAPRI) experiment, analogous to CASP [69]. In comparison to protein-protein docking, RNP docking has received less attention from computational method developers. However, the number of groups participating in the CAPRI for scoring RNP complexes, as

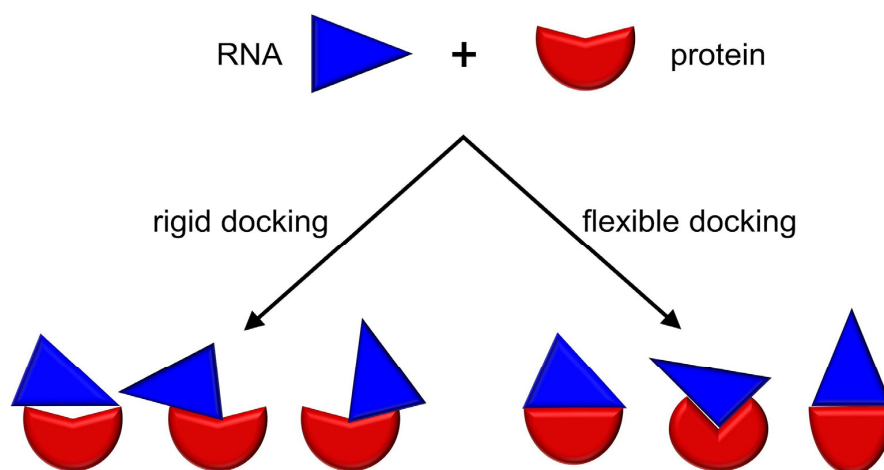
well as the number of available methods for RNP docking, has increased steadily over the years [67,70]. Most of the methods for RNP docking have been developed as modifications to existing protein-protein docking methods, in order to accept nucleic acids as receptors and/or ligands. In order to adapt protein-protein docking methods to RNP docking, the following modifications are necessary: (i) a representation for RNA molecules has to be added to the docking algorithm for data handling purposes (some of the protein-protein docking methods handle only amino acid residues) and (ii) the scoring function has to enable evaluating RNA-protein interactions (some of the protein-protein docking methods can handle non-protein molecules such as RNA but they are unable to take into account RNA-specific interactions). Table 1 lists various RNP docking tools that are freely available as standalone programs or web servers.

**Table 1.** Comparison of existing RNP docking methods. The majority of these methods are modified from existing protein-protein docking methods. The type of docking algorithm (rigid or flexible) and their availability (web server and/or standalone) are indicated.

Name	Modified from Protein-Protein Docking Method	Docking Method (Rigid/Flexible)	Availability		References
			Web Server	Standalone	
3dRPC	×	Rigid	✓	✓	[64,71]
ClusPro	✓	Rigid	✓	×	[59]
FTDock	✓	Rigid	×	✓	[72]
GRAMM	✓	Rigid	✓	✓	[73]
Hex	✓	Rigid	✓	✓	[74]
ICM	✓	Rigid	×	✓	[75]
NPDock	×	Rigid	✓	×	[60]
PatchDock	✓	Rigid	✓	✓	[76]
PEPSI-DOCK	✓	Rigid	×	✓	[77]
pyDock	✓	Rigid	✓	✓	[78]
RosettaDock	✓	Rigid	✓	✓	[79]
ZDOCK	✓	Rigid	✓	✓	[80]
ATTRACT	✓	Flexible	✓	✓	[81]
HADDOCK	✓	Flexible	✓	✓	[61,82]
HDOCK	×	Flexible	✓	×	[83]
PIPER	✓	Flexible	×	✓	[84]
Prime	✓	Flexible	×	✓	[85]

Existing docking methods can be divided into two general classes: (i) rigid and (ii) flexible (Figure 2). Upon RNP complex formation, protein, RNA or both may undergo conformational changes in the backbone (large-scale domain motions and movements in disordered regions) and/or the sidechains. Flexible docking methods attempt to account for these conformational changes in order to predict near-native biological associations. Rigid docking methods do not account explicitly for conformational changes in the structure of the input protein and/or RNA but they may represent these structures in a ‘fuzzy’ way to embody the uncertainty of the bound conformation.

Rigid body docking methods are usually the first method of choice, especially if little is known beyond the structures of the components. By virtue of compromising on computing cost for conformational flexibility, such methods are capable of exploring a bigger search space to identify potential binding sites on the protein and RNA molecules. For example, FTDock developed by the Katchalski-Katzir group generates orthogonal grids to represent the components for docking and then performs a global scan of the search space (translational and rotational) to generate a quick approximate solution of the prediction problem [72]. The results can be, however, only as accurate as allowed by the difference between unbound and bound conformations. In other words, any conformational changes associated with binding contribute to the deterioration of the accuracy of models generated by rigid docking. Thus, apart from the first quick screen of docking solution, rigid docking is recommended only for cases where no major conformational changes are expected to occur.



**Figure 2.** Comparison of rigid and flexible docking methods. A protein and an RNA molecule have been schematically represented as a red and a cyan figure, respectively.

Flexible docking can simulate conformational changes; however, its utility is also correlated with the degree of the structural change between the unbound and bound forms. It is a method of choice for cases in which the conformational search space is relatively small or for the purpose of local structure refinement. In fact, none of the typical docking methods can reliably predict the structures of complexes that undergo large and complex conformational changes during binding—for this, multi-scale methods are a recommended solution (see below). An example of a popular flexible docking method is HADDOCK, developed by the Bonvin group. It can take the information from user-defined restraints (for example, from experimental data) to drive the docking [86], as well as take into account the flexibility of both the protein and the nucleic acid [87]. Unlike many other methods, HADDOCK can use the nucleic acid both as a receptor, as well as a ligand. A user-defined set of residues are allowed to be conformationally flexible during the docking process. Thus, HADDOCK is capable of performing well in cases where there are local rearrangements in and around these residues. The program is available as a web server [61] and also as a standalone tool [82]. HDock, developed by Huang group, differs from the other existing methods, in being able to accept both protein structures and sequences as input [83]. However, it can only accept input RNA structures (and not sequences) for the docking protocol. In case of a protein sequence input, HDock performs a sequence similarity search against the PDB using the HHSuite package [88] to identify a homologous structure, which is then used as the template for a subsequent MODELLER-based modelling step [89]. The theoretical model of the protein structure is then used for the docking step.

#### 4. Other Methods for Three Dimensional Structure Prediction of RNP Complexes

The 3D structures for RNP complexes can be modelled by methods that do not involve explicit docking, or in which the docking is only a minor component of calculations.

Template-based modelling of complexes (in which the structure of one complex is modelled based on the structure of a related complex used as a template) is widely used for protein-protein interactions [90] but less investigated for the modelling of RNP complexes. Protein RNA Interaction Modelling (PRIME) [91] is one such method for predicting RNP complex structures, based on structurally similar complexes with experimentally determined structures. The structural similarity of the individual components to the those in the “template” structure is important to ascertain the binding mode with certainty. It performs well in cases where “free” docking fails (demonstrated by the authors by testing the method on the RNP docking benchmark set of the Fernandez-Recio group) and is capable of accounting for conformational changes upon binding [91]. RStrucFam, from the Sowdhamini group, is a method for predicting cognate RNA partners for RBPs, based on their sequence similarities with the protein components of known RNP complex structures [92].

Another strategy in the prediction of RNP complex structures is MD simulations [93]. The protein and RNA can be driven to bind by placing them in appropriate force fields with restraints but in general, MD is incapable of accurately simulating binding events and large conformational changes that occur on time-scales larger than microseconds [94]. On the other hand, near-native docked poses generated by macromolecular docking can be used as starting structures for MD, allowing the system to readjust to the minimal energy. Moreover, including MD force fields in scoring can improve docking [95]. Work done by the Kameda group in 2016 demonstrated an improved accuracy for RNP complex prediction by a rigid protein-protein docking method [95], ZDOCK, which originally was not parameterized for RNA. The following modifications in the ZDOCK method were introduced by the authors: (i) derivation of the physicochemical properties of nucleic acids (such as partial charge of atoms, Van der Waals radii etc.) from MD simulation force fields for nucleic acids (such as, AMBER and CHARMM) and (ii) introduction of electrostatic interactions into the scoring function based on the AMBER94 force field. The Wang group applied a combination of docking and MD followed by binding energy calculations to identify the binding mode of RNA to carbon storage regulator A protein (CsrA), which was previously unknown [96]. The authors calculated binding free energies using molecular mechanics combined with the generalized Born and surface area continuum solvation (MM/GBSA) method [97]. This study has identified the binding pockets in CsrA that could be targeted by small molecules to prevent RNA binding, without interacting with the RNA [96].

Another class of methods have been developed to model multi-scale resolution models of large macromolecular complexes. These methods often represent components as rigid bodies or as flexible forms, depending on the availability of information about their structure and dynamics. They usually rely on additional information encoded in the form of spatial restraints to define interactions between the components. The M3 framework from the Carlomagno group uses sparse and hybrid experimental data to model structures of macromolecular complexes, starting from 3D structures of individual domains, monomers and subcomplexes. These components can undergo rearrangements and yet retain their overall 3D structure upon complex formation [98]. The experimental data are used to derive interatomic distances and/or molecular shapes, to guide the modelling process. The Integrative Modelling Platform (IMP) software package from the Sali group [99] is an example of a hybrid method. It performs modelling of simple biomolecules, as well as large macromolecular assemblies, by integrating data from various experiments. The preparation of the input data for IMP requires advanced knowledge of the system analysed. Recently, our group developed PyRy3D (<http://genesilico.pl/pyry3d/>), a multiscale modelling method that enables the construction of models for very large macromolecular complexes with components of known or unknown 3D structure. PyRy3D aims at cases of data-driven modelling of complexes where the users have only limited knowledge about the system analysed and wish to test alternative hypotheses. PyRy3D applies a MC search to sample the space of solutions restricted by various spatial restraints that determine 3D shapes or interactions between subunits of complexes. It has been used to model RNPs such as the complex between the 2'-5'-oligoadenylate synthetase OAS1 and the 3'-terminal region of the West Nile Virus (WNV) RNA genome [100].

## 5. Standalone Scoring Methods for RNA-Protein Complexes

For the modelling of RNP complex structures, it is often useful to consider predictions from different modelling methods, followed by a ranking of the alternative poses using external scores. Scoring functions are essential for distinguishing between models of various accuracy, in particular, to discriminate between models that are close enough to the “true” structure to provide useful functional information and models that are inaccurate and could lead to wrong conclusions. Standalone scoring methods are invaluable for comparison of models obtained with different modelling approaches that rely on different internal scoring functions. The scoring of 3D structural models may be based on the following considerations: (i) verification to what extent the given decoys agree with independently

obtained information (for example, with experimental data or with independent computational predictions) and (ii) evaluation of the overall quality of fitting between the protein and RNA structures.

For scoring of models with an independent dataset, a large number of approaches exist. A variety of experimental methods can be applied to study RNA, protein and RNA-protein complex structures, which can be translated into spatial restraints. Furthermore, numerous computational methods are available for predicting RNA-binding residues in proteins [101,102]. Once the models of RNP complexes become available, scoring based on external information (including experimental data) can be performed with various computational tools. As an example, FILTREST3D [103] allows scoring of decoys based on a combination of distance restraints with other factors such as local or global structure or molecular shape. It also allows the use of logical operators to enable sets of alternative restraints. pyDockRST also uses the percentage of satisfied distance restraints derived from experimental data, along with electrostatics and desolvation binding energy, to score and rank docking poses [104].

In the absence of independent data, structural models of RNP complexes can be evaluated with generic scoring functions, which can be broadly categorized into three types: force field, empirical and knowledge-based. Force fields are the functional form and parameter sets used to calculate the potential energy of a system of atoms or coarse-grained points. Empirical scoring functions are based on counting the number of various types of interactions at the interface of RNP complexes. Knowledge-based scoring functions, also known as statistical potentials are mathematical functions derived based on statistical observations of interactions at the interface of known RNP complexes. A list of methods available for computational scoring of RNP complexes is presented in Table 2.

**Table 2.** List of scoring methods for RNP docking. The representation of the molecules (all-atom or coarse-grained), the type of statistical function and the availability of these methods (web server and/or standalone) have been listed in this table.

Name	Structure Representation	Scoring Method	Decoy Discrimination Threshold (RMSD)	Availability as a Standalone Tool	Reference
Varani's H-bonding potential	All-atom	H-bonding potential	<3 Å	×	[105]
Varani's all-atom potential	All-atom	All-atom distance-dependent	<5 Å	×	[106]
Fernandez's potential	Coarse-grained	Pairwise residue-ribonucleotide propensity	<10 Å	×	[107]
dRNA	All-atom	Volume-fraction corrected DFIRE energy function	NA *	×	[108]
DARS-RNP and QUASI-RNP	Coarse-grained	Quasi-chemical potential and decoys as the reference state potentials	<10–15 Å	✓	[62]
Zacharias' potential	Coarse-grained	Distance-dependent, coarse-grained force field for protein–RNA interactions.	<8 Å	×	[109]
Wang's potentials	Coarse-grained	Pairwise residue-ribonucleotide propensity with secondary structure information	<10 Å	×	[63]
Deck-RP	Coarse-grained	Distance and environment dependent	<15 Å	✓	[64]
ITScore-PR	All-atom	Pairwise distance dependent atomic interaction potential	<10 Å	✓	[65]
RPRANK	Coarse-grained	Pairwise residue-nucleotide RMSD	<10 Å	✓	[71]

\* data not available



The Varani group was one of the first groups to develop scoring functions, for RNA-protein interactions. Initially, they developed a statistical potential based on hydrogen bonding (H-bond) at the RNP interfaces [105]. However, H-bonds represent only a portion of various types of interactions occurring at the RNP interfaces [110,111]. Later, the same group developed an all-atom, distance-dependent statistical potential for predicting sequence-specific recognition between RNA and protein [106]. The Varani's all-atom potential treats the interactions between chemically similar atoms (based on the CHARMM atom definition) in the same way and as a result, it contains only a distance-dependent multiple bin term. A coarse-grained, distance-dependent pairwise residue-ribonucleotide propensity was derived by Fernandez-Recio group to score the docking poses [107]. In their approach, the entire residue is represented as a single interaction centre and therefore it uses a single bin (i.e., the presence or absence of interaction) to calculate the potential. Zhou and co-workers developed dRNA [108]; a volume-fraction corrected distance-scaled, finite ideal-gas reference (DFIRE) energy function for RNP interactions [112]. However, the dRNA method requires known RNP complex structures as templates and has limited applications when the RBPs used have novel binding modes different from those in the template structures. Our group developed two knowledge-based potentials: the quasi-chemical potential (QUASI-RNP) and the Decoys As the Reference State potential (DARS-RNP) [62]. These statistical potentials use a coarse-grained representation, which ignores molecular details and hence is insensitive to minor conformational inaccuracies. The reference states used in these two potentials differ. While QUASI-RNP uses mole fractions of residues as a reference state, DARS-RNP uses decoys. Both these potentials have the same mathematical base except for the reference states and use multiple bins for distances as well as orientations.

The Zacharias group developed a distance-dependent, coarse-grained force field for RNP interactions [109]. During a testing of this potential, the authors observed during the testing of this potential that it allowed moderate conformational changes. The potential allows protein-RNA docking followed by energy minimization in the rotational and translational degrees of freedom of the binding partners [108]. Wang and co-workers developed four different pairwise residue-nucleotide potentials. These potentials were derived from the pairwise residue-nucleotide propensities with or without considering the protein and/or RNA secondary structural elements [63]. The authors concluded that the RNA secondary structure information contributed more significantly than the protein secondary structure in discriminating the correct 3D structures of RNP complexes.

The Xiao group developed a scoring function, Distance- and Environment-dependent Coarse-grained and Knowledge-based potential for RNP complexes (DECK-RP), to evaluate the docked poses generated by RPDock [64]. DECK-RP combined the advantages of both Wang's potential [63] and DARS-RNP [62]. The secondary structure context was considered for calculating the pairwise propensities. DECK-RP uses a reference state, which includes a decoy-based component and a mol-fraction corrected component. The Zou group developed a scoring function, ITScore-PR, based on the atomic distance-dependent potentials derived from known RNP complex structures, which uses a physics-based iterative algorithm [65]. The authors demonstrated that many algorithms performed better for rigid body docking in which the components were derived from bound structure rather than for flexible docking where the components were derived from the unbound structures. The Xiao group developed a new knowledge-based potential, RPRANK. The conformation differences between residue-base pairs between standard pairs from native structures and decoys were used to calculate the statistical potential [71].

The approximate utility of various potentials, with respect to the accuracy of docking decoys analysed, is indicated as a decoy discrimination threshold in Table 2. For instance, some of the fine-grained or high-resolution (all-atom) methods are useful for discriminating decoys that exhibit root mean square deviation (RMSD) to the reference structure  $<3$  or  $<5$  Å (decoys that are less accurate exhibit essentially random score, not related to the accuracy) and some coarse-grained or low-resolution

methods can discriminate decoys up to ~10–15 Å from the reference structure but they are usually not appropriate for discriminating between decoys that are very close to the native structure.

The availability of different docking methods and scoring functions allows for various combinations to be applied. Matching the docking procedure with a scoring function of a similar resolution is recommended. For example, we developed NPdock, a web server for low-resolution RNP structure prediction [60], which first performs rigid body RNP docking using GRAMM [73] and then scores the decoys with different coarse-grained statistical potentials QUASI-RNP and DARS-RNP, also developed by our group [62]. Likewise, pyDockWEB generates rigid-body docking orientations by FTDock and evaluates them by the pyDock scoring function [78]. On the other hand, high-resolution potentials that operate at the atomic level (e.g., the ones developed by the Varani group) are expected to work well only with docking poses that are expected to be very close to the native structure, which often can be generated only by the flexible docking methods that allow for precise modelling of atomic interactions.

An important consideration in the selection of decoys is not only scoring individual models but also finding ensembles of similar models [113]. In general, an ensemble of structurally similar models with very good scores is often indicative of a more accurate prediction than a single model with the best score. However, several clusters of solutions with similar scores may exist. Hence, docking methods often report not just one best-scored model but representatives of several largest clusters (usually up to three or five). Various clustering strategies exist and, so far, no comprehensive study was performed to identify the best approach that could work with decoy datasets generated with different docking methods and scored with different functions. The strategy adopted in our method NPdock is based on grouping of decoys with low RMSD values. First, an RMSD matrix is generated for all pairs of decoys. Second, the structures with RMSD values below a specified cut-off value are grouped into one cluster, which is removed from the dataset and the process is iterated. The program reports the medoids of the biggest three clusters as well as the model with the overall best score [62].

## 6. RNA-Protein Three-Dimensional Structure Datasets for Benchmarking the Computational Docking Methods and Their Applications

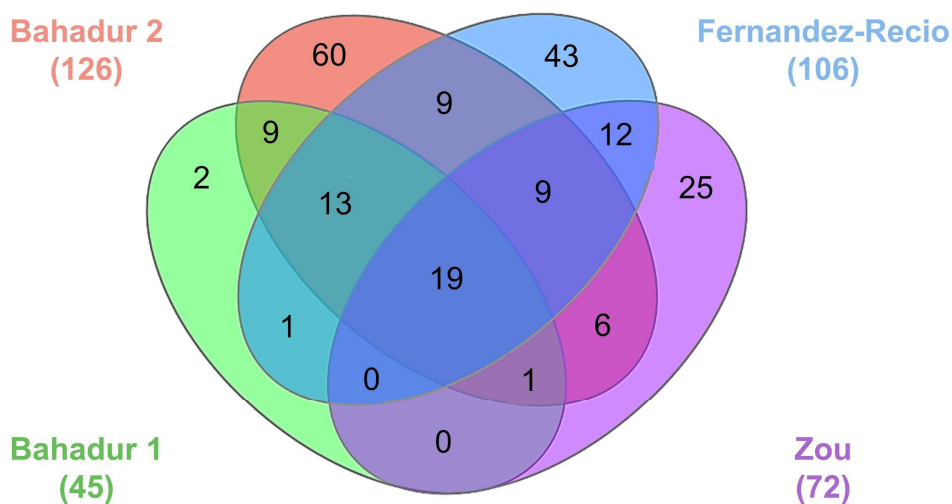
Evaluating the performance of existing docking and scoring methods requires datasets of 3D atomic coordinates of RNP complexes that serve as references (“true” structures) for validating the predictions. In this section, we discuss the various benchmark datasets curated by different groups (Table 3, Figure 3). The different benchmarks employ different criteria for selecting the complex structures and the corresponding unbound structures. The first benchmark dataset for RNP docking was assembled by the Bahadur group, which only included structures determined by X-ray crystallography [114]. This benchmark consists of 45 non-redundant RNP complexes and their corresponding unbound structures. There are nine unbound-unbound test cases for which both the protein and RNA are available in an unbound form and 36 unbound-bound test cases for which only protein is available in unbound form (Table 3). This benchmark is divided into four structural classes: (A) complexes with tRNA (16 cases), (B) complexes with ribosomal proteins (three cases), (C) complexes with duplex RNA (10 cases) and (D) complexes with single-stranded RNA (16 cases). In addition, this benchmark divides the dataset into three categories based on the conformational changes undergone by the interface C $\alpha$  atoms: (R) rigid body ( $i\text{-rmsd}_{C\alpha} < 1.5 \text{ \AA}$ ) with 34 cases, (S) semi flexible ( $1.5 \text{ \AA} \leq i\text{-rmsd}_{C\alpha} \leq 3.0 \text{ \AA}$ ) with eight cases and (X) full flexible ( $i\text{-rmsd}_{C\alpha} \geq 3.0 \text{ \AA}$ ) with three cases. Here,  $i\text{-rmsd}_{C\alpha}$  is defined as the RMSD of the interface C $\alpha$  atoms after superposing the bound and unbound structures. The authors have calculated the interface residues using the PRince webserver [115]. The benchmark from the Bahadur group was later updated with 126 non-redundant RNP complexes [116]. This includes 21 unbound-unbound cases, 95 unbound-bound cases and 10 are bound-unbound cases. Of the 21 unbound-unbound types, 12 are pseudo-unbound where the RNAs are taken from a different RNP complex. The dataset is divided into four structural classes with 28, 5,

40 and 53 cases in classes A, B, C and D, respectively. The current version of this benchmark consists of 72, 25 and 19 cases in R, S and X categories.

**Table 3.** List of RNP docking benchmarks. The number of unbound-unbound, unbound-bound and bound-unbound test cases is listed in this table.

Benchmark	Number of Test Cases	Unbound-Unbound	Unbound-Bound	Bound-Unbound	References
Bahadur group 1	45	36	9	0	[114]
Bahadur group 2	126	95	21	10	[116]
Fernandez-Recio group	106	81	25	0	[117]
Zou group	72	52	17	3	[118]

An extended benchmark including both experimental structures and homology models was curated by the Fernandez-Recio group with 106 test cases. In this benchmark dataset, 71 out of 106 entries were taken from crystallography or NMR experiments, while 35 entries were built using homology modelling [117]. Of the experimental structures available, there are nine unbound-unbound cases and 62 unbound-bound forms. Among the nine unbound-unbound cases, four cases are pseudo-unbound where the RNAs are taken from a different RNP complex. The homology-modelled cases in this benchmark consist of five unbound-model, eight model-unbound, 19 model-bound and three model-model protein-RNA cases. The benchmark is divided into three categories for docking predictions based on the conformational changes undergone by the C $\alpha$  and P atoms at the interface of RNP complexes: easy ( $i\text{-rmsd}_{C\alpha+P} \leq 2.5 \text{ \AA}$ ) with 64 cases, intermediate ( $2.5 \text{ \AA} \leq i\text{-rmsd}_{C\alpha+P} \leq 5.0 \text{ \AA}$ ) with 24 cases and difficult ( $i\text{-rmsd}_{C\alpha+P} > 5 \text{ \AA}$ ) with 18 cases.



**Figure 3.** The relationship between RNP docking benchmarks datasets. Nineteen RNP structures are common to all four benchmarks. Two hundred nine RNP structures are represented by all four benchmarks together.

The Zou group developed another docking benchmark for RNP complexes with 72 cases, of which, 52 are unbound-unbound cases, 17 are unbound-bound cases and three are bound-unbound cases [118]. In bound-unbound cases, only RNA structures are available in an unbound form. Based on the conformational changes undergone by the interface C $\alpha$  and C4' atoms and the fraction of native contacts this dataset is classified into three categories: easy ( $i\text{-rmsd}_{C\alpha+C4'} \leq 1.5 \text{ \AA}$  or  $f_{\text{nat}} \geq 0.8$ ) with 49 cases, medium ( $1.5 \text{ \AA} < i\text{-rmsd}_{C\alpha+C4'} \leq 4.0 \text{ \AA}$  and  $0.4 \leq f_{\text{nat}} < 0.8$ ) with 16 cases and difficult ( $i\text{-rmsd}_{C\alpha+C4'} > 4.0 \text{ \AA}$  or  $f_{\text{nat}} < 0.4$ ) with seven cases.

The docking benchmarks discussed in this section were widely used, i.e., to test docking methods [79,83], to develop knowledge-based scoring functions for studying RNP interactions [65,119,120], to predict RNA-binding sites in proteins [115,121,122], to investigate the role of water molecules [123], to find the binding hot-spots and to predict binding affinities [124]. Besides, the availability of bound and unbound structures assists in the development of new physicochemical and structural parameters for quantifying the changes occurring in RNA-protein interaction sites upon binding [110,125–127].

A number of studies covering the general properties of RNP complex structures used the benchmark dataset developed by the Bahadur group, i.e., due to the stringent criteria on the resolution (better than 3.0 Å) and redundancy (sequence identity  $\leq$  35%). Studies of RNA-protein binding sites of complexes in this dataset included the analysis of the role of water molecules in RNP complex formation [123], the study of sequence conservation at RNP interfaces [124] and the quantification of solvent accessibility at interfaces of RNP complexes [125]. The Zou group benchmark used a more relaxed resolution cut-off of 4.0 Å compared to other docking benchmarks. It included more low-resolution structures, which may not be suitable for the development of high-resolution scoring functions or evaluation of physicochemical parameters. However, the low-resolution structures may be useful as templates for comparative modelling or for development of coarse-grained methods. The benchmark developed by Fernandez-Recio group went even further, beyond pairs of bound and unbound structures determined experimentally for the same protein and RNA molecules, to include structures obtained by homology modelling in cases, where experimental structural data existed only for related components [117]. This dataset allows to address some challenging problems, for which low-resolution structures are sufficient. However, caution must be applied while using these structures, as errors in structures obtained by modelling may affect the conclusions.

## 7. Datasets for Ribonucleic Acid-Protein Binding Affinity Prediction and Their Applications

RNA-Protein interactions are often affinity-driven processes, where the specificity of binding is determined by the conformation adopted by the molecule to bind with its partner. The free energy required for RNA and/or protein molecules, to adopt the particular conformation required for binding, is a determining factor in such binding events.

The first RNP affinity benchmark was developed in 2013 by the Liu group with 73 cases [128]. A comprehensive and up-to-date affinity benchmark for RNP complexes is still missing. A dataset of affinity values for alanine substitutions in protein components of RNP complexes was curated by Bahadur group for 14 RNP complexes [124]. The changes in affinity values upon alanine substitutions is indicative of the role of the corresponding residue in the binding process. The knowledge of affinity values helps in defining the active residues for guided docking of RNP complexes. The dataset reports 94 experimental affinity values for 14 native structures and 80 variants bearing single residue alanine substitutions [124]. This dataset was later expanded by the Deng group to include 49 RNP complexes [129]. The dataset reports 334 experimental affinity values for 49 native structures, 254 variants with alanine substitutions and 31 other substitutions. The dbAMEPNI dataset curated by Mitchell and Zhu groups reports affinity data for 51 RNP complexes [130]. The dataset includes experimental affinity values for 193 alanine substitutions in RNPs.

It is important to draw our readers' attention to the fact that the bioinformatics methods discussed in this review are for predicting the possibilities in which a given protein and an RNA interact with each other and they assume that their interaction does happen. These methods are not appropriate for ascertaining if the protein and the RNA interact or not, or for predicting which RNA binds to a given protein or vice versa. Predicting the binding affinity would help us determine whether the molecules bind to each other. Methods for computational prediction of binding affinities are however still in their infancy. They require high-quality datasets of experimentally determined affinity values, which are currently sparse. The affinity dataset curated by the Bahadur group was used in the development of an algorithm to predict hot-spot residues within RNA-binding sites and is available as a web server

HotSPRing [124]. The same dataset was used by Pires and Ascher to develop the mCSM-NA web server, which uses graph-based signatures to predict the impact of single residue substitutions on nucleic acid binding affinity [131]. The Deng group have developed PrabHot, another method to predict the hot spot residues at RNA-protein interfaces [129].

## 8. Conclusions

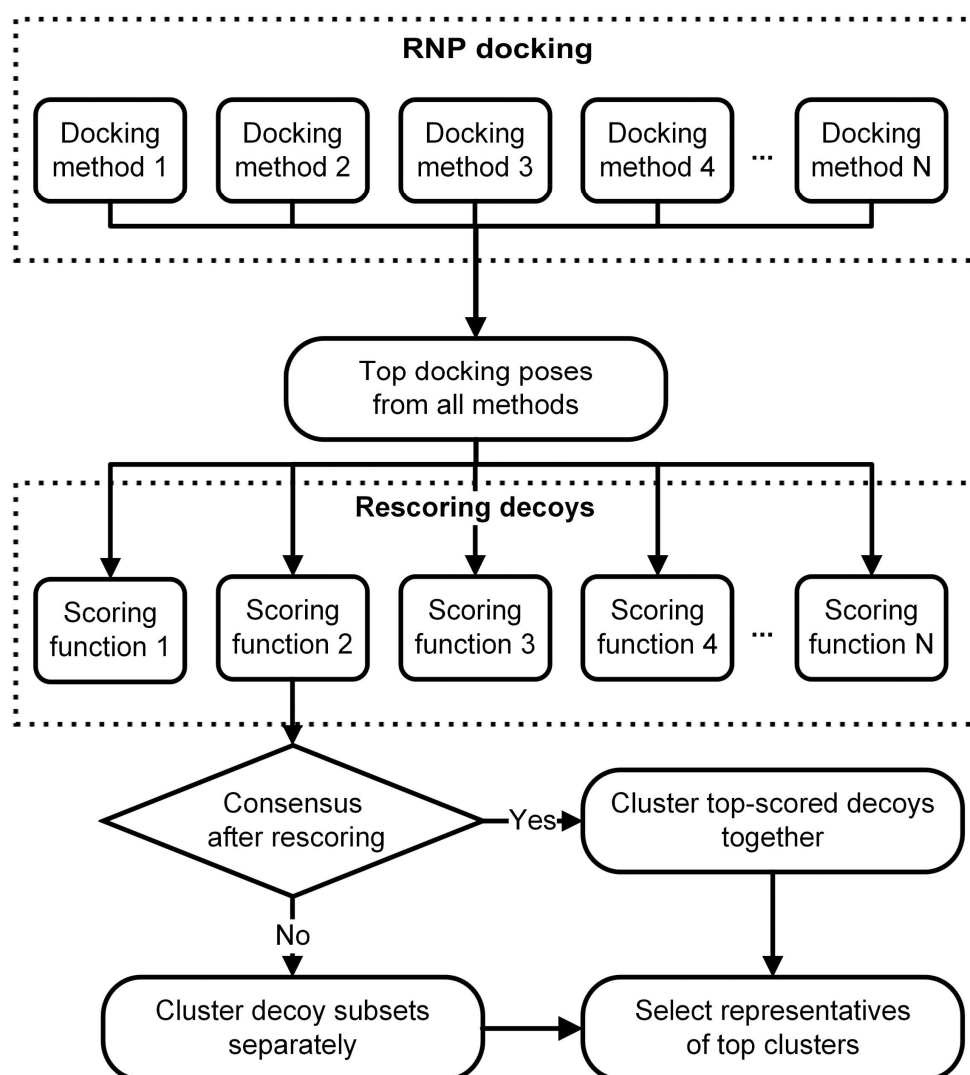
Over the past decade, there has been a growing interest in investigating RNP interactions. This is apparent from the increasing number of structures of RNP complexes that have been deposited in the PDB database per year (197 in 2017 vs. 87 in 2007), as well as a greater number of publications that appear each year in the PubMed database, with the keyword “RNA-binding proteins” (1405 in 2017 vs. 1278 in 2007). However, computational methods for prediction of RNP complex structures using information from structures of the individual components or directly from sequences, have been sought after due to the difficulties associated with the experimental determination of their structures.

Computational prediction of RNP 3D structures can provide important information in cases where standard approaches for experimental structure determination fail. In our own work, we encountered numerous RNP systems, for which we or our collaborators attempted to crystallize the complex but failed to obtain diffraction-quality crystals. Consequently, the available data (in some cases including the crystal structure of the protein partner in the apo form) were used to guide macromolecular docking and modelling of RNA-protein interactions, providing functional insight that could not be obtained from structures of the components in isolation from each other. One example includes the BsMiniIII endonuclease, which crystallized only in the apo form but not in complex with its dsRNA substrate and for which an RNP complex structure was modelled [132] and used to guide the successful engineering of substrate preference [133]. For CMTr2 methyltransferase, we were unable to obtain sufficient amounts of protein for crystallization and had to model the structure of the complex with a 5'-capped RNA substrate to obtain insights into the mechanism of substrate recognition [134]. The structure of the archaeal tRNA methyltransferase Trm10 was obtained only in the RNA-free form and the protein-tRNA complex had to be modelled, with additional experimental data as restraints [135]. Finally, for several bacterial rRNA methyltransferases including ErmC' [136] RlmH [137] and NpmA [138] only the protein structure was known and while the ribosome structure was known, the determination of the RNA-protein (or protein-ribosome) complex structure has proven unsuccessful, hence the structural insight into their mechanism of action had to be obtained by docking/modelling.

The approaches for modelling RNP complexes discussed in this paper, though capable of providing practically useful predictions, suffer from various limitations. One of the biggest drawbacks concerning RNA modelling is the relative scarcity of experimentally determined RNA and RNP complex structures that can be used for training and testing the methods as well as templates in comparative modelling approaches. The value of RNP structures for the community of computational biologists can be illustrated by the repeated calls from organizers of the CAPRI and RNA Puzzles initiatives to experimental researchers for proposing the newly determined structures as prediction targets. One can hope that the significance of this problem will wane with the growing interest in RNA structural biology, leading to an increase in the availability of solved structures solved each year.

In order to obtain biologically, chemically and physically relevant predictions, it may be advantageous to combine various existing methods for docking and scoring. Such meta-prediction was successfully applied in structural bioinformatics for modelling protein structures [139] and for protein-protein docking [140]. Thus, RNP docking can be performed using different methods and the top scored docking poses from each of these studies can be selected. The top decoys can be then re-scored using various scoring functions. The top scored decoys from the various scoring experiments can be chosen for further analysis, for example, clustering. If the scoring methods reach consensus, then poses obtained from different methods can be clustered together. In the absence of a consensus scoring,

top models proposed by different methods can be suggested as alternative solutions. The proposed RNP meta-docking workflow is represented schematically in Figure 4.



**Figure 4.** Schematic representation of the workflow for RNP docking. The docking strategy presented here combines the strengths of several docking and scoring methods.

One of the biggest challenges in docking (and in particular in RNA-protein docking) is the molecular flexibility [125] and the computational complexity associated with flexible docking. The current methods for scoring models of RNP complexes are quite accurate for evaluating poses generated by rigid docking methods. However, generating conformations that are closer to the bound conformation than the starting unbound structure and discriminating these conformations from all the others is a daunting task for computational methods. The existing computational docking algorithms seldom take into account conformational changes that may occur upon binding of the ligand to the receptor, in RNA and/or protein component(s). Similarly, the available scoring methods also have limited discriminative power to identify near-native structures when the binding protein and/or RNA undergoes large conformational changes. A possible solution to this issue is to combine the existing tools that enable template-free modelling of the protein and RNA components, with scoring functions for the assessment of intermolecular contacts. As a first step towards this approach, modelling techniques have been developed that accept models obtained by for example, rigid body docking of

“unbound” protein and RNA structures as input and perform only local refolding of protein and RNA molecules directly involved in interactions.

The conceptual similarity of successful algorithms for structural modelling of protein and RNA 3D structures [141] suggests the feasibility of combining them into unified modelling methods. Combining theoretical predictive methods with low-resolution experimental analyses is also expected to provide synergy to such attempts. Recently, it was demonstrated that the structures of many large RNP complexes, such as the spliceosome, may be modelled using cryo-EM maps as molecular envelopes. The structures of components could be fitted into such envelopes, using restraints from biochemical experiments and other bioinformatics-based predictions [142]. In order to achieve this, new multi-resolution modelling methods and new ways of encoding experimental data are required [143]. We hope that the recent surge of interest in studying RNP interactions will encourage both biologists and software developers alike, to use bioinformatics tools for obtaining structural insights into the biological systems, guided by available experimental data, as well as in proposing and developing new algorithms and their user-friendly implementations.

**Funding:** The research on RNA structure and interactions in the Bujnicki laboratory was funded primarily by the Foundation for Polish Science (FNP, grants TEAM/2009-4/2 and TEAM/2016-3/18) and by the European Research Council (ERC, StG grant RNA + P = 123D) and it is currently supported by the Polish National Science Center (NCN, grant MAESTRO 2017/26/A/NZ1/01083). J.M.B. was also supported by the “Ideas for Poland” fellowship from the FNP. C.N. was supported by IIMCB statutory funds. P.G. was supported by the Foundation for Polish Science grant TEAM/2016-3/18.

**Acknowledgments:** We thank J. Mahita, Sunandan Mukherjee and A. Gandhimathi for critical reading of the manuscript and their useful comments. We thank J. Mahita and Soma Deoghuria for linguistic feedback on the manuscript. We also thank our colleagues working on RNA-protein interactions for their inspiration, in particular, Michał Boniecki (IIMCB).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Morris, K.V.; Mattick, J.S. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15*, 423–437. [[CrossRef](#)] [[PubMed](#)]
2. Koonin, E.V. RNA Worlds: From Life’s Origins to Diversity in Gene Regulation. *Q. Rev. Biol.* **2012**, *87*, 66. [[CrossRef](#)]
3. Chen, Y.; Varani, G. Protein families and RNA recognition. *FEBS J.* **2005**, *272*, 2088–2097. [[CrossRef](#)] [[PubMed](#)]
4. Noller, H.F. Ribosomal RNA and translation. *Annu. Rev. Biochem.* **1991**, *60*, 191–227. [[CrossRef](#)] [[PubMed](#)]
5. Guthrie, C. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science* **1991**, *253*, 157–163. [[CrossRef](#)] [[PubMed](#)]
6. Reichert, V.L.; Le Hir, H.; Jurica, M.S.; Moore, M.J. 5’ exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev.* **2002**, *16*, 2778–2791. [[CrossRef](#)] [[PubMed](#)]
7. Glisovic, T.; Bachorik, J.L.; Yong, J.; Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **2008**, *582*, 1977–1986. [[CrossRef](#)] [[PubMed](#)]
8. Johnson JM, E. al Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **2003**, *302*, 2141–2144. [[CrossRef](#)] [[PubMed](#)]
9. Minvielle-Sebastia, L.; Keller, W. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* **1999**, *11*, 352–357. [[CrossRef](#)]
10. Valente, L.; Nishikura, K. ADAR Gene Family and A-to-I RNA Editing: Diverse Roles in Posttranscriptional Gene Regulation. Available online: <https://www.ncbi.nlm.nih.gov/pubmed/16096031/> (accessed on 21 July 2018).
11. Hogg, J.R.; Collins, K. Structured non-coding RNAs and the RNP Renaissance. *Curr. Opin. Chem. Biol.* **2008**, *12*, 684–689. [[CrossRef](#)] [[PubMed](#)]
12. Peterlin, B.M.; Brogie, J.E.; Price, D.H. 7SK snRNA: A noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 92–103. [[CrossRef](#)] [[PubMed](#)]

13. Swain, A.; Misulovin, Z.; Pherson, M.; Gause, M.; Mihindukulasuriya, K.; Rickels, R.A.; Shilatifard, A.; Dorsett, D. Drosophila TDP-43 RNA-Binding Protein Facilitates Association of Sister Chromatid Cohesion Proteins with Genes, Enhancers and Polycomb Response Elements. *PLoS Genet.* **2016**, *12*, e1006331. [[CrossRef](#)] [[PubMed](#)]
14. Lukong, K.E.; Chang, K.-W.; Khandjian, E.W.; Richard, S. RNA-binding proteins in human genetic disease. *Trends Genet.* **2008**, *24*, 416–425. [[CrossRef](#)] [[PubMed](#)]
15. Cooper, T.A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777–793. [[CrossRef](#)] [[PubMed](#)]
16. Lunde, B.M.; Moore, C.; Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 479–490. [[CrossRef](#)] [[PubMed](#)]
17. Cléry, A.-T.; Allain, F.H. From Structure to function of RNA binding domains. In *Madame Curie Bioscience Database [Internet]*; Landes Bioscience: Austin, TX, USA, 2013.
18. Maris, C.; Dominguez, C.; Allain, F.H.-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118–2131. [[CrossRef](#)] [[PubMed](#)]
19. Siomi, H.; Matunis, M.J.; Matthew Michael, W.; Dreyfuss, G. The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Res.* **1993**, *21*, 1193–1198. [[CrossRef](#)] [[PubMed](#)]
20. Ryter, J.M.; Schultz, S.C. Molecular basis of double-stranded RNA-protein interactions: Structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* **1998**, *17*, 7505–7513. [[CrossRef](#)] [[PubMed](#)]
21. Lu, D.; Searles, M.A.; Klug, A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **2003**, *426*, 96–100. [[CrossRef](#)] [[PubMed](#)]
22. Steitz, T.A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253. [[CrossRef](#)] [[PubMed](#)]
23. Sperling, J.; Azubel, M.; Sperling, R. Structure and function of the Pre-mRNA splicing machine. *Structure* **2008**, *16*, 1605–1615. [[CrossRef](#)] [[PubMed](#)]
24. Licatalosi, D.D.; Darnell, R.B. RNA processing and its regulation: Global insights into biological networks. *Nat. Rev. Genet.* **2010**, *11*, 75–87. [[CrossRef](#)] [[PubMed](#)]
25. Burd, C.G.; Matunis, E.L.; Dreyfuss, G. The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. *Mol. Cell. Biol.* **1991**, *11*, 3419–3424. [[CrossRef](#)] [[PubMed](#)]
26. Ban, N.; Nissen, P.; Hansen, J.; Moore, P.B.; Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **2000**, *289*, 905–920. [[CrossRef](#)] [[PubMed](#)]
27. Ke, A.; Doudna, J.A. Crystallization of RNA and RNA-protein complexes. *Methods* **2004**, *34*, 408–414. [[CrossRef](#)] [[PubMed](#)]
28. Scott, L.G.; Hennig, M. RNA Structure Determination by NMR. *Methods Mol. Biol.* **2008**, *452*, 29–61. [[PubMed](#)]
29. Garman, E.F. Developments in X-ray crystallographic structure determination of biological macromolecules. *Science* **2014**, *343*, 1102–1108. [[CrossRef](#)] [[PubMed](#)]
30. Lapinaite, A.; Simon, B.; Skjaerven, L.; Rakwalska-Bange, M.; Gabel, F.; Carlomagno, T. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **2013**, *502*, 519–523. [[CrossRef](#)] [[PubMed](#)]
31. Duss, O.; Yulikov, M.; Jeschke, G.; Allain, F.H.-T. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nat. Commun.* **2014**, *5*, 3669. [[CrossRef](#)] [[PubMed](#)]
32. Patel, T.R.; Chojnowski, G.; Koul, A.; McKenna, S.A.; Bujnicki, J.M. Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods* **2017**, *118–119*, 146–162. [[CrossRef](#)] [[PubMed](#)]
33. Cook, K.B.; Kazan, H.; Zuberi, K.; Morris, Q.; Hughes, T.R. RBPDB: A database of RNA-binding specificities. *Nucleic Acids Res.* **2011**, *39*, D301–D308. [[CrossRef](#)] [[PubMed](#)]
34. Baltz, A.G.; Munschauer, M.; Schwanhäusser, B.; Vasilic, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **2012**, *46*, 674–690. [[CrossRef](#)] [[PubMed](#)]
35. Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B.M.; Strein, C.; Davey, N.E.; Humphreys, D.T.; Preiss, T.; Steinmetz, L.M.; et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **2012**, *149*, 1393–1406. [[CrossRef](#)] [[PubMed](#)]
36. Castello, A.; Hentze, M.W.; Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol. Metab.* **2015**, *26*, 746–757. [[CrossRef](#)] [[PubMed](#)]
37. Zhao, H.; Yang, Y.; Janga, S.C.; Kao, C.C.; Zhou, Y. Prediction and validation of the unexplored RNA-binding protein atlas of the human proteome. *Proteins* **2014**, *82*, 640–647. [[CrossRef](#)] [[PubMed](#)]



38. Gerstberger, S.; Hafner, M.; Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **2014**, *15*, 829–845. [[CrossRef](#)] [[PubMed](#)]
39. Ghosh, P.; Sowdhamini, R. Genome-wide survey of putative RNA-binding proteins encoded in the human proteome. *Mol. Biosyst.* **2016**, *12*, 532–540. [[CrossRef](#)] [[PubMed](#)]
40. Ghosh, P.; Murugavel, P.; Sowdhamini, R. hRBPome: A central repository of all known human RNA-binding proteins. *BioRxiv* **2018**. [[CrossRef](#)]
41. Liao, Y.; Castello, A.; Fischer, B.; Leicht, S.; Föehr, S.; Frese, C.K.; Ragan, C.; Kurscheid, S.; Pagler, E.; Yang, H.; et al. The cardiomyocyte RNA-binding proteome: Links to intermediary metabolism and heart disease. *Cell Rep.* **2016**, *16*, 1456–1469. [[CrossRef](#)] [[PubMed](#)]
42. Conrad, T.; Albrecht, A.-S.; de Melo Costa, V.R.; Sauer, S.; Meierhofer, D.; Ørom, U.A. Serial interactome capture of the human cell nucleus. *Nat. Commun.* **2016**, *7*, 11212. [[CrossRef](#)] [[PubMed](#)]
43. Liepelt, A.; Naarmann-de Vries, I.S.; Simons, N.; Eichelbaum, K.; Föhr, S.; Archer, S.K.; Castello, A.; Usadel, B.; Krijgsveld, J.; Preiss, T.; et al. Identification of RNA-binding proteins in macrophages by interactome capture. *Mol. Cell. Proteom.* **2016**, *15*, 2699–2714. [[CrossRef](#)] [[PubMed](#)]
44. Johnson, M.S.; Srinivasan, N.; Sowdhamini, R.; Blundell, T.L. Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* **1994**, *29*, 1–68. [[CrossRef](#)] [[PubMed](#)]
45. Hardin, C.; Pogorelov, T.V.; Luthey-Schulten, Z. Ab initio protein structure prediction. *Curr. Opin. Struct. Biol.* **2002**, *12*, 176–181. [[CrossRef](#)]
46. Tozzini, V. Multiscale modeling of proteins. *Acc. Chem. Res.* **2010**, *43*, 220–230. [[CrossRef](#)] [[PubMed](#)]
47. Vakser, I.A. Protein-protein docking: From interaction to interactome. *Biophys. J.* **2014**, *107*, 1785–1793. [[CrossRef](#)] [[PubMed](#)]
48. Moreira, I.S.; Fernandes, P.A.; Ramos, M.J. Protein-protein docking dealing with the unknown. *J. Comput. Chem.* **2010**, *31*, 317–342. [[CrossRef](#)] [[PubMed](#)]
49. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking: A review. *Biophys. Rev.* **2017**, *9*, 91–102. [[CrossRef](#)] [[PubMed](#)]
50. Miao, Z.; Westhof, E. RNA structure: Advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **2017**, *46*, 483–503. [[CrossRef](#)] [[PubMed](#)]
51. Seetin, M.G.; Mathews, D.H. RNA Structure Prediction: An Overview of Methods. *Methods Mol. Biol.* **2012**, *905*, 99–122. [[CrossRef](#)] [[PubMed](#)]
52. Dawson, W.K.; Bujnicki, J.M. Computational modeling of RNA 3D structures and interactions. *Curr. Opin. Struct. Biol.* **2016**, *37*, 22–28. [[CrossRef](#)] [[PubMed](#)]
53. Madan, B.; Kasprzak, J.M.; Tuszynska, I.; Magnus, M.; Szczepaniak, K.; Dawson, W.K.; Bujnicki, J.M. Modeling of Protein–RNA Complex Structures Using Computational Docking Methods. *Methods Mol. Biol.* **2016**, *1414*, 353–372. [[CrossRef](#)] [[PubMed](#)]
54. Moulton, J.; Pedersen, J.T.; Judson, R.; Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins* **1995**, *23*, ii–v. [[CrossRef](#)] [[PubMed](#)]
55. Cruz, J.A.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J.M.; Chen, S.-J.; Cao, S.; Das, R.; Ding, F.; Dokholyan, N.V.; Flores, S.C.; et al. RNA-puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* **2012**, *18*, 610–625. [[CrossRef](#)] [[PubMed](#)]
56. Moulton, J.; Fidelis, K.; Kryshchuk, A.; Schwede, T.; Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)-Round XII. *Proteins* **2018**, *86* (Suppl. 1), 7–15. [[CrossRef](#)] [[PubMed](#)]
57. Miao, Z.; Adamiak, R.W.; Antczak, M.; Batey, R.T.; Becka, A.J.; Biesiada, M.; Boniecki, M.J.; Bujnicki, J.M.; Chen, S.-J.; Cheng, C.Y.; et al. RNA-puzzles round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **2017**, *23*, 655–672. [[CrossRef](#)] [[PubMed](#)]
58. Vajda, S.; Hall, D.R.; Kozakov, D. Sampling and scoring: A marriage made in heaven. *Proteins* **2013**, *81*, 1874–1884. [[CrossRef](#)] [[PubMed](#)]
59. Comeau, S.R.; Gatchell, D.W.; Vajda, S.; Camacho, C.J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **2004**, *32*, W96–W99. [[CrossRef](#)] [[PubMed](#)]
60. Tuszynska, I.; Magnus, M.; Jonak, K.; Dawson, W.; Bujnicki, J.M. NPDock: a web server for protein–nucleic acid docking. *Nucleic Acids Res.* **2015**, *43*, W425–W430. [[CrossRef](#)] [[PubMed](#)]
61. Van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastriitis, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [[CrossRef](#)] [[PubMed](#)]

62. Tuszynska, I.; Bujnicki, J.M. DARS-RNP and QUASI-RNP: New statistical potentials for protein-RNA docking. *BMC Bioinform.* **2011**, *12*, 348. [[CrossRef](#)] [[PubMed](#)]
63. Li, C.H.; Cao, L.B.; Su, J.G.; Yang, Y.X.; Wang, C.X. A new residue-nucleotide propensity potential with structural information considered for discriminating protein-RNA docking decoys. *Proteins* **2012**, *80*, 14–24. [[CrossRef](#)] [[PubMed](#)]
64. Huang, Y.; Liu, S.; Guo, D.; Li, L.; Xiao, Y. A novel protocol for three-dimensional structure prediction of RNA-protein complexes. *Sci. Rep.* **2013**, *3*, 1887. [[CrossRef](#)] [[PubMed](#)]
65. Huang, S.-Y.; Zou, X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* **2014**, *42*, e55. [[CrossRef](#)] [[PubMed](#)]
66. Bastard, K.; Saladin, A.; Prévost, C. Accounting for large amplitude protein deformation during in silico macromolecular docking. *Int. J. Mol. Sci.* **2011**, *12*, 1316–1333. [[CrossRef](#)] [[PubMed](#)]
67. Lensink, M.F.; Wodak, S.J. Docking and scoring protein interactions: CAPRI 2009. *Proteins* **2010**, *78*, 3073–3084. [[CrossRef](#)] [[PubMed](#)]
68. Zacharias, M. Accounting for conformational changes during protein–protein docking. *Curr. Opin. Struct. Biol.* **2010**, *20*, 180–186. [[CrossRef](#)] [[PubMed](#)]
69. Janin, J.; Henrick, K.; Moult, J.; Ten Eyck, L.; Sternberg, M.J.E.; Vajda, S.; Vakser, I.; Wodak, S.J. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins Struct. Funct. Genet.* **2003**, *52*, 2–9. [[CrossRef](#)] [[PubMed](#)]
70. Lensink, M.F.; Wodak, S.J. Docking, scoring, and affinity prediction in CAPRI. *Proteins* **2013**, *81*, 2082–2095. [[CrossRef](#)] [[PubMed](#)]
71. Huang, Y.; Li, H.; Xiao, Y. Using 3dRPC for RNA–protein complex structure prediction. *Biophys. Rep.* **2016**, *2*, 95–99. [[CrossRef](#)] [[PubMed](#)]
72. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.A.; Aflalo, C.; Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2195–2199. [[CrossRef](#)] [[PubMed](#)]
73. Tovchigrechko, A.; Vakser, I.A. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* **2006**, *34*, W310–W314. [[CrossRef](#)] [[PubMed](#)]
74. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.-D.; Ritchie, D.W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **2010**, *38*, W445–W449. [[CrossRef](#)] [[PubMed](#)]
75. Arnautova, Y.A.; Abagyan, R.; Totrov, M. Protein-RNA docking using ICM. *J. Chem. Theory Comput.* **2018**. [[CrossRef](#)] [[PubMed](#)]
76. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [[CrossRef](#)] [[PubMed](#)]
77. Neveu, E.; Ritchie, D.W.; Popov, P.; Grudinin, S. PEPsi-Dock: A detailed data-driven protein–protein interaction potential accelerated by polar Fourier correlation. *Bioinformatics* **2016**, *32*, i693–i701. [[CrossRef](#)] [[PubMed](#)]
78. Jiménez-García, B.; Pons, C.; Fernández-Recio, J. pyDockWEB: A web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. *Bioinformatics* **2013**, *29*, 1698–1699. [[CrossRef](#)] [[PubMed](#)]
79. Guilhot-Gaudeffroy, A.; Froidevaux, C.; Azé, J.; Bernauer, J. Protein-RNA complexes and efficient automatic docking: expanding RosettaDock possibilities. *PLoS ONE* **2014**, *9*, e108928. [[CrossRef](#)] [[PubMed](#)]
80. Pierce, B.G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **2014**, *30*, 1771–1773. [[CrossRef](#)] [[PubMed](#)]
81. De Vries, S.J.; Schindler, C.E.M.; Chauvot de Beauchêne, I.; Zacharias, M. A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.* **2015**, *108*, 462–465. [[CrossRef](#)] [[PubMed](#)]
82. Dominguez, C.; Boelens, R.; Alexandre, M.J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [[CrossRef](#)] [[PubMed](#)]
83. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDock: A web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373. [[CrossRef](#)] [[PubMed](#)]
84. Kozakov, D.; Brenke, R.; Comeau, S.R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins* **2006**, *65*, 392–406. [[CrossRef](#)] [[PubMed](#)]

85. *A Powerful and Innovative Package for Accurate Protein Structure Predictions*; Schrödinger, LLC Prime: New York, NY, USA, 2018.
86. Bonvin, A.M.J.J.; Karaca, E.; Kastritis, P.L.; Rodrigues, J.P.G.L.M. Defining distance restraints in HADDOCK. *Nat. Protoc.* **2018**, *13*, 1503. [[CrossRef](#)] [[PubMed](#)]
87. Van Dijk, M.; van Dijk, A.D.J.; Hsu, V.; Boelens, R.; Bonvin, A.M.J.J. Information-driven protein-DNA docking using HADDOCK: It is a matter of flexibility. *Nucleic Acids Res.* **2006**, *34*, 3317–3325. [[CrossRef](#)] [[PubMed](#)]
88. Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **2011**, *9*, 173–175. [[CrossRef](#)] [[PubMed](#)]
89. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [[CrossRef](#)] [[PubMed](#)]
90. Kundrotas, P.J.; Zhu, Z.; Janin, J.; Vakser, I.A. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9438–9441. [[CrossRef](#)] [[PubMed](#)]
91. Zheng, J.; Kundrotas, P.J.; Vakser, I.A.; Liu, S. Template-Based Modeling of Protein-RNA Interactions. *PLoS Comput. Biol.* **2016**, *12*, e1005120. [[CrossRef](#)] [[PubMed](#)]
92. Ghosh, P.; Mathew, O.K.; Sowdhamini, R. RStrucFam: A web server to associate structure and cognate RNA for RNA-binding proteins from sequence information. *BMC Bioinform.* **2016**, *17*, 411. [[CrossRef](#)] [[PubMed](#)]
93. Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *Wiley Interdiscip. Rev. RNA* **2017**, *8*. [[CrossRef](#)] [[PubMed](#)]
94. Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R.A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; et al. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118*, 4177–4338. [[CrossRef](#)] [[PubMed](#)]
95. Iwakiri, J.; Hamada, M.; Asai, K.; Kameda, T. Improved Accuracy in RNA-Protein Rigid Body Docking by Incorporating Force Field for Molecular Dynamics Simulation into the Scoring Function. *J. Chem. Theory Comput.* **2016**, *12*, 4688–4697. [[CrossRef](#)] [[PubMed](#)]
96. Ren, X.; Zeng, R.; Tortorella, M.; Wang, J.; Wang, C. Structural Insight into Inhibition of CsrA-RNA Interaction Revealed by Docking, Molecular Dynamics and Free Energy Calculations. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
97. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [[CrossRef](#)] [[PubMed](#)]
98. Karaca, E.; Rodrigues, J.P.G.L.M.; Graziadei, A.; Bonvin, A.M.J.J.; Carlomagno, T. M3: An integrative framework for structure determination of molecular machines. *Nat. Methods* **2017**, *14*, 897–902. [[CrossRef](#)] [[PubMed](#)]
99. Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **2012**, *10*, e1001244. [[CrossRef](#)] [[PubMed](#)]
100. Deo, S.; Patel, T.R.; Chojnowski, G.; Koul, A.; Džananović, E.; McEleney, K.; Bujnicki, J.M.; McKenna, S.A. Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2' 5'-oligoadenylate synthetase family. *J. Struct. Biol.* **2015**, *190*, 236–249. [[CrossRef](#)] [[PubMed](#)]
101. Yan, J.; Friedrich, S.; Kurgan, L. A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues. *Brief. Bioinform.* **2016**, *17*, 88–105. [[CrossRef](#)] [[PubMed](#)]
102. Miao, Z.; Westhof, E. A Large-Scale Assessment of Nucleic Acids Binding Site Prediction Programs. *PLoS Comput. Biol.* **2015**, *11*, e1004639. [[CrossRef](#)] [[PubMed](#)]
103. Gajda, M.J.; Tuszyńska, I.; Kaczor, M.; Bakulina, A.Y.; Bujnicki, J.M. FILTREST3D: discrimination of structural models using restraints from experimental data. *Bioinformatics* **2010**, *26*, 2986–2987. [[CrossRef](#)] [[PubMed](#)]
104. Chelliah, V.; Blundell, T.L.; Fernández-Recio, J. Efficient restraints for protein-protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment. *J. Mol. Biol.* **2006**, *357*, 1669–1682. [[CrossRef](#)] [[PubMed](#)]
105. Chen, Y.; Kortemme, T.; Robertson, T.; Baker, D.; Varani, G. A new hydrogen-bonding potential for the design of protein-RNA interactions predicts specific contacts and discriminates decoys. *Nucleic Acids Res.* **2004**, *32*, 5147–5162. [[CrossRef](#)] [[PubMed](#)]
106. Zheng, S.; Robertson, T.A.; Varani, G. A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J.* **2007**, *274*, 6378–6391. [[CrossRef](#)] [[PubMed](#)]

107. Pérez-Cano, L.; Solernou, A.; Pons, C.; Fernández-Recio, J. Structural prediction of protein-RNA interaction by computational docking with propensity-based statistical potentials. In *Biocomputing 2010*; World Scientific: Singapore, 2009; pp. 293–301. ISBN 9789814299473.
108. Zhao, H.; Yang, Y.; Zhou, Y. Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. *Nucleic Acids Res.* **2011**, *39*, 3017–3025. [[CrossRef](#)] [[PubMed](#)]
109. Setny, P.; Zacharias, M. A coarse-grained force field for Protein–RNA docking. *Nucleic Acids Res.* **2011**, *39*, 9118–9129. [[CrossRef](#)] [[PubMed](#)]
110. Barik, A.C.N.; Pilla, S.P.; Bahadur, R.P. Molecular architecture of protein-RNA recognition sites. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2738–2751. [[CrossRef](#)] [[PubMed](#)]
111. Treger, M.; Westhof, E. Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recognit.* **2001**, *14*, 199–214. [[CrossRef](#)] [[PubMed](#)]
112. Zhou, H.; Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* **2002**, *11*, 2714–2726. [[CrossRef](#)] [[PubMed](#)]
113. Lorenzen, S.; Zhang, Y. Identification of near-native structures by clustering protein docking conformations. *Proteins Struct. Funct. Bioinf.* **2007**, *68*, 187–194. [[CrossRef](#)] [[PubMed](#)]
114. Barik, A.; C, N.; P, M.; Bahadur, R.P. A protein-RNA docking benchmark (I): Nonredundant cases. *Proteins* **2012**, *80*, 1866–1871. [[CrossRef](#)] [[PubMed](#)]
115. Barik, A.; Mishra, A.; Bahadur, R.P. PRince: a web server for structural and physicochemical analysis of protein-RNA interface. *Nucleic Acids Res.* **2012**, *40*, W440–W444. [[CrossRef](#)] [[PubMed](#)]
116. Nithin, C.; Mukherjee, S.; Bahadur, R.P. A non-redundant protein–RNA docking benchmark version 2.0. *Proteins Struct. Funct. Bioinf.* **2017**, *85*, 256–267. [[CrossRef](#)] [[PubMed](#)]
117. Pérez-Cano, L.; Jiménez-García, B.; Fernández-Recio, J. A protein-RNA docking benchmark (II): Extended set from experimental and homology modeling data. *Proteins* **2012**, *80*, 1872–1882. [[CrossRef](#)] [[PubMed](#)]
118. Huang, S.-Y.; Zou, X. A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J. Comput. Chem.* **2013**, *34*, 311–318. [[CrossRef](#)] [[PubMed](#)]
119. Yan, Z.; Wang, J. Optimizing scoring function of protein-nucleic acid interactions with both affinity and specificity. *PLoS One* **2013**, *8*, e74443. [[CrossRef](#)] [[PubMed](#)]
120. Pérez-Cano, L.; Romero-Durana, M.; Fernández-Recio, J. Structural and energy determinants in protein-RNA docking. *Methods* **2017**, *118–119*, 163–170. [[CrossRef](#)] [[PubMed](#)]
121. Miao, Z.; Westhof, E. Prediction of nucleic acid binding probability in proteins: A neighboring residue network based score. *Nucleic Acids Res.* **2015**, *43*, 5340–5351. [[CrossRef](#)] [[PubMed](#)]
122. Wang, K.; Jian, Y.; Wang, H.; Zeng, C.; Zhao, Y. RBind: Computational network method to predict RNA binding sites. *Bioinformatics* **2018**. [[CrossRef](#)] [[PubMed](#)]
123. Barik, A.; Bahadur, R.P. Hydration of protein–RNA recognition sites. *Nucleic Acids Res.* **2014**, *42*, 10148–10160. [[CrossRef](#)] [[PubMed](#)]
124. Barik, A.; Nithin, C.; Karampudi, N.B.R.; Mukherjee, S.; Bahadur, R.P. Probing binding hot spots at protein-RNA recognition sites. *Nucleic Acids Res.* **2016**, *44*, e9. [[CrossRef](#)] [[PubMed](#)]
125. Mukherjee, S.; Bahadur, R.P. An account of solvent accessibility in protein-RNA recognition. *Sci. Rep.* **2018**, *8*, 10546. [[CrossRef](#)] [[PubMed](#)]
126. Cheng, W.; Yan, C. A graph approach to mining biological patterns in the binding interfaces. *J. Comput. Biol.* **2017**, *24*, 31–39. [[CrossRef](#)] [[PubMed](#)]
127. Iwakiri, J.; Kameda, T.; Asai, K.; Hamada, M. Analysis of base-pairing probabilities of RNA molecules involved in protein–RNA interactions. *Bioinformatics* **2013**, *29*, 2524–2528. [[CrossRef](#)] [[PubMed](#)]
128. Yang, X.; Li, H.; Huang, Y.; Liu, S. The dataset for protein-RNA binding affinity. *Protein Sci.* **2013**, *22*, 1808–1811. [[CrossRef](#)] [[PubMed](#)]
129. Pan, Y.; Wang, Z.; Zhan, W.; Deng, L. Computational identification of binding energy hot spots in protein–RNA complexes using an ensemble approach. *Bioinformatics* **2017**, *34*, 1473–1480. [[CrossRef](#)] [[PubMed](#)]
130. Liu, L.; Xiong, Y.; Gao, H.; Wei, D.-Q.; Mitchell, J.C.; Zhu, X. dbAMEPNI: A database of alanine mutagenic effects for protein–nucleic acid interactions. *Database* **2018**, *2018*. [[CrossRef](#)] [[PubMed](#)]
131. Pires, D.E.V.; Ascher, D.B. mCSM–NA: predicting the effects of mutations on protein–nucleic acids interactions. *Nucleic Acids Res.* **2017**, *45*, W241–W246. [[CrossRef](#)] [[PubMed](#)]

132. Główny, D.; Pianka, D.; Sulej, A.A.; Kozłowski, Ł. P.; Czarnecka, J.; Chojnowski, G.; Skowronek, K.J.; Bujnicki, J.M. Sequence-specific cleavage of dsRNA by Mini-III RNase. *Nucleic Acids Res.* **2015**, *43*, 2864–2873. [[CrossRef](#)] [[PubMed](#)]
133. Główny, D.; Kurkowska, M.; Czarnecka, J.; Szczepaniak, K.; Pianka, D.; Kappert, V.; Bujnicki, J.M.; Skowronek, K.J. Identification of protein structural elements responsible for the diversity of sequence preferences among Mini-III RNases. *Sci. Rep.* **2016**, *6*, 38612. [[CrossRef](#)] [[PubMed](#)]
134. Smietanski, M.; Werner, M.; Purta, E.; Kaminska, K.H.; Stepinski, J.; Darzynkiewicz, E.; Nowotny, M.; Bujnicki, J.M. Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. *Nat. Commun.* **2014**, *5*, 3004. [[CrossRef](#)] [[PubMed](#)]
135. Van Laer, B.; Roovers, M.; Wauters, L.; Kasprzak, J.M.; Dyzma, M.; Deyaert, E.; Kumar Singh, R.; Feller, A.; Bujnicki, J.M.; Droogmans, L.; et al. Structural and functional insights into tRNA binding and adenosine N1-methylation by an archaeal Trm10 homologue. *Nucleic Acids Res.* **2016**, *44*, 940–953. [[CrossRef](#)] [[PubMed](#)]
136. Maravić, G.; Bujnicki, J.M.; Feder, M.; Pongor, S.; Flögel, M. Alanine-scanning mutagenesis of the predicted rRNA-binding domain of ErmC' redefines the substrate-binding site and suggests a model for protein-RNA interactions. *Nucleic Acids Res.* **2003**, *31*, 4941–4949. [[CrossRef](#)] [[PubMed](#)]
137. Purta, E.; Kaminska, K.H.; Kasprzak, J.M.; Bujnicki, J.M.; Douthwaite, S. YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA* **2008**, *14*, 2234–2244. [[CrossRef](#)] [[PubMed](#)]
138. Husain, N.; Obranic, S.; Kosciński, L.; Seetharaman, J.; Babic, F.; Bujnicki, J.M.; Maravic-Vlahovick, G.; Sivaraman, J. Structural basis for the methylation of A1408 in 16S rRNA by a panaminoglycoside resistance methyltransferase NpmA from a clinical isolate and analysis of the NpmA interactions with the 30S ribosomal subunit. *Nucleic Acids Res.* **2011**, *39*, 1903–1918. [[CrossRef](#)] [[PubMed](#)]
139. Kurowski, M.A.; Bujnicki, J.M. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.* **2003**, *31*, 3305–3307. [[CrossRef](#)] [[PubMed](#)]
140. Anashkina, A.A.; Kravatsky, Y.; Kuznetsov, E.; Makarov, A.A.; Adzhubei, A.A. Meta-server for automatic analysis, scoring and ranking of docking models. *Bioinformatics* **2017**. [[CrossRef](#)] [[PubMed](#)]
141. Rother, K.; Rother, M.; Boniecki, M.; Puton, T.; Bujnicki, J.M. RNA and protein 3D structure modeling: Similarities and differences. *J. Mol. Model.* **2011**, *17*, 2325–2336. [[CrossRef](#)] [[PubMed](#)]
142. Jurica, M.S. Detailed close-ups and the big picture of spliceosomes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 315–320. [[CrossRef](#)] [[PubMed](#)]
143. Flores, S.C.; Bernauer, J.; Shin, S.; Zhou, R.; Huang, X. Multiscale modeling of macromolecular biosystems. *Brief. Bioinform.* **2012**, *13*, 395–405. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).